HistoEncoder: a digital pathology foundation model for prostate cancer

Joona Pohjonen[1,*], Abderrahim-Oussama Batouche[1,2,*], Antti Rannikko[1,3,4], Kevin Sandeman[1,5], Andrew Erickson[1], Esa Pitkänen[7,6,4,†], and Tuomas Mirtti[1,4,8,†]

[1]Research Program in Systems Oncology, Faculty of Medicine, University of Helsinki
[2]Doctoral Program in Computer Science, University of Helsinki
[3]Department of Urology, Helsinki University Hospital
[4]iCAN Digital Precision Cancer Medicine Flagship, Finland
[5]Department of Pathology, Division of Laboratory Medicine, Skåne University Hospital, Malmö, Sweden
[6]Research Program in Applied Tumor Genomics, Faculty of Medicine, University of Helsinki
[7]Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki
[8]Department of Pathology, Helsinki University Hospital

Abstract

Foundation models are trained on massive amounts of data to distinguish complex patterns and can be adapted to a wide range of downstream tasks with minimal computational resources. Here, we develop a foundation model for prostate cancer digital pathology called HistoEncoder by pre-training on 48 million prostate tissue tile images. We demonstrate that HistoEncoder features extracted from tile images with similar histological patterns map closely together in the feature space. HistoEncoder outperforms models pre-trained with natural images, even without fine-tuning or with 1000 times less training data. We describe two use cases that leverage the capabilities of HistoEncoder by fine-tuning the model with a limited amount of data and computational resources. First, we show how HistoEncoder can be used to automatically annotate large-scale datasets with high accuracy. Second, we combine histomics with commonly used clinical nomograms, significantly improving prostate cancer-specific death survival models. Foundation models such as HistoEncoder can allow organizations with limited resources to build effective clinical software tools without needing extensive datasets or significant amounts of computing.

1 Introduction

Neural network-based solutions [1] have achieved impressive results with tissue diagnostics, often surpassing human counterparts in consistency, speed and accuracy [2–4]. For instance, a multiple instance learning model by Esteva *et al.*, trained and validated with clinical trial histological images and data, showed that prostate morphological features learnt by the model contain predictive information beyond conventional nomograms [5]. A subsequent AI biomarker, predicting the benefit of adding androgen deprivation therapy to radiation therapy, primarily includes conventional Gleason patterns and image features extracted with a neural network as key components of the model [6].

Despite promising results, recent work has demonstrated that neural networks perform substantially worse on datasets not used during training [7–11]. A major driver of this performance reduction is dataset shift [12, 13], where neural networks fail to generalize to data from a new clinical setting that differs from the training data. Fine-tuning pre-trained neural networks for downstream tasks has been shown to improve model robustness and reduce uncertainty [14]. Many recent publications have confirmed this by leveraging neural networks pre-trained on natural images [15–26]. Still, due to the considerable domain shift between histological and natural images, transfer learning from a neural network pre-trained with natural images offers little benefit to performance [27]. Thus, there is a need for neural networks pre-trained with histological images.

Recent advances in self-supervised learning [28–32] have resulted in the emergence of foundation models [33–36]. Foundation models are trained on large-scale datasets, leveraging unlabelled samples via self-supervised learning [37, 38], and can be then easily adapted for downstream tasks, even without any additional training [31, 32].

Here, we train foundation models on 48 million tile images from thousands of histological slide images with prostate tissue (fig. 1). These foundation models outperform models pre-trained with natural images by a large margin. Additionally, we describe two workflows leveraging foundation models. The first workflow describes a method for automatically annotating large-scale tissue image datasets, which we evaluate by annotating the largest publicly available histological dataset [39] with high accuracy. The second workflow integrates annotated significant histology features with commonly used clinical prognostic nomograms to improve the prediction of prostate cancer-specific mortality.

2 Materials and methods

2.1 *HistoEncoder models*

In this study, we train cross-covariance image transformers (XCiT) [42] with a self-supervised learning method DINO [31, 32], which leverages discriminative signals between groups of images to learn useful features using self-distillation. Instead of focusing on maximizing the fine-tuning performance on a downstream task [28, 29], the encoder model training aims to maximize the representativeness of extracted features [32]. We selected XCiT due to linear complexity in the number of tokens, allowing efficient processing of high-resolution images

[*] These authors contributed equally.
[†] Correspondence:
esa.pitkanen@helsinki.fi, tuomas.mirtti@helsinki.fi
Code: `https://github.com/jopo666/HistoEncoder`

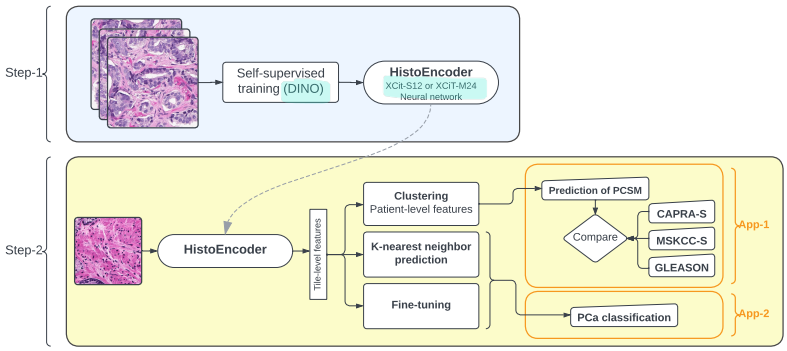


Figure 1: An overview of the HistoEncoder workflow. HistoEncoder utilizes the cross-covariance transformer (XCiT) as the backbone network. The models are pre-trained with 48 million prostate tissue tile images from 1,307 patients in a self-supervised manner (Step 1), and fine-tuned to cancer classification and mortality prediction tasks, or used without fine-tuning via a KNN classifier (Step 2).

Table 1: Datasets used in this study.

Name	Medical centre	Patients	Slides	Type	Tile images*	Reference
HelsinkiProstate	Helsinki University Hospital	1,307	11,226	Biopsy & Organ section	898.4 million	–
Helsinki30†	Helsinki University Hospital	30	465	Organ section	75.2 million	[13]
Helsinki60†	Helsinki University Hospital	60	863	Organ section	146.0 million	[13]
HelsinkiTMA	Helsinki University Hospital	432	1,769	Tissue microarray	0.5 million	–
PESO	Radboud University Medical Center	40	137	Region of interest	13,287	[40, 41]
Karolinska	Karolinska Institutet	2113	5,456	Biopsy	3.2 million	[39]
Radboud	Radboud University Medical Center		5,160	Biopsy	2.6 million	[39]

*Non-overlapping 256 × 256 tiles with >50% tissue; †Part of HelsinkiProstate dataset

[42], which are commonly encountered in digital pathology. Two XCiT backbones are trained on prostate tissue samples, `prostate-s` and `prostate-m`, based on the small and medium-sized XCiT model variants `XCiT-S12` and `XCiT-M24`, respectively.

Many recent models for clinical digital pathology image analysis have been pre-trained with natural images, often using supervised training methods [15–26]. Although it is common to use models pre-trained with supervised training methods, this does not always produce useful embedded features, when compared to models trained with self-supervised methods (see [31] and Supplementary Tab. 2, 3, and 4) . Thus, to make comparisons between models trained with histological images and natural images more fair, all comparisons are against XCiT-S12 and XCiT-M24 models trained on natural images using the self-supervised method DINO [31, 43]. These models are denoted as `natural-s` and `natural-m`, respectively.

2.2 *Workflows leveraging encoder models*

We describe two workflows that leverage the image features from our encoder models. The first is for automatic pre-annotation of large imaging datasets, and the second is for combining histomics data with other data modalities. Command line interfaces and a Python module for running these workflows can be found from `https://github.com/jopo666/HistoEncoder`.

2.2.1 *Automatic annotation of large-scale slide image datasets*

First, all slide images in the dataset are cut into small tile images, for example with `HistoPrep` [44]. Second, features for all tile images are extracted with either `prostate-s` or `prostate-m` encoder models. Third, all extracted features for the tile images are clustered into n clusters with mini-batch K-means. Given that HistoEncoder produces similar embedded features for tile images with similar histological patterns, clusters can be labelled based on visual inspection of the prevailing histological pattern. After automatically pre-annotating all tile images in a dataset with HistoEncoder, a pathologist can visually inspect a subset of the tile images in each cluster. Now, if a given cluster contains only a single histological pattern, such as stroma or epithelium, all tile images assigned to this cluster can be labelled based on the prevailing pattern. If a cluster contains too much variability, the tile images in this cluster can be further clustered and labelled to produce clusters with less variability. The process can then be repeated until enough tile images have been labelled for the task at hand.

2.2.2 *Combining histomics with other data modalities*

Here, we present a workflow to integrate HistoEncoder image features with other data modalities. When additional data modalities, like spatial transcriptomics data [45] or tissue microarray (TMA) spot labels, provide information for smaller tissue re-

gions, feature vectors can be directly generated for these regions using the `prostate-s` or `prostate-m` encoder models.

In contrast, if the data modalities to be integrated contain information for larger tissue regions or whole or multiple slide images, the slide images must be cut into smaller tile images that are compatible with HistoEncoder. A common example is patient-level clinical data, where data from each patient pertains to all tile images from the same patient. In this case, we will first automatically annotate and cluster all tile images following the workflow in section 2.2.1. Each cluster contains tile images with similar histological patterns. Second, we will calculate, for each patient, the proportion of tile images assigned to each cluster ("cluster fractions"). Cluster fractions represent a patient-specific summary of histological patterns observed in the slide images.

To give an example, in a dataset of tissue biopsy slide images and associated clinical data, we can derive patient-level cluster fractions by analyzing all tile images extracted from each biopsy slide. If a specific cluster predominantly contains tile images with high-grade cancer, the fraction of this cluster for a given patient would represent the proportion of high-grade cancer observed in that patient's biopsy slides.

2.3 *Datasets*

All datasets leveraged in this study are summarized in table 1. All slide images are processed with `HistoPrep` [44].

2.3.1 *Training dataset*

Training data, designated as HelsinkiProstate, for the prostate tissue encoders, consists of patients who have undergone prostate biopsies or radical prostatectomy in Helsinki University Hospital between 2013 and 2021. In total, there are 1,307 patients with 5,642 and 5,584 tissue slides of needle biopsies, and tissue sections from radical prostatectomy (table 1). To create the HelsinkiProstate dataset, each slide is cut into 640×640 pixel tile images at magnifications 20x, 10x, and 5x with 20% overlap between neighbouring tile images. This produces several hundred million tile images, which are then preprocessed with `HistoPrep` [44] to remove tiles containing non-tissue areas such as pen markings or other artefacts. Once irrelevant tile images have been filtered out, all remaining tile images are run through a prostate cancer classifier model from [13]. To create a balanced training dataset, all 16 million tile images with a prediction score $\hat{y} > 0.2$, and 32 million randomly sampled tile images with score $\hat{y} \leq 0.2$ were selected to comprise the HelsinkiProstate dataset.

2.3.2 *Evaluation datasets*

HelsinkiTMA dataset contains 1,769 TMA spots from 432 prostate cancer patients who underwent radical prostatectomy between 1983 and 1998 at the Helsinki University Hospital. All slide images are cut into 512×512 with 25% overlap between neighbouring tiles. The patients have a median follow-up time of 19.0 years. All 432 patients have Gleason grade information available, and 238 have complete clinical data for calculating CAPRA-S [46] score and Memorial Sloan Kettering Cancer Center 15-year survival probability [47], denoted as MSKCC-S. Kaplan-Meier survival curves [48] for Gleason grade groups and both clinical nomograms are presented in Supplementary Fig. 7.

Helsinki30 and Helsinki60 datasets [13] contain whole slide images (normal size and whole mounts) from 30 and 60 patients who have undergone radical prostatectomy at the Helsinki University Hospital between the years 2014 and 2021. All slide images in both datasets have been annotated by pathologists, and classified as cancerous or benign. These datasets are also part of the training (HelsinkiProstate) dataset (table 1).

Several publicly available datasets are used in this study. The test set of the PESO dataset [40, 41] contains 137 tile images of $2,500 \times 2,500$ pixels from 37 different slide images annotated as either cancerous or benign. PANDA development dataset [39] contains 10,616 prostate biopsy slides from 2,113 patients from Radboud University Medical Center and Karolinska Institute, which we denote in this study as Radboud and Karolinska.

2.4 *Fine-tuning the encoder models*

To fine-tune the encoders for downstream tasks, the partial fine-tuning setup from [28] is used, with the following modifications. RandAugment [49] is replaced with StrongAugment [13] with 2, 3 or 4 operations with 0.5, 0.3 and 0.2 probability, respectively. Each image is flipped vertically and/or horizontally and a perspective operation is applied with a scale from 0 to 0.1 with a probability of 0.5. For partial fine-tuning [28], only the last 0.5, 2, 4 or 8 XCiT-S12 model blocks are fine-tuned, where 0.5 denotes fine-tuning the last fully connected layer of the last model block. There are 1.182, 3.550, 7.120 and 14.260 million fine-tuned parameters for 0.5, 2, 4, and 8 model blocks, and 25.868 million for the full XCiT-S12 model. For fine-tuning the encoders without training, a K-nearest neighbour (KNN) classifier is used with $k = 20$. The KNN-classifier is fit with the features extracted from the training dataset, and label predictions correspond to the proportion of nearest neighbours with a positive label. Each experiment is repeated five times, and the mean and standard deviation of the area under the receiver operating curve (AUROC) are reported.

For the PESO dataset, randomly resized crops with scale [0.01, 0.2] are randomly sampled from the 2,500x2,500 pixel region of interest images during fine-tuning, and an epoch is defined as 51,200 training samples. For Karolinska and Radboud datasets, training images are randomly sampled from 384x384 pixel tile images with 20% overlaps, and an epoch is defined as 262,144 training samples, or the maximum amount of tile images included in model training. When limiting the number of training images from the PESO dataset, region of interest images from 1, 2, 4, 8, 16 or 32 histological slides are used for training. When limiting the number of training data in the Karolinska and Radboud datasets, only 4,096, 8,192, 16,384, 32,768, 65,536, 131,072, 262,144 or 524,288 randomly selected tile images, or all 972,800 and 782,336 images in the Karolinska and Radboud datasets, are used for training. In both training data limitation experiments, the last two encoder blocks are fine-tuned.

For augmentations and transformations, `strong-augment` (0.1.0) and `albumentations` (1.3.0) [50] Python packages were used.

2.4.1 *Prostate cancer-specific death survival models*

To create survival models using HistoEncoder features, we first cluster the HelsinkiTMA dataset and obtain histological cluster memberships for each patient as described in section 2.2.2 with the `prostate-m` encoder and number of clusters set to 32.

Penalized Cox proportional hazards models are then used to predict prostate cancer-specific death, with a penalizer of 0.001 and an $L1$-ratio of 0.5. From the 32 patient-level clusters, six clusters are selected based on a parameter importance analysis. All models are then trained on the patient Gleason grade, CAPRA-S or MSKCC-S, with or without HistoEncoder cluster features. Each model is trained 1,000 times, where 25% of the samples are set aside as a test set using stratified random splits. Model performance is evaluated with a concordance score, time-dependent AUC score between 1 and 23 years, and a net benefit [51] curve for predicting prostate cancer-specific death at 15 years.

3 RESULTS

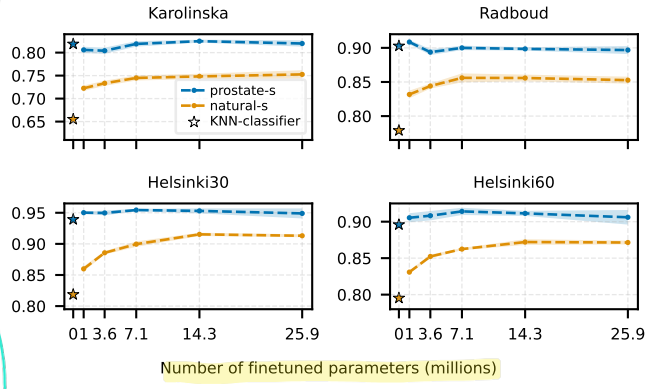
3.1 *Fine-tuning a prostate cancer classifier*

In this section, we report the performance of both `prostate-s` and `natural-s` encoder models in classifying prostate cancer in tissue images. The models are fine-tuned using either the PESO, or Karolinska and Radboud datasets. Both encoder models are fine-tuned while limiting either the number of fine-tuned parameters or the amount of training data.

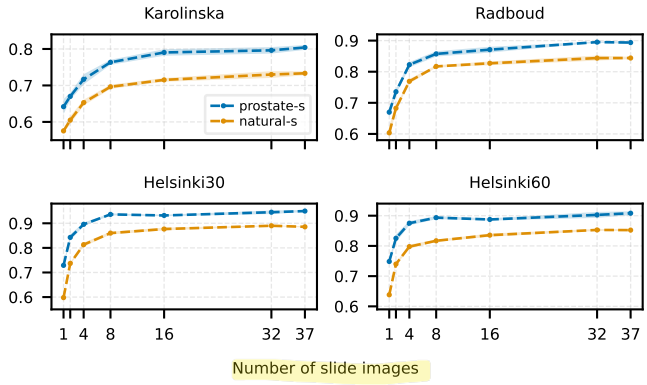
While limiting the number of fine-tuned parameters (fig. 2a and Supplementary Fig. 6a), `prostate-s` based models significantly outperform `natural-s` based models in all evaluation datasets, regardless how many parameters are fine-tuned. `Prostate-s` based models also saturate quickly and only require minimal fine-tuning to achieve the best performance. `Prostate-s` based models seem to overfit to the training datasets (Supplementary Fig. 6a), and would likely benefit from higher regularisation.

While limiting the number of training images (fig. 2b and Supplementary Fig. 6b) , `prostate-s` based models significantly outperform `natural-s` based models in all evaluation datasets. `Prostate-s` based models achieve comparable or better performance with only 16 regions of interest crops from four distinct histological slides, when compared to `natural-s` based models trained on the full PESO dataset (fig. 2b). In particular, `prostate-s` based models achieve significantly better performance with only 1,024 tile images when compared to `natural-s` based models trained with all 972,800 and 782,336 tile images in the Karolinska and Radboud datasets (Supplementary Fig. 6b),respectively. `Prostate-s` based models also saturate quickly with diminishing returns after 8,192 tile images.

Remarkably, a KNN-classifier fitted with features extracted from `prostate-s` outperforms a fully fine-tuned `natural-s` model on all evaluation datasets (fig. 2a and Supplementary Fig. 6).`Prostate-s` outperform models trained in a supervised manner with ImageNet data (Supplementary Tab. 2,3, and 4).



(a) Limiting the number of fine-tuned encoder blocks.



(b) Limiting the amount of data used to train the models.

Figure 2: AUROC scores for the prostate cancer classifiers fine-tuned from pre-trained `prostate-s` and `natural-s` encoder models using the PESO dataset. `Prostate-s` based models achieve higher AUROC scores than `natural-s` based models with less fine-tuned parameters (*a*) and training data (*b*). A simple KNN classifier requiring no training significantly outperforms a fully fine-tuned `natural-s` encoder model on all evaluation datasets.

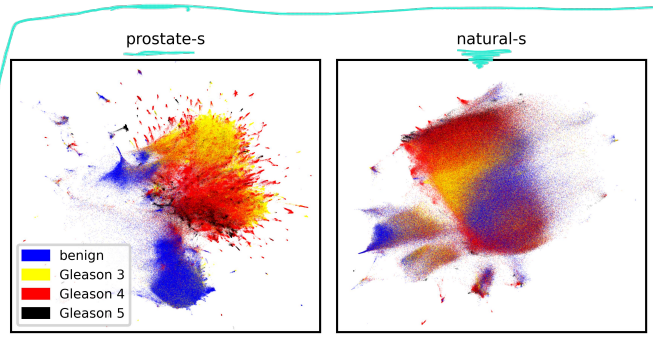
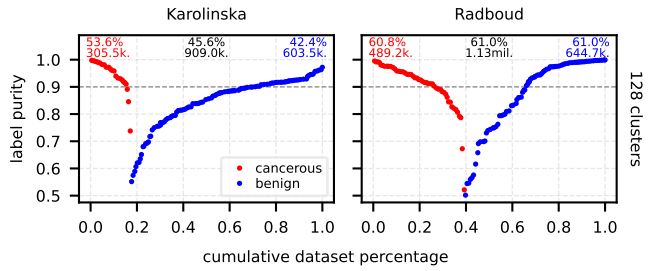
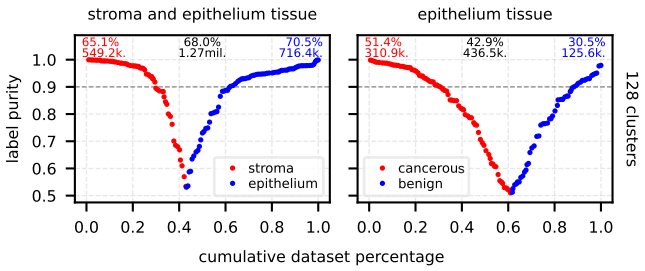


Figure 3: UMAP representation of the features extracted from the epithelium tissue in the Radboud dataset. Benign and cancerous epithelium, as well as different Gleason grades, form clear clusters in the features extracted with the `prostate-s` encoder, while features extracted with the `natural-s` model are more mixed.



(a) Cluster label purities the Karolinska and Radboud datasets.



(b) More detailed label purities for the Radboud dataset.

Figure 4: Cluster label purities for different labels in the Karolinska and Radboud datasets. A significant proportion of tile images in these datasets are contained in clusters with greater than 90% label purity for cancerous vs benign tissue (a), stroma vs. epithelium tissue (b left), and cancerous vs. benign epithelium tissue (b right).

3.2 *Identifying histological patterns as HistoEncoder feature clusters*

Tissue image features derived from the `prostate-s` model create clusters that clearly separate benign from cancerous epithelium (fig. 3). This representation also allows differentiating cancer grades (Gleason). Conversely, features extracted with the natural-s model show a more diffuse pattern, making it harder to distinguish tissue types.

To quantify whether similar features are extracted for tile images with similar histological patterns, the `prostate-m` encoder is used to automatically label the Karolinska and Radboud datasets with the annotation workflow (section 2.2.1). After assigning each tile image to a cluster, the proportion of different labels in each feature cluster can be assessed. In the Karolinska dataset, 45.6% of all tile images, 53.3% of cancerous epithelium, and 42.4% of benign epithelium and stroma are contained in clusters with a label purity above 90% (fig. 4). In the Radboud dataset, a clear majority of images belong to high-purity clusters (68.8% all tiles, 74.7% cancerous epithelium, and 63.9% benign epithelium and stroma). In total, 2.1 million 384×384 tile images in the PANDA dataset could be classified into cancerous and benign tissue with over 90% accuracy after visually inspecting only 256 tile image clusters.

Annotations in the Radboud dataset are more granular than those in the Karolinska dataset, distinguishing between stroma and benign epithelium. In the Radboud dataset, 68.0% of all tile images, 65.1% of stroma, and 70.5% of epithelium are contained in clusters with a label purity above 90% (fig. 4b). For the harder task of classifying epithelium tissue as cancerous or benign, 42.9% of all tile images, 51.4% of cancerous, and 30.5% of benign epithelium are contained in clusters with a label purity above 90%. Varying the number of annotated clusters, we observe that as few as eight clusters are enough to find clusters with label purity over 90% (Supplementary Fig. 8).

3.3 *Predicing prostate cancer-specific mortality*

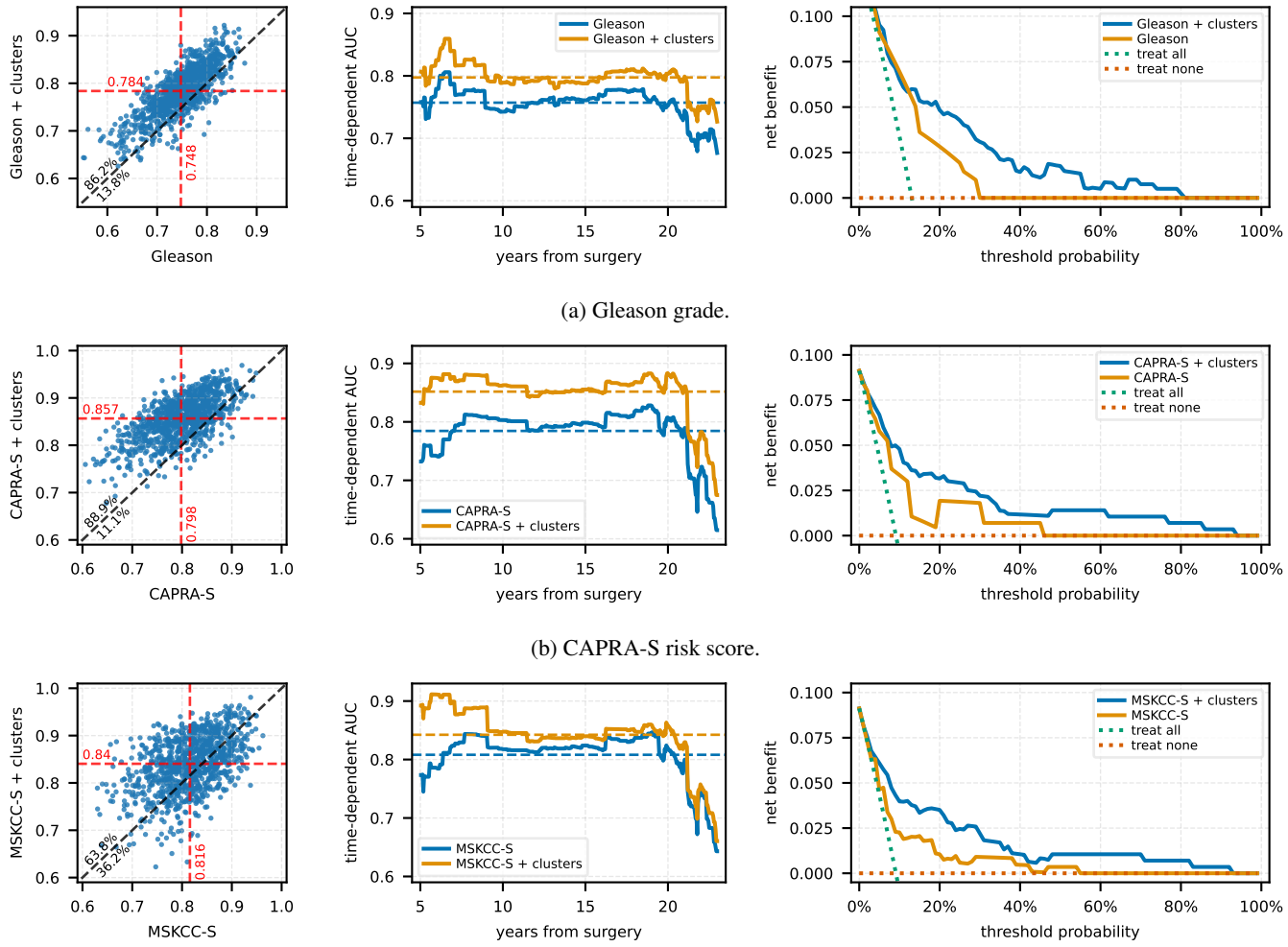
We next combine information extracted with HistoEncoder from histological slide images with clinical data to build multimodal survival models predicting prostate cancer-specific death. The `prostate-m` encoder is used to collect patient-level feature cluster frequencies ($n = 32$ clusters; section 2.2.2) for the tissue microarray spots in the HelsinkiTMA dataset. These cluster frequencies represent the histological pattern distribution for each patient and can be simply concatenated with clinical data. Next, six feature clusters are selected and Cox proportional hazards models are fitted (see section 2.4.1). A visual assessment of the top six feature clusters suggests more atypia in the cellular morphologies in clusters associated with worse prognoses (Supplementary Fig. 9).Certain subtypes and growth patterns, such as cribriform gland architecture and mucinous pattern are recurrent in the clusters with the highest HRs. Extracellular matrix formation, individual cell clusters and lymphocyte composition seem to vary even between the top six clusters.

In 1,000 random stratified splits, survival models augmented with the patient-level cluster frequencies achieve higher concordance scores than the baseline models in 84.9%, 89.2%, and 67.4% of the splits with Gleason grade, CAPRA-S and MSKCC-S, respectively (fig. 5, left). HistoEncoder-augmented models also achieve higher mean time-dependent AUC scores over a 1 to 23-year period (fig. 5, center). Finally, the augmented models yield a higher net benefit predicting prostate cancer death at 15 years over a wide threshold probability (fig. 5, right).

4 DISCUSSION

In this work, we introduce HistoEncoder, a foundation model for prostate cancer digital histopathology. Foundation models hold significant promise to contribute to precision cancer medicine. While pre-training a foundation model takes considerable computational resources, the model can then be fine-tuned to a multitude of specific tasks [52–54]. Importantly, fine-tuning often requires orders of magnitude less data and resources than pre-training [52, 55, 56].

We pre-trained HistoEncoder with 48 million tile images extracted from prostate tissue images and showed how the model could distinguish malignant and benign tissues, Gleason grades, as well as stromal and epithelial tissues without any annotated information on tissue types being available during training. Previously, foundation models for digital histopathology have been typically pre-trained with natural images [57–59]. In this study, we demonstrated how pre-training with domain-specific images yielded substantially better performance than using natural images for prostate histopathology tasks. In line with our findings, recent foundation models trained on tissue images have had ex-



(a) Gleason grade.

(b) CAPRA-S risk score.

(c) MSKCC-S prediction of being alive at 15 years.

Figure 5: Head-to-head concordance comparisons between 1,000 stratified random splits (*left*), time-dependent AUC scores (*centre*) and net benefit curves for prediction of prostate cancer-specific death at 15 years (*right*). Including patient-level feature cluster percentages improves concordance, time-dependent AUC score and provides a higher net benefit over a wide threshold probability for the Gleason grade (*a*), CAPRA-S (*b*) and MSKCC-S (*c*) nomograms.

cellent performance in digital histopathology downstream tasks such as cancer detection and survival prediction [60, 61]. These models have however not challenged existing cancer-specific clinical prediction models to the extent that we address prostate cancer.

Here, we demonstrated extending and applying HistoEncoder in two clinically relevant tasks. First, we created a classifier which achieved high performance in predicting the presence of prostate cancer. We observed significantly better prediction accuracy, and compute and data efficiency with HistoEncoder models trained with tissue images compared to identical models trained with natural images, also in unsupervised cancer grading task. Second, we applied HistoEncoder in the integration of tissue imaging and clinical data to compare with commonly used nomograms for prostate cancer-specific mortality prediction. This approach resulted in a model which is able to consistently and robustly outperform the commonly applied risk classification systems in clinical use (*i.e.*, Gleason grading, CAPRA-S and MSKCC-S systems). Notably, only a handful of annotated cases and a personal computer were sufficient to fine-tune the HistoEncoder models for this purpose.

Previously, Pinckaers *et al.* developed a CNN model trained on biochemical recurrence (BCR) outcomes in a case-control setting [62]. A biomarker derived from this ImageNet-pretrained ResNet50-D model demonstrated significant predictivity in a multivariable model including preoperative PSA, ISUP Gleason grade, pathological stage and surgical margin status. In an external validation cohort, however, the conventional ISUP Gleason grade outperformed the biomarker. This highlights the challenge of developing generalizable models for real-world clinical applications. Despite extensive data augmentation, models pre-trained with natural images may struggle to learn features that would consistently surpass Gleason's grading across diverse patient cohorts.

In contrast, our results suggest that HistoEncoder learns visualisable and comprehensible features (Supplementary Fig. 9) beyond Gleason-grade patterns that are useful in clinical tasks such as predicting prostate cancer mortality. There is a plethora of previous work showing that explainable features beyond conventional histopathological classification systems are prognostic or predictive of cancer-patient outcome [63–65]. However, reports visualizing comprehensible AI-derived features for a human expert and clinically meaningful multimodal approaches are scarce. Here we showed that HistoEncoder was able to learn such image features without expert guidance or labeled data.

One of the weaknesses of our study is the lack of an external validation cohort in the survival analysis, and therefore it remains to be validated whether these features would be present also in other cohorts. Future efforts should hence evaluate whether HistoEncoder is able to extract coherent, predictive features across multiple cohorts, as this could yield a clinically applicable approach to stratify patients to subgroups based on survival probability beyond Gleason grades. Ultimately, complex machine learning models need to provide additional clinical value beyond current classification systems such as Gleason scoring, preferably with explainable features, in order to be adopted in everyday clinical practice.

4.1 Extending HistoEncoder

We provide HistoEncoder as an easy-to-use Python package containing both the models pre-trained with prostate tissue images, as well as a standalone software tool to extract features and cluster tissue tile images. HistoEncoder can thus be extended to specific tasks involving any histopathological images.

As proof of concept, we were able to address cancer detection and mortality prediction with HistoEncoder using limited data and computing resources from a single laptop computer. This exemplifies the utility of such foundation models in lowering the barrier to creating clinical software tools. The `prostate-s` and `prostate-m` encoder models were able to compress information from large tissue areas into a feature vector, which was subsequently combined with other data modalities. If the other data modalities, such as spatial transcriptomics data [66] or TMA spot labels, contain information for small tissue regions, feature vectors can be directly extracted for these regions with HistoEncoder models.

Taken together, our results demonstrate how HistoEncoder can provide clinically relevant insights from tissue images. We also highlight the importance of training data representing domain-specific data. Foundation models such as HistoEncoder allow computational methods to be quickly developed for precision cancer medicine tasks with small amounts of domain-specific data. In this study, we did not evaluate whether HistoEncoder would be useful as a foundation for models involving other cancers than prostate cancer, or in a pan-cancer setting. An exciting direction would be to explore the utility of foundation models in the analysis of multiple clinically relevant modalities and as part of multimodal predictor and interpreter models [67–69]. It is crucial to evaluate HistoEncoder against tissue-agnostic foundation models pre-trained specifically on histopathology images [60, 61, 70]. Future multimodal models are likely to allow leveraging tissue imaging on a large scale in conjunction with other clinical and molecular profiling data, leading to improved tools for cancer diagnosis, prognosis and treatment.

ACKNOWLEDGEMENTS

This work was supported by Cancer Foundation Finland [304667, 191118], Jane and Aatos Erkko Foundation [290520], Research Council of Finland [322675] and Hospital District of Helsinki and Uusimaa [TYH2018214, TYH2018222, TYH2019235, TYH2019249]. The authors also wish to acknowledge FIMM Digital Microscopy and Molecular Pathology Unit supported by HiLIFE and Biocenter Finland for imaging services, and CSC – IT Center for Science, Finland for generous computational resources on the LUMI supercomputer (LUMI Extreme Scale project).

ETHICS STATEMENT

Ethical approvals for the use of human tissue material and clinicopathologic data were obtained from the Institutional Ethics Committee of Hospital District of Helsinki and Uusimaa (§70/16.5.2018; HUS/419/2018) and by the National Supervisory Authority for Welfare and Health (VALVIRA, D:no V/38176/2018). According to the national and European Union

legislation on noninterventional medical research, the study was conducted without informed individual patient consent by permission of the Hospital District of Helsinki and Uusimaa (§105/21.12.2018; HUS/419/2018). The experiments conformed to the principles set out in the WMA Declaration of Helsinki and the Department of Health and Human Services Belmont Report.

DATA AVAILABILITY

The original human subject and sample-related data is available upon request, provided the institutional ethical approval and research permit allow data sharing according to the particular request. Restrictions may apply to the availability of the internal Helsinki University Hospital datasets, which cannot be made publicly available due to general data protection regulations, national legislation and institutional guidelines. Publicly available datasets can be accessed through their respective publications. For data inquiries, please contact `tuomas.mirtti@helsinki.fi`.

DECLARATION OF COMPETING INTERESTS

The authors have no interests to declare.

REFERENCES

[1] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[2] Jeroen Van der Laak, Geert Litjens, and Francesco Ciompi. Deep learning in histopathology: the path to the clinic. *Nature medicine*, 27(5):775–784, 2021.

[3] Amelie Echle, Niklas Timon Rindtorff, Titus Josef Brinker, Tom Luedde, Alexander Thomas Pearson, and Jakob Nikolas Kather. Deep learning in cancer pathology: a new generation of clinical biomarkers. *British journal of cancer*, 124(4):686–696, 2021.

[4] Kaustav Bera, Kurt A Schalper, David L Rimm, Vamsidhar Velcheti, and Anant Madabhushi. Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology. *Nature reviews Clinical oncology*, 16(11):703–715, 2019.

[5] Andre Esteva, Jean Feng, Douwe van der Wal, Shih-Cheng Huang, Jeffry P. Simko, Sandy DeVries, Emmalyn Chen, Edward M. Schaeffer, Todd M. Morgan, Yilun Sun, Amirata Ghorbani, Nikhil Naik, Dhruv Nathawani, Richard Socher, Jeff M. Michalski, Mack Roach, Thomas M. Pisansky, Jedidiah M. Monson, Farah Naz, James Wallace, Michelle J. Ferguson, Jean-Paul Bahary, James Zou, Matthew Lungren, Serena Yeung, Ashley E. Ross, Michael Kucharczyk, Luis Souhami, Leslie Ballas, Christopher A. Peters, Sandy Liu, Alexander G. Balogh, Pamela D. Randolph-Jackson, David L. Schwartz, Michael R. Girvigian, Naoyuki G. Saito, Adam Raben, Rachel A. Rabinovitch, Khalil Katato, Howard M. Sandler, Phuoc T. Tran, Daniel E. Spratt, Stephanie Pugh, Felix Y. Feng, Osama Mohamad, and NRG Prostate Cancer AI Consortium. Prostate cancer therapy personalization via multimodal deep learning on randomized phase iii clinical trials. *npj Digital Medicine*, 5(1):71, 2022.

[6] Daniel E. Spratt, Siyi Tang, Yilun Sun, and et al. Artificial intelligence predictive model for hormone therapy use in prostate cancer. *PREPRINT (Version 1) available at Research Square*, 2023. 21 April 2023.

[7] Kunal Nagpal, Davis Foote, Yun Liu, Po-Hsuan Cameron Chen, Ellery Wulczyn, Fraser Tan, Niels Olson, Jenny L Smith, Arash Mohtashamian, James H Wren, et al. Development and validation of a deep learning algorithm for improving gleason scoring of prostate cancer. *NPJ digital medicine*, 2(1):1–10, 2019.

[8] Thomas de Bel, Meyke Hermsen, Jesper Kers, Jeroen van der Laak, and Geert Litjens. Stain-transforming cycle-consistent generative adversarial networks for improved segmentation of renal histopathology. In *International Conference on Medical Imaging with Deep Learning–Full Paper Track*, 2018.

[9] Yun Liu, Timo Kohlberger, Mohammad Norouzi, George E Dahl, Jenny L Smith, Arash Mohtashamian, Niels Olson, Lily H Peng, Jason D Hipp, and Martin C Stumpe. Artificial intelligence-based breast cancer nodal metastasis detection: Insights into the black box for pathologists. *Archives of pathology & laboratory medicine*, 143(7):859–868, 2019.

[10] Gabriele Campanella, Arjun R Rajanna, Lorraine Corsale, Peter J Schüffler, Yukako Yagi, and Thomas J Fuchs. Towards machine learned quality control: A benchmark for sharpness quantification in digital pathology. *Computerized Medical Imaging and Graphics*, 65:142–151, 2018.

[11] Julia K Winkler, Christine Fink, Ferdinand Toberer, Alexander Enk, Teresa Deinlein, Rainer Hofmann-Wellenhof, Luc Thomas, Aimilios Lallas, Andreas Blum, Wilhelm Stolz, et al. Association between surgical skin markings in dermoscopic images and diagnostic performance of a deep learning convolutional neural network for melanoma recognition. *JAMA dermatology*, 155(10):1135–1141, 2019.

[12] Adarsh Subbaswamy and Suchi Saria. From development to deployment: dataset shift, causality, and shift-stable models in health ai. *Biostatistics*, 21(2):345–352, 2020.

[13] Joona Pohjonen, Carolin Stürenberg, Atte Föhr, Reija Randen-Brady, Lassi Luomala, Jouni Lohi, Esa Pitkänen, Antti Rannikko, and Tuomas Mirtti. Augment like there's no tomorrow: Consistently performing neural networks for medical imaging. *arXiv preprint arXiv:2206.15274*, 2022.

[14] Dan Hendrycks, Kimin Lee, and Mantas Mazeika. Using pre-training can improve model robustness and uncertainty. In *International Conference on Machine Learning*, pages 2712–2721. PMLR, 2019.

[15] Chetan L. Srinidhi, Ozan Ciga, and Anne L. Martel. Deep neural network models for computational histopathology: A survey. *Medical Image Analysis*, 67:101813, 2021.

[16] Yuri Tolkach, Tilmann Dohmgörgen, Marieta Toma, and Glen Kristiansen. High-accuracy prostate cancer pathology using deep learning. *Nature Machine Intelligence*, 2(7):411–418, 2020.

[17] Marcel Gehrung, Mireia Crispin-Ortuzar, Adam G Berman, Maria O'Donovan, Rebecca C Fitzgerald, and Florian

Markowetz. Triage-driven diagnosis of barrett's esophagus for early detection of esophageal adenocarcinoma using deep learning. *Nature medicine*, 27(5):833–841, 2021.

[18] William Lotter, Abdul Rahman Diab, Bryan Haslam, Jiye G Kim, Giorgia Grisot, Eric Wu, Kevin Wu, Jorge Onieva Onieva, Yun Boyer, Jerrold L Boxerman, et al. Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. *Nature Medicine*, 27(2):244–249, 2021.

[19] Ming Y Lu, Tiffany Y Chen, Drew FK Williamson, Melissa Zhao, Maha Shady, Jana Lipkova, and Faisal Mahmood. Ai-based pathology predicts origins for cancers of unknown primary. *Nature*, 594(7861):106–110, 2021.

[20] Jean Ogier du Terrail, Armand Leopold, Clément Joly, Constance Béguier, Mathieu Andreux, Charles Maussion, Benoît Schmauch, Eric W Tramel, Etienne Bendjebbar, Mikhail Zaslavskiy, et al. Federated learning for predicting histological response to neoadjuvant chemotherapy in triple-negative breast cancer. *Nature medicine*, 29(1):135–146, 2023.

[21] Chi-Long Chen, Chi-Chung Chen, Wei-Hsiang Yu, Szu-Hua Chen, Yu-Chan Chang, Tai-I Hsu, Michael Hsiao, Chao-Yuan Yeh, and Cheng-Yu Chen. An annotation-free whole-slide training approach to pathological classification of lung cancer types using deep learning. *Nature communications*, 12(1):1193, 2021.

[22] Yongju Lee, Jeong Hwan Park, Sohee Oh, Kyoungseob Shin, Jiyu Sun, Minsun Jung, Cheol Lee, Hyojin Kim, Jin-Haeng Chung, Kyung Chul Moon, et al. Derivation of prognostic contextual histopathological features from whole-slide images of tumours via graph deep learning. *Nature Biomedical Engineering*, pages 1–15, 2022.

[23] Artem Shmatko, Narmin Ghaffari Laleh, Moritz Gerstung, and Jakob Nikolas Kather. Artificial intelligence in histopathology: enhancing cancer research and clinical oncology. *Nature cancer*, 3(9):1026–1038, 2022.

[24] Jana Lipkova, Tiffany Y Chen, Ming Y Lu, Richard J Chen, Maha Shady, Mane Williams, Jingwen Wang, Zahra Noor, Richard N Mitchell, Mehmet Turan, et al. Deep learning-enabled assessment of cardiac allograft rejection from endomyocardial biopsies. *Nature medicine*, 28(3):575–582, 2022.

[25] Chengkuan Chen, Ming Y Lu, Drew FK Williamson, Tiffany Y Chen, Andrew J Schaumberg, and Faisal Mahmood. Fast and scalable search of whole-slide images via self-supervised deep learning. *Nature Biomedical Engineering*, 6(12):1420–1434, 2022.

[26] Ming Y Lu, Drew FK Williamson, Tiffany Y Chen, Richard J Chen, Matteo Barbieri, and Faisal Mahmood. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature biomedical engineering*, 5(6):555–570, 2021.

[27] Maithra Raghu, Chiyuan Zhang, Jon Kleinberg, and Samy Bengio. Transfusion: Understanding transfer learning for medical imaging. *Advances in neural information processing systems*, 32, 2019.

[28] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners, 2021.

[29] Zhenda Xie, Zheng Zhang, Yue Cao, Yutong Lin, Jianmin Bao, Zhuliang Yao, Qi Dai, and Han Hu. Simmim: A simple framework for masked image modeling, 2022.

[30] Jinghao Zhou, Chen Wei, Huiyu Wang, Wei Shen, Cihang Xie, Alan Yuille, and Tao Kong. ibot: Image bert pre-training with online tokenizer, 2022.

[31] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers, 2021.

[32] Maxime Oquab, Timothée Darcet, Theo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision, 2023.

[33] Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. On the opportunities and risks of foundation models, 2022.

[34] Mathilde Caron, Piotr Bojanowski, Julien Mairal, and Armand Joulin. Unsupervised pre-training of image features on non-curated data, 2019.

[35] Priya Goyal, Mathilde Caron, Benjamin Lefaudeux, Min Xu, Pengchao Wang, Vivek Pai, Mannat Singh, Vitaliy Liptchinsky, Ishan Misra, Armand Joulin, and Piotr Bojanowski. Self-supervised pretraining of visual features in the wild, 2021.

[36] Priya Goyal, Quentin Duval, Isaac Seessel, Mathilde Caron, Ishan Misra, Levent Sagun, Armand Joulin, and Piotr Bojanowski. Vision models are more robust and fair when pretrained on uncurated images without supervision, 2022.

[37] Jiajun Li, Tiancheng Lin, and Yi Xu. Sslp: Spatial guided self-supervised learning on pathological images. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II 24*, pages 3–12. Springer, 2021.

[38] Chetan L Srinidhi and Anne L Martel. Improving self-supervised learning with hardness-aware dynamic curriculum learning: An application to digital pathology. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 562–571, 2021.

[39] Wouter Bulten, Kimmo Kartasalo, Po-Hsuan Cameron Chen, Peter Ström, Hans Pinckaers, Kunal Nagpal, Yuannan Cai, David F Steiner, Hester van Boven, Robert Vink, et al. Artificial intelligence for diagnosis and gleason grading of prostate cancer: the panda challenge. *Nature medicine*, pages 1–10, 2022.

[40] Wouter Bulten, Péter Bándi, Jeffrey Hoven, Rob van de Loo, Johannes Lotz, Nick Weiss, Jeroen van der Laak, Bram van Ginneken, Christina Hulsbergen-van de Kaa, and Geert Litjens. Peso: Prostate epithelium segmentation

on h&e-stained prostatectomy whole slide images, 2018. Accessed: 05.03.2021.

[41] Jonathan I Epstein, Lars Egevad, Mahul B Amin, Brett Delahunt, John R Srigley, and Peter A Humphrey. The 2014 international society of urological pathology (isup) consensus conference on gleason grading of prostatic carcinoma. *The American journal of surgical pathology*, 40(2):244–252, 2016.

[42] Alaaeldin El-Nouby, Hugo Touvron, Mathilde Caron, Piotr Bojanowski, Matthijs Douze, Armand Joulin, Ivan Laptev, Natalia Neverova, Gabriel Synnaeve, Jakob Verbeek, and Hervé Jegou. Xcit: Cross-covariance image transformers, 2021.

[43] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[44] Joona Pohjonen and Valeria Ariotta. Histoprep: Preprocessing large medical images for machine learning made easy! https://github.com/jopo666/HistoPrep, 2022.

[45] Patrik L Ståhl, Fredrik Salmén, Sanja Vickovic, Anna Lundmark, José Fernández Navarro, Jens Magnusson, Stefania Giacomello, Michaela Asp, Jakub O Westholm, Mikael Huss, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 353(6294):78–82, 2016.

[46] Matthew R Cooperberg, Joan F Hilton, and Peter R Carroll. The capra-s score: a straightforward tool for improved prediction of outcomes after radical prostatectomy. *Cancer*, 117(22):5039–5046, 2011.

[47] Dynamic prostate cancer nomogram: Coefficients. `https://www.mskcc.org/nomograms/prostate/post_op/coefficients`. Accessed: 2023-05-02.

[48] Edward L Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.

[49] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 702–703, 2020.

[50] A Buslaev, A Parinov, E Khvedchenya, V. I. Iglovikov, and A. A. Kalinin. Albumentations: fast and flexible image augmentations. *ArXiv e-prints*, 2018.

[51] Andrew J Vickers, Ben Van Calster, and Ewout W Steyerberg. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *bmj*, 352, 2016.

[52] Fa Wang, Z. Zhuang, F. Gao, and et al. TMO-Net: an explainable pretrained multi-omics model for multi-task learning in oncology. *Genome Biology*, 25:149, 2024.

[53] D. Truhn, J. N. Eckardt, D. Ferber, et al. Large language models and multimodal foundation models for precision oncology. *npj Precision Oncology*, 8:72, 2024.

[54] M. A. Wójcik. Foundation models in healthcare: Opportunities, biases and regulatory prospects in europe. In Andrea Kő, Enrico Francesconi, Gabriele Kotsis, A Min Tjoa, and Ismail Khalil, editors, *Electronic Government and the Information Systems Perspective. EGOVIS 2022*, Lecture Notes in Computer Science, pages 32–46, Cham, 2022. Springer International Publishing.

[55] Wasif Khan, Seowung Leem, Kyle B See, Joshua K Wong, Shaoting Zhang, and Ruogu Fang. A comprehensive survey of foundation models in medicine. *arXiv preprint arXiv:2406.10729*, 2024.

[56] Hee E Kim, Alejandro Cosa-Linan, Nandhini Santhanam, Mahboubeh Jannesari, Mate E Maros, and Thomas Ganslandt. Transfer learning for medical image classification: a literature review. *BMC medical imaging*, 22(1):69, 2022.

[57] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[58] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[59] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021.

[60] Eugene Vorontsov, Alican Bozkurt, Adam Casson, George Shaikovski, Michal Zelechowski, Kristen Severson, Eric Zimmermann, James Hall, Neil Tenenholtz, Nicolo Fusi, et al. A foundation model for clinical-grade computational pathology and rare cancers detection. *Nature medicine*, pages 1–12, 2024.

[61] Xiyue Wang, Junhan Zhao, Eliana Marostica, Wei Yuan, Jietian Jin, Jiayu Zhang, Ruijiang Li, Hongping Tang, Kanran Wang, Yu Li, et al. A pathology foundation model for cancer diagnosis and prognosis prediction. *Nature*, pages 1–9, 2024.

[62] Hans Pinckaers, Jolique van Ipenburg, Jonathan Melamed, Angelo De Marzo, Elizabeth A Platz, Bram van Ginneken, Jeroen van der Laak, and Geert Litjens. Predicting biochemical recurrence of prostate cancer with artificial intelligence. *Communications Medicine*, 2(1):64, 2022.

[63] Amelie Echle, Niklas Timon Rindtorff, Titus Josef Brinker, Tom Luedde, Alexander Thomas Pearson, and Jakob Nikolas Kather. Deep learning in cancer pathology: a new generation of clinical biomarkers. *British journal of cancer*, 124(4):686–696, 2021.

[64] Sacheth Chandramouli, Patrick Leo, George Lee, Robin Elliott, Christine Davis, Guangjing Zhu, Pingfu Fu, Jonathan I Epstein, Robert Veltri, and Anant Madabhushi. Computer extracted features from initial h&e tissue biopsies predict disease progression for prostate cancer patients on active surveillance. *Cancers*, 12(9):2708, 2020.

[65] Michaela Unger and Jakob Nikolas Kather. A systematic analysis of deep learning in genomics and histopathology for precision oncology. *BMC Medical Genomics*, 17(1):48, 2024.

[66] Patrik L Ståhl, Fredrik Salmén, Sanja Vickovic, Anna Lundmark, José Fernández Navarro, Jens Magnusson, Stefania Giacomello, Michaela Asp, Jakub O Westholm, Mikael Huss, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 353(6294):78–82, 2016.

[67] Hong-Yu Zhou, Yizhou Yu, Chengdi Wang, Shu Zhang, Yuanxu Gao, Jia Pan, Jun Shao, Guangming Lu, Kang Zhang, and Weimin Li. A transformer-based representation-learning model with unified processing of multimodal input for clinical diagnostics. *Nature biomedical engineering*, 7(6):743–755, 2023.

[68] Andrew H Song, Guillaume Jaume, Drew FK Williamson, Ming Y Lu, Anurag Vaidya, Tiffany R Miller, and Faisal Mahmood. Artificial intelligence for digital and computational pathology. *Nature Reviews Bioengineering*, 1(12):930–949, 2023.

[69] Michael Moor, Oishi Banerjee, Zahra Shakeri Hossein Abad, Harlan M Krumholz, Jure Leskovec, Eric J Topol, and Pranav Rajpurkar. Foundation models for generalist medical artificial intelligence. *Nature*, 616(7956):259–265, 2023.

[70] Richard J Chen, Tong Ding, Ming Y Lu, Drew FK Williamson, Guillaume Jaume, Andrew H Song, Bowen Chen, Andrew Zhang, Daniel Shao, Muhammad Shaban, et al. Towards a general-purpose foundation model for computational pathology. *Nature Medicine*, 30(3):850–862, 2024.

Table S1: AUROC scores for a K-nearest neighbor prostate cancer classifier, fitted with features extracted from the PESO dataset.

Model	Parameters	Training method	Training data	Karolinska	Radboud	Helsinki30	Helsinki60
XCiT-S12	25.9	self-supervised	HelsinkiProstate	0.819	**0.903**	0.939	0.896
XCiT-M24	83.9	self-supervised	HelsinkiProstate	**0.824**	0.898	**0.943**	**0.901**
XCiT-S12	25.9	self-supervised	ImageNet	0.655	0.779	0.819	0.795
XCiT-M24	83.9	self-supervised	ImageNet	0.683	0.790	0.816	0.806
EfficientNet-B0	4.0	supervised	ImageNet	0.668	0.774	0.550	0.552
EfficientNet-B1	6.5	supervised	ImageNet	0.666	0.753	0.545	0.54
EfficientNet-B2	7.7	supervised	ImageNet	0.671	0.767	0.558	0.563
EfficientNet-B3	10.7	supervised	ImageNet	0.661	0.740	0.549	0.548
EfficientNet-B4	17.5	supervised	ImageNet	0.652	0.747	0.547	0.553
ResNet-18	11.2	supervised	ImageNet	0.682	0.762	0.547	0.553
ResNet-50	23.5	supervised	ImageNet	0.684	0.746	0.525	0.521
ResNet-101	42.5	supervised	ImageNet	0.618	0.676	0.531	0.531
XCiT-S12	25.9	supervised	ImageNet	0.645	0.690	0.707	0.714
XCiT-M24	83.9	supervised	ImageNet	0.647	0.720	0.752	0.739

Table S2: AUROC scores for a K-nearest neighbor prostate cancer classifier, fitted with features extracted from the Karolinska dataset.

Model	Parameters	Training method	Training data	PESO	Radboud	Helsinki30	Helsinki60
XCiT-S12	25.9	self-supervised	HelsinkiProstate	0.936	0.916	**0.949**	**0.908**
XCiT-M24	83.9	self-supervised	HelsinkiProstate	**0.940**	**0.918**	0.945	0.902
XCiT-S12	25.9	self-supervised	ImageNet	0.857	0.811	0.821	0.808
XCiT-M24	83.9	self-supervised	ImageNet	0.837	0.790	0.796	0.790
EfficientNet-B0	4.0	supervised	ImageNet	0.685	0.778	0.526	0.544
EfficientNet-B1	6.5	supervised	ImageNet	0.707	0.778	0.529	0.551
EfficientNet-B2	7.7	supervised	ImageNet	0.692	0.770	0.523	0.544
EfficientNet-B3	10.7	supervised	ImageNet	0.744	0.784	0.541	0.551
EfficientNet-B4	17.5	supervised	ImageNet	0.704	0.749	0.541	0.547
ResNet-18	11.2	supervised	ImageNet	0.734	0.770	0.531	0.553
ResNet-50	23.5	supervised	ImageNet	0.710	0.718	0.529	0.530
ResNet-101	42.5	supervised	ImageNet	0.643	0.706	0.511	0.522
XCiT-S12	25.9	supervised	ImageNet	0.784	0.683	0.664	0.648
XCiT-M24	83.9	supervised	ImageNet	0.736	0.738	0.664	0.631

Table S3: AUROC scores for a K-Nearest neighbor prostate cancer classifier, fitted with features extracted from the Radboud dataset.

Model	Parameters	Training method	Training data	PESO	Karolinska	Helsinki30	Helsinki60
XCiT-S12	25.9	self-supervised	HelsinkiProstate	**0.934**	**0.854**	**0.959**	**0.919**
XCiT-M24	83.9	self-supervised	HelsinkiProstate	**0.934**	0.852	0.956	0.910
XCiT-S12	25.9	self-supervised	ImageNet	0.875	0.757	0.831	0.818
XCiT-M24	83.9	self-supervised	ImageNet	0.865	0.756	0.819	0.804
EfficientNet-B0	4.0	supervised	ImageNet	0.768	0.740	0.545	0.545
EfficientNet-B1	6.5	supervised	ImageNet	0.764	0.736	0.538	0.537
EfficientNet-B2	7.7	supervised	ImageNet	0.745	0.721	0.537	0.537
EfficientNet-B3	10.7	supervised	ImageNet	0.773	0.718	0.559	0.550
EfficientNet-B4	17.5	supervised	ImageNet	0.676	0.653	0.499	0.501
ResNet-18	11.2	supervised	ImageNet	0.786	0.737	0.567	0.569
ResNet-50	23.5	supervised	ImageNet	0.780	0.711	0.551	0.543
ResNet-101	42.5	supervised	ImageNet	0.690	0.683	0.517	0.526
XCiT-S12	25.9	supervised	ImageNet	0.805	0.714	0.763	0.746
XCiT-M24	83.9	supervised	ImageNet	0.797	0.723	0.776	0.732

Table S4: Cox proportional hazards model coefficients and p-values with Gleason grade and HistoEncoder feature cluster frequencies.

Variable	coefficient	exp(coefficient)	SE(coefficient)	CI, lower 95%	CI, upper 95%	*p*
GG1 (ref.)						
GG2	2.01	7.45	0.83	0.39	3.63	0.02
GG3	1.19	3.29	0.88	-0.54	2.92	0.18
GG4	2.59	13.31	0.84	0.94	4.24	<0.005
GG5	1.91	6.74	0.95	0.04	3.78	0.05
Cluster 25	3.16	23.47	0.70	1.79	4.52	<0.005
Cluster 9	3.06	21.40	0.97	1.16	4.97	<0.005
Cluster 6	2.64	13.98	1.39	-0.09	5.36	0.06
Cluster 13	-1.78	0.17	2.60	-6.88	3.32	0.49
Cluster 21	2.41	11.09	1.07	0.32	4.49	0.02
Cluster 16	4.89	133.20	1.61	1.74	8.05	<0.005

Concordance=0.81; L1-ratio=0.5; penalizer=0.001; observations=398; events=52

Table S5: Cox proportional hazards model coefficients and *p*-values with CAPRA-S nomogram and HistoEncoder feature cluster frequencies.

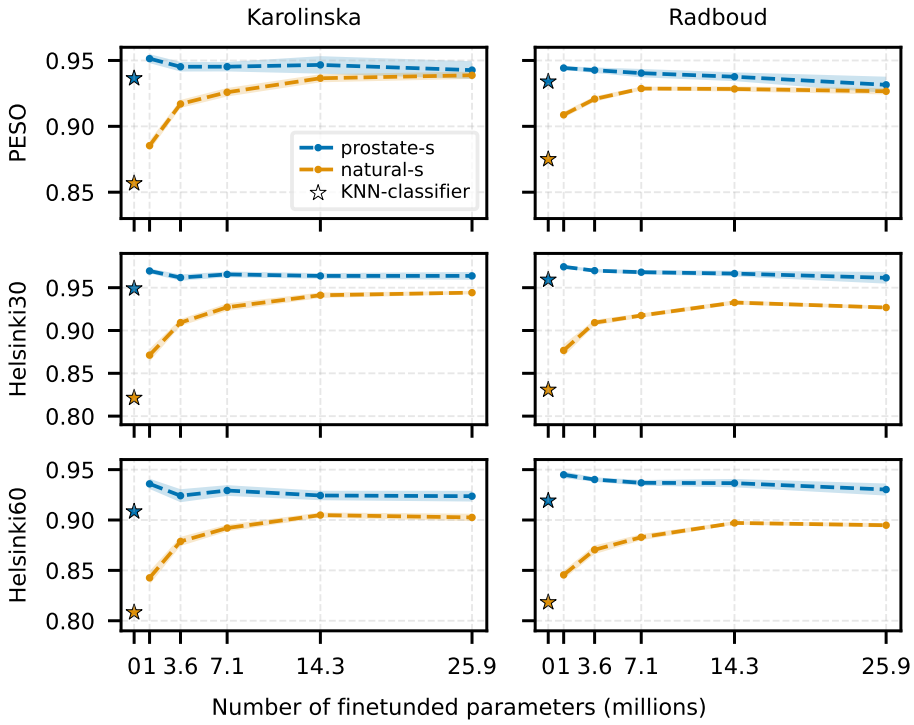
Variable	coefficient	exp(coefficient)	SE(coefficient)	CI, lower 95%	CI, upper 95%	*p*
CAPRA-S	0.45	1.58	0.11	0.24	0.67	<0.005
Cluster 25	0.60	1.82	0.13	0.34	0.85	<0.005
Cluster 9	0.43	1.54	0.11	0.21	0.66	<0.005
Cluster 6	0.34	1.40	0.12	0.10	0.58	0.01
Cluster 13	0.07	1.07	0.27	-0.46	0.59	0.80
Cluster 21	0.18	1.20	0.14	-0.09	0.45	0.19
Cluster 16	0.33	1.39	0.34	-0.34	1.00	0.33

Concordance=0.89; L1-ratio=0.5; penalizer=0.001; observations=285; events=26

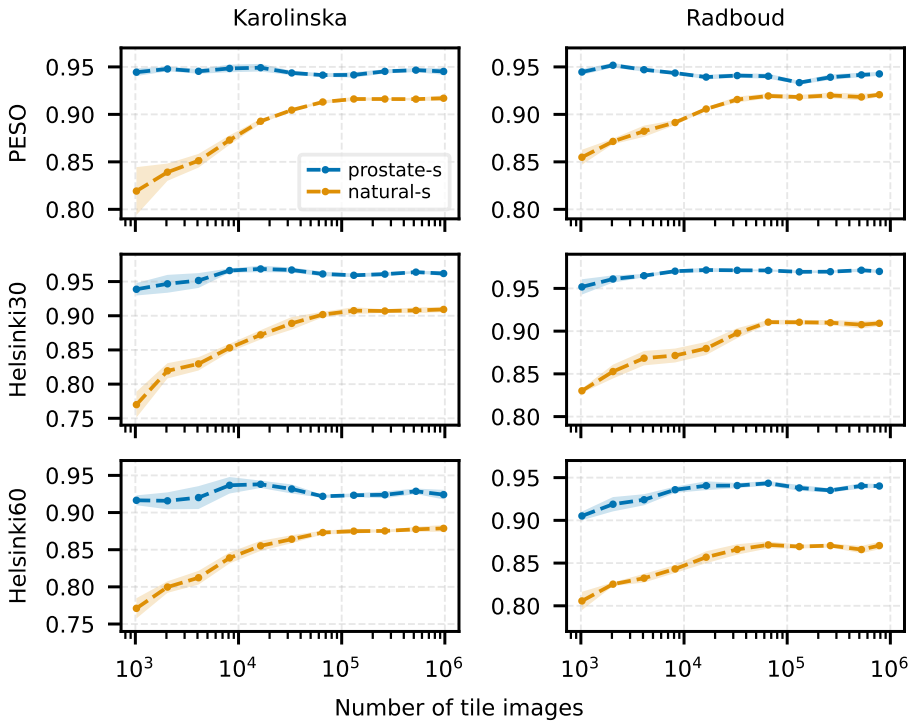
Table S6: Cox proportional hazards model coefficients and *p*-values with MSKCC-S nomogram and feature cluster frequencies.

Variable	coefficient	standard deviation	CI, lower 95%	CI, upper 95%	*p*	
MSKCC-S	-0.54	0.58	0.14	-0.82	-0.26	<0.005
Cluster 25	0.57	1.76	0.12	0.32	0.81	<0.005
Cluster 9	0.42	1.53	0.11	0.20	0.65	<0.005
Cluster 6	0.40	1.48	0.12	0.16	0.63	<0.005
Cluster 13	0.01	1.01	0.26	-0.50	0.52	0.98
Cluster 21	0.17	1.19	0.13	-0.09	0.43	0.20
Cluster 16	0.30	1.35	0.36	-0.40	1.00	0.40

Concordance=0.86; L1-ratio=0.5; penalizer=0.001; observations=285; events=26



(a) Limiting the number of fine-tuned encoder blocks.



(b) Limiting the amount of training data.

Figure S1: AUROC scores for the prostate cancer classifiers fine-tuned from `prostate-s` and `natural-s` encoder models using the Karolinska and Radboud datasets. `Prostate-s` based models achieve higher AUROC scores than `natural-s` based models with less fine-tuned parameters (*a*) and training data (*b*). A KNN-classifier achieves similar or significantly better performance than fully fine-tuned `natural-s` based models. When limiting the number of tile images (*b*), `prostate-s` based models significantly outperform `natural-s` based models even trained with 1000 times less data.

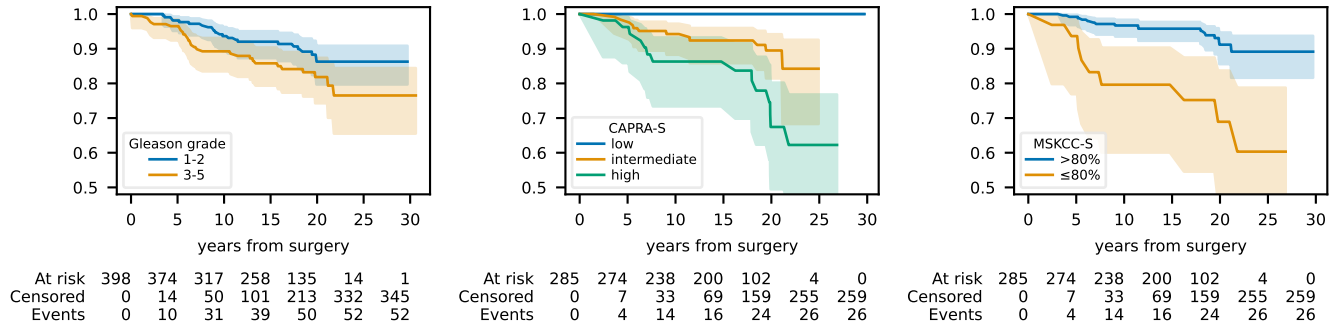


Figure S2: Kaplan-Meier survival curve estimations for Gleason grades and clinical nomograms in HelsinkiTMA cohort used for the prostate cancer death survival models.

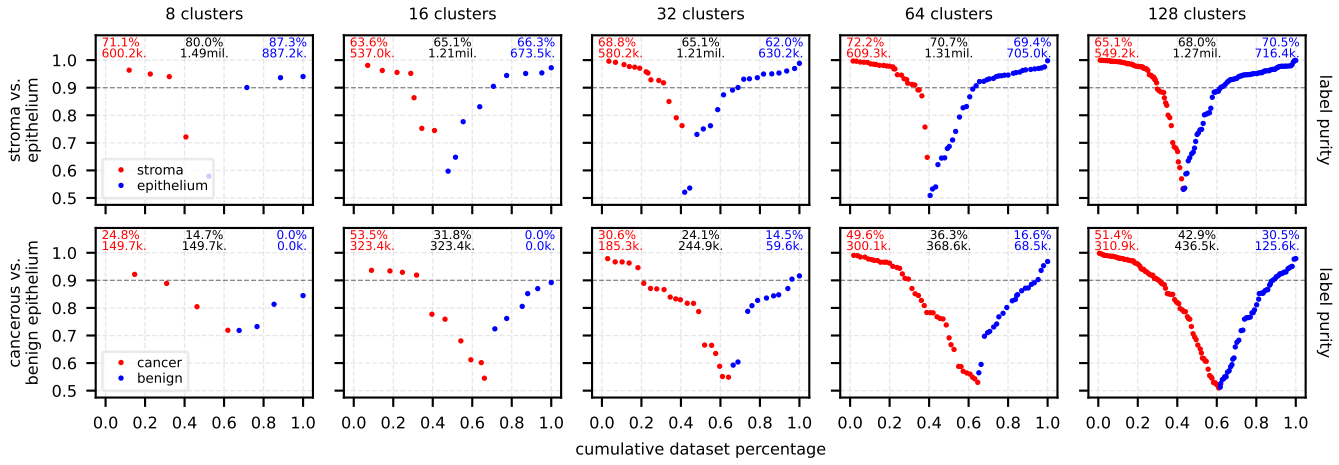


Figure S3: Cluster purities for the Radboud dataset with a varying number of clusters.

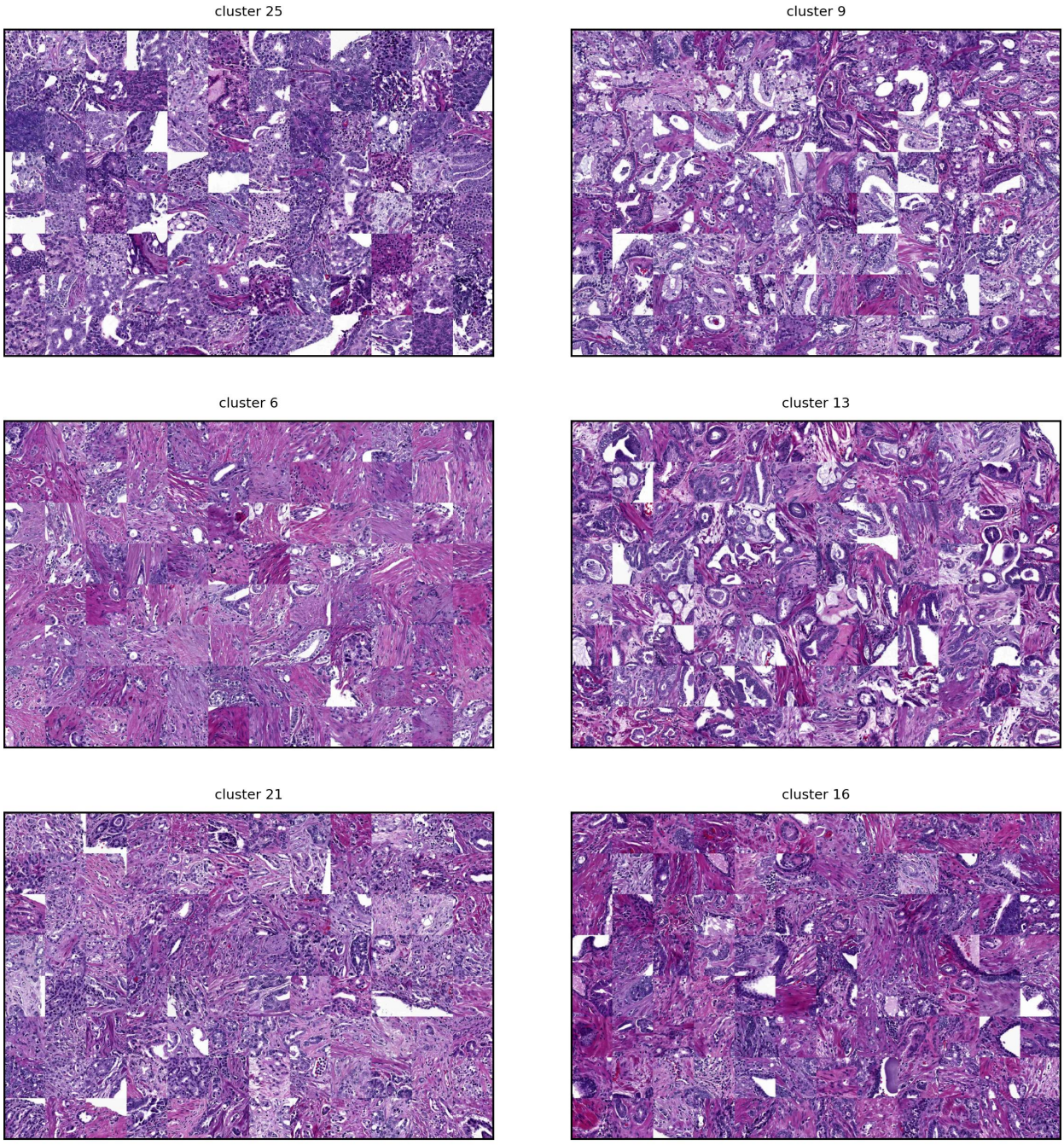


Figure S4: Randomly sampled tile images from the clusters used in the prostate cancer death survival models.