Pan-cancer multimodal foundation models to predict neoadjuvant immunotherapy response

Inez den Hond, 6149545
Writing assignment Bioinformatics & Biocomplexity, Universiteit Utrecht
February 2, 2024

Supervisor: dr. J. Teuwen,
Netherlands Cancer Institute
Examiner: dr. ir. J. de Ridder, UMC Utrecht

B.1.2 – Abstract

In this study, the main objective is to develop a large, pan-cancer multimodal foundation model that can be used to predict response to neoadjuvant immunotherapy, an unmet need in the field of oncology. The past years, neoadjuvant immunotherapy has emerged as a groundbreaking paradigm in oncology, transforming treatment strategies by exploiting the immune system before surgery to enhance therapeutic efficacy. Ongoing clinical trials highlight the potential of neoadjuvant immune checkpoint blockade as a promising strategy for many different cancer types. Clinical trials and other studies aim to find predictive biomarkers for neoadjuvant immunotherapy response. Nowadays, increasing amounts of biological data are collected as patients are diagnosed and treated in the hospital. Deep learning (DL) can be used to unlock the full potential of these datasets by learning intrinsic patterns in the data for cancer diagnosis, prognosis, and many other use cases. However, most DL methods are trained on a small amount of labelled data which limits the generalization capability of these models. Outside the medical domain, the rise of foundation models (FM) has led to major successes in natural language processing and computer vision. FMs are billion-parameter models that are trained on vast amounts of data with self-supervised learning, a DL method that allows the FM to develop a general understanding of unlabelled data. We aim to develop a large, multimodal FM that includes both histopathology and transcriptomics data from a large patient database. Furthermore, we will exploit our FM to predict response to neoadjuvant immunotherapy based on ongoing clinical trials. The development of a large, multimodal FM is unprecedented in oncology and will pave the way for personalized medicine.

B.1.3 – Layman's summary

Cancer is a disease caused by uncontrolled division of abnormal, tumour cells that can occur in different parts of the body. Traditional cancer treatments include surgery, chemotherapy, radiation, or immunotherapy. As the immune system is supressed in cancer cells, the body's natural defence mechanisms are disturbed. Immunotherapy aims to empower the immune system to recognize and destroy cancer cells more effectively. Recently, it has been found that giving immunotherapy prior to surgery even further improves the ability of the body to

combat the cancer. This is called neoadjuvant immunotherapy. Currently, many clinical trials are conducted to properly investigate the effect of neoadjuvant immunotherapy in different cancer types and so far, promising results have been delivered. However, not every patient responds well to this treatment approach, and it is important to decide as early as possible what patients can benefit from neoadjuvant therapy and what patients may have more benefit from an alternative treatment approach. Response to treatment is dependent on many patient characteristics. A personalized approach can identify for each patient whether he or she will be a responder and lead to more effective treatment decisions. Artificial intelligence (AI) methods can learn underlying patterns from large, medical datasets and can aid in cancer diagnosis, prognosis, and treatment strategies. As more data becomes available, AI models become bigger, and the most recent trend are foundation models (FM). The unique strength of foundation models compared to more traditional AI methods is that they are able to leverage the large datasets available in the hospital setting. Medical foundation models are often trained on one data type, for example on microscopic tissue images or genetic data. In this proposal, we aim to develop a foundation model that is trained on multiple, complementary data types. This multimodal approach will improve the FM's understanding of the underlying cancer biology. Finally, we will bring our multimodal FM into practice by predicting neoadjuvant immunotherapy response, which will ultimately contribute to a treatment approach tailored to each individual patient.

B.1.4 – Keywords

Foundation models – multimodal – neoadjuvant immunotherapy – precision oncology – self-supervised learning

B.2.1.a – Extensive background

Neoadjuvant immunotherapy

For patients with cancer, neoadjuvant immunotherapy has revolutionized the landscape of cancer treatment. Neoadjuvant immunotherapy encompasses the administration of immune checkpoint inhibitors prior to surgery, with the aim of strengthening the body's immune response against cancer cells (1). Compared to adjuvant immunotherapy, neoadjuvant immune checkpoint blockade (ICB) is thought to initiate a stronger and broader T-cell response as the tumour is still fully present. A larger immune response will lead to a reduction in tumour size which improves resectability (2,3). The field of neoadjuvant immunotherapy has progressed rapidly over the past few years and many clinical trials are going on to unlock the full potential of this novel treatment approach (1). ICB typically targets programmed death 1 (PD-1), programmed death ligand 1 (PD-L1) and cytotoxic T-lymphocyte antigen 4 (CTLA-4). The efficacy of these drugs in the preoperative setting was first seen in mouse studies and has been further validated in translational studies and clinical trials, with promising results for improved patient survival in several cancer types (4). Currently, there are two U.S. Food and Drug Administration (FDA)-approved neoadjuvant ICB treatment regimens for both resectable non-small-cell lung cancer (NSCLC) (5) and early triple-negative breast cancer (TNBC) (6), with many more on the way. Boydell et al. give a selection of ongoing neoadjuvant immunotherapy phase III trials (7).

However, not every patient benefits from neoadjuvant immunotherapy. Therefore, the next challenge is to identify at baseline what patients are likely to respond and who will likely benefit from a different treatment to avoid unnecessary treatment with toxic immunotherapy (3). Neoadjuvant immunotherapy is often given to treatment naïve patients where the cancer is in a resectable stage, so it is important to carefully select eligible patients as to ensure the best treatment for each patient. The plethora of clinical trials for neoadjuvant immunotherapy provide a goldmine of information about biomarkers that may contribute to a patient being a responder or non-responder to neoadjuvant ICB (4,7). The NICHE-1 and NICHE-2 phase II clinical trials have named mismatch repair deficiency (dMMR) as a promising biomarker for response to neoadjuvant ICB in early stage colon cancer (8,9), although only a small subset of patients have dMMR (7). So far, no predictive biomarker has been validated for neoadjuvant ICB response. Suggested potential biomarkers are PD-1/PD-L1 expression, the status of the tumour microenvironment (TME), presence of regulatory T cells, high tumour mutational burden, microsatellite instability, and T cell receptor diversity. It is highly unlikely there will be a single predictive biomarker for neoadjuvant immunotherapy response, as there are many complex interactions between the tumour and the immune system. The cancer immunogram depicts those interactions and states that many components of the immune system play a role in the response to cancer (3,4,7,10,11).

Deep learning for precision oncology

As cancer patients are diagnosed and treated in the hospital, massive amounts of biological information are collected about each patient. The understanding of individual patient's characteristics can help advance personalized therapy. Therefore, there lies an important opportunity to learn from these different data modalities, including clinical, radiological, histological and omics data, that complement each other. There is a challenge in how to effectively integrate this high-dimensional patient data, but deep learning (DL) shows enormous potential to advance clinical and biological understanding by identifying patterns and extracting hidden information in large biomedical datasets. Development of such multimodal DL methods allows for the discovery of data-driven biomarkers and pushes forward the rise of precision oncology (7,12–14).

Many unimodal AI models have been developed already to aid in cancer discovery, diagnosis, prognosis, and treatment (15). Boehm et al. provide an extensive overview with examples for radiology, histopathology, and genomics (12). Only a small number of these algorithms eventually gets approved for clinical use. In the field of radiology, many DL methods have been approved for clinical use, mainly for cancer detection and diagnosis (16). For adjuvant ICB, a radiomics-based biomarker was found to be associated with clinical outcome for pan-cancer patients treated with anti-PD-1 and PD-L1 (17). Compared to radiology images, H&E-stained whole slide images (WSI) provide more information about the tumour microenvironment. As these are acquired during routine clinical practice and thus abundantly available, histology images provide a promising avenue for DL-based cancer biomarker identification (13). Microsatellite instability (MSI) status, a possible predictive biomarker for response to neoadjuvant ICB, can be predicted directly from H&E stained whole slide images (WSI) for colorectal cancer (18), and regulatory approved diagnostic DL methods exist for breast cancer (19), colorectal cancer (19,20), and prostate cancer (21,22). In an exploratory study, Chen et al. have developed a machine learning method that can predict response to neoadjuvant ICB for breast cancer patients based on expression of immunological genes (23).

Multimodal DL methods provide great potential to integrate these different data types and develop cancer classifiers based on not one but multiple modalities. Much effort has been put into developing such multimodal DL methods as reviewed in Boehm et al. (12), and recently more multimodal models include histopathology images. In the field of neoadjuvant ICB, a recent study proposed a computed tomography (CT) based DL model that could predict pathological response to neoadjuvant immunotherapy in NSCLC. Based on a small cohort of 274 patients, they used both CT images and clinical characteristics (24). Vale-Silva et al. presented MultiSurv, a deep learning method that predicts pan-cancer survival based on different modalities including omics, clinical and imaging data from 11,081 patients from the publicly available The Cancer Genome Atlas (TCGA, (25)) (26). For each modality a specific subnetwork was trained that learns feature representations. After fusion of these representations, the final model predicts the survival probabilities for each patient. Best

performance was achieved when combining clinical data with either gene expression or DNA methylation data. MultiSurv is a promising example of how large-scale, high-dimensional, multimodal data can be leveraged to a powerful modelling approach.

The performance of DL methods is dependent on large datasets to properly capture relevant underlying patterns in the data. DL methods are often trained in a supervised manner where the model is trained on a labelled dataset to learn the relationship between an input (e.g. histopathological images) and an output (e.g. tumour subtype classification), to eventually predict the output for new, unseen input. The lack of large, publicly available, annotated datasets in the medical domain hampers the development of large machine learning models. Creation of such datasets requires manual annotation by medical experts which is a time-consuming process. Especially for computational pathology, extensive labelling of WSI on the patch level is exhaustive and requires expertise knowledge (27).

Good performance has been reached with supervised classification of breast cancer metastasis with a small, pixel-level labelled dataset of 400 WSI (28). However, the small sample size may restrain its clinical applicability due to a limited representation of clinical variation. Campanella et al. propose a framework that can train classification models without pixel-level labelled WSI (27). They created a large dataset containing 44,732 uncurated WSI from different institutions, thus representing clinical variability and artifacts encountered in the hospital setting. Only a small part of their dataset is publicly available. Instead of supervised learning as described above, they used the slide-level annotations to perform multiple instance learning (MIL), which is a type of weakly supervised learning (WSL). WSL is a DL paradigm where the model is trained using data that is labelled at a less precise level than for supervised learning. WSL is often used in scenarios where obtaining fully labelled datasets is challenging. MIL addresses this issue by giving each tile of the WSI the same high-level, weak label (e.g. diseased or healthy). This slide-level diagnosis is readily available from diagnostic reports and easy to come by, in contrast to the pixel- or tile-level annotation. The MIL classifier is first trained at the tile level and then aggregates the tile-level predictions to give a slide-level classification. This framework showed that pixel-level annotated WSI are not necessary to reach clinical grade performance for cancer diagnosis, which allows for the full use of biologically available data that carries only slide-level labels. PORPOISE is another WSL method, where a pan-cancer DL-based multimodal fusion algorithm was trained on H&E WSI and multiple molecular profile features to predict cancer survival on 6,592 WSI slides from TCGA (29).

Outside the medical domain there are lots of large, annotated datasets available, for example ImageNet (30), a database of currently around 15 million natural images. ImageNet can be deployed for DL tasks in the medical domain with transfer learning, a DL method to transfer knowledge from a general, pre-trained model (e.g. trained on ImageNet) to a model finetuned on a smaller, annotated dataset for a specific downstream medical task (e.g. classification from

histopathology slides). For example, both MultiSurv and PORPOISE use an ImageNet pretrained model (ResNeXt50 and ResNet50 respectively) to extract the feature representation from WSI for the downstream classification task (26,29). However, there are fundamental differences between natural and medical images. Histopathology images are very large in size, have no canonical orientation, have a different colour scheme, and the classification task is often dependent on small details instead of larger structures compared to natural images. This may restrain knowledge transfer as the features learned by the pretrained model may not be applicable to the medical domain (31–33).

Foundation models

Recently, foundation models (FM) have emerged as a new generation of AI frameworks that can learn to perform diverse tasks using large, unannotated datasets. FMs show amazing capacity to obtain general knowledge from extensive datasets that allows them to perform many different tasks. FMs are pretrained on large amounts of unlabelled data where the idea is to learn the intrinsic representations from datasets without explicit task-specific labels. This pretraining phase is crucial to gain a broad understanding of the patterns and general features of the data. After pretraining, FMs can be finetuned for specific tasks, making them adaptable for a wide range of downstream applications (34,35).

Examples of well-known FMs include BERT (36) and GPT (37) for natural language, and CLIP (38) is a multimodal FM that integrates natural language and images. Foundation models are built with self-supervised learning (SSL), where the model extracts meaningful representations from the input data by automatically creating tasks that enable the model to generate its own learning signals (34,39). There are multiple ways to perform this pretext task in SSL, including contrastive learning, masked modelling, and self-distillation. Contrastive learning involves training FMs to differentiate positive and negative pairs and is used in methods like SimCLR (40) and MoCov2 (41). Self-distillation, as seen with DINO (42), relies on a teacher-student framework, where a student learns to reconstruct its own perturbed input from the teacher. Masked modelling, as seen in BERT, tasks FMs with reconstructing masked portions of the input data. Masked modelling can also be applied to computer vision, for example with iBOT (43). iBOT is a self-supervised framework that uses masked image modelling and a vision transformer (ViT) architecture. ViT have shown remarkable performance for image understanding and processing compared to convolutional neural networks (CNN), for example ResNet (34,39).

In the medical domain, self-supervised learning is increasingly utilised to unlock the full potential of large-scale, unannotated datasets. FMs trained on biological datasets allows the model to learn the underlying patterns, thereby potentially enhancing diagnostic capabilities and many other medical tasks, without the need for exhaustive manual annotation of these large datasets. There are loads of unannotated medical datasets available (histology slides, radiology images, sequencing data, electronic health records, and protein sequences) and self-

supervised learning provides a promising avenue to use these datasets to the fullest extent (32). In the field of computational histopathology, much effort has been put towards developing self-supervised learning methods to use the large amounts of unlabelled WSI available. Using contrastive SSL, CTransPath is a histopathology FM trained on 30,000 WSI from TCGA and PAIP (44), another large publicly available WSI dataset. Based on CTransPath, Wagner et al. developed a transformer-based algorithm to predict MSI status from CRC histopathology slides (45). They have used CTransPath as the FM and used a MIL method for the biomarker prediction task. However, their method was based on resected slides instead of biopsy material which is preferred for a diagnostic method.

Filiot et al. were the first to develop a histopathology FM, Phikon, with masked image modelling using iBOT (31). Phikon was trained using 6,000 unlabelled WSI slides from TCGA. For equivalent model architectures (ViT-B), they showed that in-domain pretraining outperformed out-of-domain pretraining methods based on ImageNet. They have benchmarked several MIIL methods like ABMIL, TransMIL, and Chowder, and used Chowder as the MIL method in the final model. On various weakly supervised slide-level classification tasks (MSI status, subtype prediction, metastasis detection) evaluated on both TCGA and external validation cohorts, Phikon outperforms CTransPath and other state-of-the-art histopathology SSL methods, both with cohort-specific and pan-cancer pretraining.

As the performance of FMs scales with the size of the dataset, recently focus has shifted towards increasing the size of the dataset for pretraining. Most SSL models for histopathology, including CTransPath and Phikon, use TCGA as their main data source that includes around 11,000 diagnostic WSI. However, data leakage can occur if FMs are pretrained on TCGA data and the subsequent downstream, classification tasks are evaluated on TCGA data as well. Last year, UNI was released, a ViT-L SSL model pretrained on 100,000 in-house pan-cancer WSI using DINOv2 (46). They have shown to outperform CTransPath on all weakly-supervised slide-level classification tasks on external test cohorts (excluding TCGA tasks that have no external test cohort), including tasks like NSCLC subtyping, breast metastasis detection, and several mutation prediction tasks. UNI uses ABMIL as the method to perform weakly supervised slide classification after extracting the features from WSI from the pretrained model.

Recently Virchow was developed, to date the largest histopathology foundation model trained on 1.5 million pan-cancer WSI from Memorial Sloan Kettering (MSK) Cancer Center. Virchow is based on a ViT-Huge architecture and trained with the DINOv2 SSL algorithm. They state that a student-teacher based SSL method like DINOv2 is more suitable for a class-imbalanced dataset like WSI than contrastive learning, as is used in CTransPath. They performed weakly supervised slide-level biomarker prediction (MSI, FGFR, and EGFR) on unseen datasets, using AGATA as the MIL method. Virchow outperformed CTransPath on the same task for every biomarker. The unprecedented performance of UNI and Virchow emphasizes the importance of pretraining FMs on massive pan-cancer histopathology datasets.

Self-supervised learning is not limited to histopathology but can be applied to different modalities as well. For example, Padegal et al. predicted vital status for colorectal cancer patients based on RNA-sequencing data using self-supervised learning (47). As gene expression data is of tabular format, they used a pre-existing SSL framework called TabNet (48). Their model was pretrained on unlabelled gene expression from TCGA and thereafter finetuned on the labelled target dataset (a subset of their train dataset). As gene expression is very high-dimensional, they used rigorous feature selection beforehand, reducing the number of features from 60,660 gene IDs to ~200. They selected the features on the labelled dataset and used those features as a filter to perform feature reduction on the unlabelled dataset, thus using the labelled dataset during the pretraining stage. In this model, clinical and copy number variation were also included making this a multimodal tabular foundation model.

Last year, a visual-language foundation model was developed specifically for histopathology, trained on large dataset of 208,414 images and their textual descriptions curated from Twitter and other public internet resources (49). The PLIP (pathology language-image pretraining) model is trained using contrastive learning between the language and image modalities and is based on CLIP (contrastive language-image pretraining). CLIP is a large, well-known visual-language foundation model (38) pretrained on a large dataset of images with captions. PLIP is evaluated on external validation datasets and was shown to outperform the baseline CLIP.

Taking all these developments into account, it can be concluded that the development of a large, multimodal foundation model for oncology has been unprecedented. Such a model will pave the way for tumour classification, clinical response prediction, and many other downstream tasks.

B.2.1.b – Overall aim

Using large-scale unlabelled histopathology and transcriptomic data from cancer patients, we aim to develop a pan-cancer multimodal foundation model that later can be used for a wide range of downstream tasks, including neoadjuvant immunotherapy response prediction. We will use self-supervised learning techniques to develop a foundation model that learns the underlying intrinsic structure of the data, that subsequently can be used for more specific downstream tasks with supervised training on a smaller, labelled dataset. Our foundation model will be the first to integrate different medical modalities on such a large scale. Using our FM, we aim to predict pathologic response to neoadjuvant immunotherapy for TNBC, NSCLC, and melanoma among others, as there are currently limited biomarkers available, and an AI-based decision method can reduce the need for extensive molecular profiling.

B.2.1.c – Objectives

Objective 1: Develop a pan-cancer multimodal foundation model for gene expression and histopathology images.

Objective 2: Validate the performance of the multimodal foundation model on different downstream classification tasks.

Objective 3: Use the multimodal foundation model to predict response to neoadjuvant therapy for cancer patients.

B.2.2.a – Design of the proposed research

We aim to develop a multimodal foundation model based on large amounts of histopathology and transcriptomics data. This foundation model can be used for a wide variety of downstream tasks, where we will focus on the prediction of neoadjuvant immunotherapy response for different cancer types. The FM will be trained on a large dataset from both in-house and publicly available datasets, and will be finetuned on both publicly available datasets and in-house clinical trial datasets.

Data collection for FM training

FM training data will be collected from the internal patient database (The Netherlands Cancer Institute) where currently 83,000 diagnostic WSI are available and a subset of TCGA cohorts (8,180 diagnostic WSI). From TCGA, we will use all available cohorts except the ones mentioned in table 1. Per patient we will collect gene expression and digital histopathological images to build our pan-cancer dataset. We will select unlabelled formalin-fixed paraffin-embedded (FFPE) haematoxylin and eosin (H&E) stained WSI. The distribution of cancer types in the internal patient database is unknown but is expected to follow the same trend seen in TCGA, where more prevalent cancer types have more entries into the database.

WSI preprocessing

WSI will be preprocessed as described by Filiot et al. in section 3.2 (31). Filiot et al. have downsampled the number of tiles across slides to obtain the same number of tiles for each cancer type. We will skip this down sampling step to increase the size of our dataset. Virchow did not use down sampling to create their pan-cancer dataset, they have thus used an imbalanced dataset with a larger occurrence for more prevalent cancer types to train their model. At a later stage, we can explore the effect of balancing the number of tiles per tissue type in the pretraining dataset.

Gene expression data preprocessing

Preprocessing of RNA-sequencing data will be performed similar to Padegal et al. (47). They have defined the fragments per kilobase of transcripts per million mapped reads upper quartile (FPKM-UQ) as the preferred method to represent the gene expression values. As gene expression data is high-dimensional, we will follow a similar feature selection strategy where first only the protein coding genes are selected. As a second step, Padegal et al. perform feature reduction on a smaller, labelled dataset to then use these features as a filter on the larger unlabelled dataset. This is unfavourable for our situation where we want to use the gene expression values from an unlabelled dataset without using a labelled dataset. Therefore, we will use all 19,962 protein coding genes.

FM training

To develop a multimodal foundation model for histopathology and transcriptomics data, we will adapt existing multimodal foundation model architectures. PLIP is a multimodal

histopathology – language FM that encodes image and text into one embedding using contrastive learning. We will adapt this approach to image and tabular data, and we aim to fuse the two modalities as early as possible. For both modalities, we will build a transformer model specific to that modality. For the histopathology images, we will use a ViT-B architecture, based on Huang et al. and Filiot et al. UNI and Virchow implemented a ViT-L architecture but also used a larger dataset. The vision transformer takes tiles of 224 x 224 pixels as input. For the tabular gene expression data, we will use a TabNet architecture. Both encoders will output a vector of similar dimension, and during training these vectors will be optimized by minimizing the contrastive loss between similar pairs. This will ensure the FM will learn from both modalities simultaneously. The FM will be trained on the in-house pan-cancer dataset. Once trained, the FM returns the learned representations for a given input of histopathology and gene expression data.

Evaluation of the FM on downstream classification tasks

Next, the pretrained FM can be used to perform downstream, supervised classification tasks. We will evaluate our FM on an extensive set of downstream tasks that are either common in related studies or fit our research question. These tasks involve publicly available datasets so we can compare the performance of our FM to other state-of-the-art SSL methods.

Data collection for finetuning tasks

We collect data from TCGA (25). An exploratory search at the TCGA repository showed 11,765 available diagnostic FFPE WSI from 32 different TCGA cohorts with multiple digital histopathological slides available per case. We will only include patients if they have both WSI and RNA-sequencing data available. Preprocessing of WSI and gene expression data will be handled similar as described above.

Finetuning tasks

We have defined a set of downstream tasks to evaluate our multimodal FM inspired by the evaluation tasks of other histology pretrained FMs like Phikon, Virchow, and UNI. In these studies, usually both slide-level and patch-level tasks are evaluated. We will only evaluate slide-level classification tasks as patch-level tasks are not directly relevant to our research question. Patch-level classification tasks are generally considered easier whereas slide-level classification resembles real-world scenarios and our intended research aim (50).

To perform the finetuning, we first extract the features from our pretrained FM to then train a MIL model on the downstream classification task using weakly supervised learning. The MIL methods discussed in the background section (Chowder, ABMIL, TransMIL) are all aimed at weakly supervised learning for just histopathological images and do not enable integration with tabular data. We will adapt the multimodal multi-instance fusion module (M3IF) that was developed to perform MIL with a cross-modal representation from both histopathological

images and tabular clinical data (51), where we will use tabular gene expression data instead of clinical data.

We have defined a number of downstream classification tasks based on TCGA cohorts, as TCGA provides both histopathological images and gene expression data. The histopathology FMs also evaluated their performance on other datasets like Camelyon16 and PAIP-CRC but they consist of only WSI and are thus not applicable to our FM. Table 1 describes the classification tasks, the TCGA cohort they are based on, and if the task is evaluated in one of the other histopathology FMs. The number of WSI gives an estimate of the number of patients as it can occur that one patient has multiple WSI entries in the database. Also, the TCGA data portal does not allow easy filtering to find the exact number of patients that have both WSI and gene expression data available. The final number of included patients may therefore be lower than the number of WSI stated in this table.

Task	Cohort	Nr. of WSI	Performed in other study
Histological subtype prediction for renal cell cancer	TCGA-KIRC, TCGA-KIRP, TCGA-KICH	940	(31)
Molecular subtype prediction for breast cancer	TCGA-BRCA	1,133	(31)
Cancer type prediction for NSCLC	TCGA-LUAD, TCGA-LUSC	1,053	(31,46)
Genomic alteration prediction (MSI)	TCGA-COAD	459	(31,46)

Table 1 – Overview of downstream TCGA-based classification tasks, the cohorts that will be used, the number of WSI that is available for those cohorts, and from what study the task has been adapted. NSCLC = non-small cell lung cancer, MSI = microsatellite instability, WSI = whole slide images, KIRC = kidney renal clear cell carcinoma, KIRP = kidney renal papillary cell carcinoma, KICH = kidney chromophobe, BRCA = breast invasive carcinoma, LUAD = lung adenocarcinoma, LUSC = lung squamous cell carcinoma, COAD = colon adenocarcinoma.

Prediction of neoadjuvant immunotherapy response

Finally, we will evaluate the performance of our foundation model on our intended downstream task: prediction of neoadjuvant immunotherapy response from histopathological images and gene expression data. The downstream classification task will be performed similar as the TCGA tasks, first extracting the learned features from the multimodal FM and

subsequently performing MIL to classify patients. Table 2 provides an overview of the ongoing clinical trial cohorts at the NKI. If other clinical trial studies are available at time of this research, they will be included in this step as well. For example, the TONIC2 trial is expected to finish in 2024. For each trial, we will select the patients that received neoadjuvant immunotherapy. The downstream, finetuning tasks is defined as predicting pathologic complete response (pCR). pCR is achieved when no more tumour cells are found after ICB. It has been shown that pCR is correlated with overall survival, but it still has to be validated as a regulatory-approved endpoint for neoadjuvant immunotherapy. As pCR is evaluated shortly after treatment with neoadjuvant ICB, it allows the physician to alter the treatment plan based on pCR (3,7).

Trial name	Cancer type	Trial phase	Nr. of patients included (so far) that received neoadjuvant ICB
NADINA (52)	Stage III melanoma	Phase III	210
OpaCIN(-neo) (53)	Stage III melanoma	Phase II	96
PRADO (54)	Stage III melanoma	Phase II	99
NICHE (8)	Early-stage colon cancer	Phase II	40
NICHE2 (9)	Early-stage colon cancer	Phase II	112
NABUCCO (55)	Stage III urothelial cancer	Phase I	24
IMCISION (56)	HNSCC	Phase Ib/II	32

Table 2 – Overview of clinical trials that will be used to predict neoadjuvant immunotherapy response. All trials are performed at the Netherlands Cancer Institute TNBC = triple negative breast cancer, ICB = immune checkpoint blockade, HNSCC = head and neck squamous cell carcinoma.

B.2.2.b – Work plan / research lines

Objective 1: Develop a pan-cancer multimodal foundation model for gene expression and histopathology images (2 years).

Deliverables:

- Literature review outlining current state-of-the-art approaches and technologies related for existing medical FMs, focussing on histopathology, transcriptomics, and multimodal efforts.

- Curation of a pan-cancer dataset from both in-house (NKI) and publicly available (TCGA) data.
- Development and evaluation of a robust, multimodal FM.
- Complete documentation of FM training procedure to enhance reproducibility.

Objective 2: Validate the performance of the multimodal foundation model on different downstream classification tasks (1 year).

Deliverables:

- Curation of dataset for different downstream classification tasks from an external dataset (TCGA).
- Finetuning of the multimodal FM on these downstream tasks
- Rigorous evaluation of the performance on these downstream tasks, including comparison with existing state-of-the-art methods.
- Complete documentation of FM finetuning procedure to enhance reproducibility.

Objective 3: Use the multimodal foundation model to predict response to neoadjuvant therapy for cancer patients (1 year).

Deliverables:

- Curation of dataset of in-house (NKI) clinical trial data.
- Finetuning of the multimodal FM to predict neoadjuvant immunotherapy response.
- Validation of the performance on the prediction task in close collaboration with physicians.

B.2.3 Feasibility & risk assessment

The practical feasibility of this research proposal is demonstrated by several key factors that together will ensure its realization. First, we have robust data availability from both in-house and publicly available datasets. The Netherlands Cancer Institute's extensive collection of clinical trial and general patient data is directly available to us, and we will use publicly available datasets like TCGA. Second, the lab's prior experience with developing self-supervised learning methods for histopathology (57) establishes a solid groundwork for the proposed research. Third, a potential research collaboration with Memorial Sloan Kettering (MSK) Cancer Center adds a layer of strength to our research as they have prior experience with developing large foundation models (58). In terms of time planning, a timeline has been outlined in the previous section, accounting for dataset curation, model development, model validation, and potential refinements.

The greatest potential risk lies in the novelty of our proposed research, being that our multimodal foundation model may encounter difficulties in capturing interactions between the different data modalities. This can lead to model instability or suboptimal performance and will be reflected in a low performance on the supervised, downstream tasks compared to the benchmark methods. To mitigate this risk, an alternative approach involves adopting a

modular model design where we will develop a separate foundation model for each modality. The representations learned by each model will then be fused before the inference task. For the histopathology data, we will use a similar model architecture and training procedure as Phikon (31). For the transcriptomics data, we will use TabNet as the self-supervised learning framework as proposed by Padegal et al. (47). For the downstream inference tasks, we will use the same approach as in our original proposed research. This modular approach aims to tackle the challenges related to integration in the foundation model while making sure the strengths of each modality are still effectively captured. The downside of this approach is that the early integration of the different modalities is sacrificed.

In the event that both the multimodal foundation model and the modular approach face unsurmountable challenges, we will revert to a more traditional DL approach, similar to the design of MultiSurv (26). As this is a supervised learning method, we will adapt MultiSurv to predict neoadjuvant immunotherapy response from a histopathology and transcriptomic sub model. Even though in this case we will no longer develop a large, multimodal foundation model, there are not many studies yet that aim at predicting response to neoadjuvant immunotherapy, and the availability of these studies at the NKI still secures the novelty of this research.

The proposed research seamlessly integrates with the available resources and expertise at the Netherlands Cancer Institute (NKI), which has been at the forefront of neoadjuvant immunotherapy. This treatment paradigm was first pioneered at the NKI and since then many clinical trials in the field are hosted by the institute's research groups. Our proposed research benefits from access to extensive clinical trial and general in-house patient data, providing a solid foundation for developing a large DL model for predicting neoadjuvant immunotherapy response. Furthermore, the NKI has access to a high-performance computing (HPC) cluster that aligns with the computational demands of the proposed research, enabling us to efficiently process large datasets for the development and finetuning of our multimodal FM. This strategic embedding promises a beneficial collaboration that can lead to important breakthroughs in cancer treatment.

B.2.4.a – Scientific impact

This research proposal holds considerable scientific significance as it describes the first multimodal foundation model that combines histopathology and genomics data. On one hand, effort has been put into developing large foundation models for histopathology (31,46,58) and transcriptomics (47). On the other hand, a few studies have investigated the potential of weakly supervised learning for pan-cancer multimodal survival prediction (26,29). We are the first to propose a general-purpose foundation model built on different modalities that can be used for a wide variety of downstream tasks beyond our intended application for prediction of neoadjuvant immunotherapy response. For example, the FM can be used for cancer diagnosis, metastasis detection, subtype classification, and many other applications.

Furthermore, our proposed research can help in understanding the workings of neoadjuvant immunotherapy by investigating the differences between responders and non-responders. This reverse translation is thought to help with uncovering the therapeutic targets of neoadjuvant immunotherapy as well as define novel predictive biomarkers for response (1). We currently propose the integration of two different modalities in a foundation model. The potential for extension with additional modalities like clinical information and radiology images paves the way for further improvement and adaptation to an ever-increasing amount of available biological data.

Beyond its immediate application, the development of a multimodal, medical foundation model contributes to the broader impact of artificial intelligence in healthcare. The ability to unravel underlying patterns across diverse medical datasets not only has implications for prediction of treatment response but sets a precedent for the use of DL in the landscape of precision medicine. Lastly, we will release the code and the weights of the model publicly to ensure full reproducibility and transparency.

B.2.4.b – Societal impact

The direct aim of this research, the development of a multimodal foundation model for predicting neoadjuvant immunotherapy response, has the potential to revolutionize cancer treatment. By enabling a personalized and precise approach, this research could lead to more effective and targeted interventions in patient care, avoiding unnecessary treatment-related side effects and improving overall patient quality of life. An efficient DL method for predicting treatment outcomes will reduce the need for extensive molecular profiling and make treatment decisions quicker and more cost-effective.

If neoadjuvant immunotherapy becomes the standard of care, such a DL framework may become essential for predicting treatment response in clinical practice. Further development and extensive clinical validation of both neoadjuvant immunotherapy and such DL methods is required before approval for clinical practice, but this research provides a promising first step in that direction. Our multimodal approach can be used beyond oncology to enable personalized medicine in other diseases as well which will impact the life of many patients.

B.2.5 – Ethical considerations

In this research proposal, both publicly available and in-house datasets are used. For the publicly available data collected from TCGA, the donors are anonymised and ethical approval from the patients has been obtained by TCGA. For the internal general patient dataset, informed consent is obtained when the patient is admitted to the hospital. For the internal clinical trial dataset, informed consent is obtained when patients are enrolled to the clinical trial by the hospital. For both the general patient and the clinical trial dataset we use anonymised data. By taking these measures, we ensure there is no personal information present in the datasets our FM is trained on and that it can never expose sensitive information.

B.2.6 – References

1. Topalian SL, Forde PM, Emens LA, Yarchoan M, Smith KN, Pardoll DM. Neoadjuvant immune checkpoint blockade: A window of opportunity to advance cancer immunotherapy. Cancer Cell. 2023 Sep 11;41(9):1551–66.

2. O'Donnell JS, Hoefsmit EP, Smyth MJ, Blank CU, Teng MWL. The Promise of Neoadjuvant Immunotherapy and Surgery for Cancer Treatment. Clin Cancer Res. 2019 Oct 1;25(19):5743–51.

3. Versluis JM, Long GV, Blank CU. Learning from clinical trials of neoadjuvant checkpoint blockade. Nat Med. 2020 Apr;26(4):475–84.

4. Mittendorf EA, Burgers F, Haanen J, Cascone T. Neoadjuvant Immunotherapy: Leveraging the Immune System to Treat Early-Stage Disease. Am Soc Clin Oncol Educ Book. 2022 Jul;(42):189–203.

5. Forde PM, Spicer J, Lu S, Provencio M, Mitsudomi T, Awad MM, et al. Neoadjuvant Nivolumab plus Chemotherapy in Resectable Lung Cancer. N Engl J Med. 2022 May 26;386(21):1973–85.

6. Schmid P, Cortes J, Dent R, Pusztai L, McArthur H, Kümmel S, et al. Event-free Survival with Pembrolizumab in Early Triple-Negative Breast Cancer. N Engl J Med. 2022 Feb 10;386(6):556–67.

7. Boydell E, Sandoval JL, Michielin O, Obeid M, Addeo A, Friedlaender A. Neoadjuvant Immunotherapy: A Promising New Standard of Care. Int J Mol Sci. 2023 Jul 24;24(14):11849.

8. Chalabi M, Fanchi LF, Dijkstra KK, Van den Berg JG, Aalbers AG, Sikorska K, et al. Neoadjuvant immunotherapy leads to pathological responses in MMR-proficient and MMR-deficient early-stage colon cancers. Nat Med. 2020 Apr;26(4):566–76.

9. Chalabi M, Verschoor YL, van den Berg J, Sikorska K, Beets G, Lent AV, et al. LBA7 Neoadjuvant immune checkpoint inhibition in locally advanced MMR-deficient colon cancer: The NICHE-2 study. Ann Oncol. 2022 Sep 1;33:S1389.

10. Blank CU, Haanen JB, Ribas A, Schumacher TN. The "cancer immunogram." Science. 2016;352(6286):658–60.

11. Wang S, Yuan P, Mao B, Li N, Ying J, Tao X, et al. Genomic features and tumor immune microenvironment alteration in NSCLC treated with neoadjuvant PD-1 blockade. NPJ Precis Oncol. 2022 Jan 13;6:2.

12. Boehm KM, Khosravi P, Vanguri R, Gao J, Shah SP. Harnessing multimodal data integration to advance precision oncology. Nat Rev Cancer. 2022 Feb;22(2):114–26.

13. Echle A, Rindtorff NT, Brinker TJ, Luedde T, Pearson AT, Kather JN. Deep learning in cancer pathology: a new generation of clinical biomarkers. Br J Cancer. 2021 Feb;124(4):686–96.

14. Steyaert S, Pizurica M, Nagaraj D, Khandelwal P, Hernandez-Boussard T, Gentles AJ, et al. Multimodal data fusion for cancer biomarker discovery with deep learning. Nat Mach Intell. 2023 Apr;5(4):351–62.

15. Bhinder B, Gilvary C, Madhukar NS, Elemento O. Artificial Intelligence in Cancer Research and Precision Medicine. Cancer Discov. 2021 Apr;11(4):900–15.

16. Swanson K, Wu E, Zhang A, Alizadeh AA, Zou J. From patterns to patients: Advances in clinical machine learning for cancer diagnosis, prognosis, and treatment. Cell. 2023 Apr;186(8):1772–91.

17. Sun R, Limkin EJ, Vakalopoulou M, Dercle L, Champiat S, Han SR, et al. A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. Lancet Oncol. 2018 Sep 1;19(9):1180–91.

18. Echle A, Grabsch HI, Quirke P, van den Brandt PA, West NP, Hutchins GGA, et al. Clinical-Grade Detection of Microsatellite Instability in Colorectal Tumors by Deep Learning. Gastroenterology. 2020 Oct;159(4):1406-1416.e11.

19. AI predicts cancer relapse from slides. Nat Biotechnol. 2022 Oct 1;40(10):1423–1423.

20. Saillard C, Dubois R, Tchita O, Loiseau N, Garcia T, Adriansen A, et al. Blind validation of MSIntuit, an AI-based pre-screening tool for MSI detection from histology slides of colorectal cancer [Internet]. medRxiv; 2022 [cited 2024 Jan 31]. p. 2022.11.17.22282460. Available from: https://www.medrxiv.org/content/10.1101/2022.11.17.22282460v1

21. Eloy C, Marques A, Pinto J, Pinheiro J, Campelos S, Curado M, et al. Artificial intelligence–assisted cancer diagnosis improves the efficiency of pathologists in prostatic biopsies. Virchows Arch. 2023;482(3):595–604.

22. Raciti P, Sue J, Retamero JA, Ceballos R, Godrich R, Kunz JD, et al. Clinical Validation of Artificial Intelligence-Augmented Pathology Diagnosis Demonstrates Significant Gains in Diagnostic Accuracy in Prostate Cancer Detection. Arch Pathol Lab Med. 2023 Oct 1;147(10):1178–85.

23. Chen J, Hao L, Qian X, Lin L, Pan Y, Han X. Machine learning models based on immunological genes to predict the response to neoadjuvant therapy in breast cancer patients. Front Immunol [Internet]. 2022 [cited 2024 Jan 25];13. Available from: https://www.frontiersin.org/articles/10.3389/fimmu.2022.948601

24. She Y, He B, Wang F, Zhong Y, Wang T, Liu Z, et al. Deep Learning for Predicting Major Pathological Response to Neoadjuvant Immunotherapy in Non-Small Cell Lung Cancer: A Retrospective Multicenter Study [Internet]. Rochester, NY; 2022 [cited 2024 Jan 17]. Available from: https://papers.ssrn.com/abstract=4070416

25. The Cancer Genome Atlas Program (TCGA) - NCI [Internet]. 2022 [cited 2024 Feb 1]. Available from: https://www.cancer.gov/ccg/research/genome-sequencing/tcga

26. Vale-Silva LA, Rohr K. Long-term cancer survival prediction using multimodal deep learning. Sci Rep. 2021 Jun 29;11(1):13505.

27. Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat Med. 2019 Aug;25(8):1301–9.

28. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. JAMA. 2017 Dec 12;318(22):2199–210.

29. Chen RJ, Lu MY, Williamson DFK, Chen TY, Lipkova J, Noor Z, et al. Pan-cancer integrative histology-genomic analysis via multimodal deep learning. Cancer Cell. 2022 Aug 8;40(8):865-878.e6.

30. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition [Internet]. 2009 [cited 2024 Jan 26]. p. 248–55. Available from: https://ieeexplore.ieee.org/document/5206848

31. Filiot A, Ghermi R, Olivier A, Jacob P, Fidon L, Mac Kain A, et al. Scaling Self-Supervised Learning for Histopathology with Masked Image Modeling [Internet]. Pathology; 2023 Jul [cited 2023 Nov 15]. Available from: http://medrxiv.org/lookup/doi/10.1101/2023.07.21.23292757

32. Krishnan R, Rajpurkar P, Topol EJ. Self-supervised learning in medicine and healthcare. Nat Biomed Eng. 2022 Dec;6(12):1346–52.

33. Raghu M, Zhang C, Kleinberg J, Bengio S. Transfusion: Understanding Transfer Learning for Medical Imaging. In: Advances in Neural Information Processing Systems [Internet]. Curran Associates, Inc.; 2019 [cited 2023 Dec 14]. Available from: https://proceedings.neurips.cc/paper_files/paper/2019/hash/eb1e78328c46506b46a4ac4a1e378b91-Abstract.html

34. Bommasani R, Hudson DA, Adeli E, Altman R, Arora S, von Arx S, et al. On the Opportunities and Risks of Foundation Models [Internet]. arXiv; 2022 [cited 2023 Sep 29]. Available from: http://arxiv.org/abs/2108.07258

35. Moor M, Banerjee O, Abad ZSH, Krumholz HM, Leskovec J, Topol EJ, et al. Foundation models for generalist medical artificial intelligence. Nature. 2023 Apr;616(7956):259–65.

36. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Burstein J, Doran C, Solorio T, editors. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) [Internet]. Minneapolis, Minnesota: Association for Computational Linguistics; 2019 [cited 2024 Jan 31]. p. 4171–86. Available from: https://aclanthology.org/N19-1423

37. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language Models are Few-Shot Learners [Internet]. arXiv; 2020 [cited 2024 Jan 31]. Available from: http://arxiv.org/abs/2005.14165

38. Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, et al. Learning Transferable Visual Models From Natural Language Supervision. In: Proceedings of the 38th International Conference on Machine Learning [Internet]. PMLR; 2021 [cited 2024 Jan 29]. p. 8748–63. Available from: https://proceedings.mlr.press/v139/radford21a.html

39. Balestriero R, Ibrahim M, Sobal V, Morcos A, Shekhar S, Goldstein T, et al. A Cookbook of Self-Supervised Learning [Internet]. arXiv; 2023 [cited 2023 Sep 29]. Available from: http://arxiv.org/abs/2304.12210

40. Chen T, Kornblith S, Norouzi M, Hinton G. A Simple Framework for Contrastive Learning of Visual Representations. In: Proceedings of the 37th International Conference on Machine Learning [Internet]. PMLR; 2020 [cited 2024 Feb 2]. p. 1597–607. Available from: https://proceedings.mlr.press/v119/chen20j.html

41. Chen X, Fan H, Girshick R, He K. Improved Baselines with Momentum Contrastive Learning [Internet]. arXiv; 2020 [cited 2024 Feb 2]. Available from: http://arxiv.org/abs/2003.04297

42. Caron M, Touvron H, Misra I, Jégou H, Mairal J, Bojanowski P, et al. Emerging Properties in Self-Supervised Vision Transformers [Internet]. arXiv; 2021 [cited 2024 Feb 2]. Available from: http://arxiv.org/abs/2104.14294

43. Zhou J, Wei C, Wang H, Shen W, Xie C, Yuille A, et al. iBOT: Image BERT Pre-Training with Online Tokenizer [Internet]. arXiv; 2022 [cited 2024 Jan 10]. Available from: http://arxiv.org/abs/2111.07832

44. Kim YJ, Jang H, Lee K, Park S, Min SG, Hong C, et al. PAIP 2019: Liver cancer segmentation challenge. Med Image Anal. 2021 Jan;67:101854.

45. Wagner SJ, Reisenbüchler D, West NP, Niehues JM, Zhu J, Foersch S, et al. Transformer-based biomarker prediction from colorectal cancer histology: A large-scale multicentric study. Cancer Cell. 2023 Sep 11;41(9):1650-1661.e4.

46. Chen RJ, Ding T, Lu MY, Williamson DFK, Jaume G, Chen B, et al. A General-Purpose Self-Supervised Model for Computational Pathology [Internet]. arXiv; 2023 [cited 2023 Nov 15]. Available from: http://arxiv.org/abs/2308.15474

47. Padegal G, Rao MK, Boggaram Ravishankar OA, Acharya S, Athri P, Srinivasa G. Analysis of RNA-Seq data using self-supervised learning for vital status prediction of colorectal cancer patients. BMC Bioinformatics. 2023 Jun 7;24(1):241.

48. Arik SO, Pfister T. TabNet: Attentive Interpretable Tabular Learning [Internet]. arXiv; 2020 [cited 2024 Jan 22]. Available from: http://arxiv.org/abs/1908.07442

49. Huang Z, Bianchi F, Yuksekgonul M, Montine TJ, Zou J. A visual–language foundation model for pathology image analysis using medical Twitter. Nat Med. 2023 Sep;29(9):2307–16.

50. Hua S, Yan F, Shen T, Zhang X. PathoDuet: Foundation Models for Pathological Slide Analysis of H&E and IHC Stains [Internet]. arXiv; 2023 [cited 2024 Jan 12]. Available from: http://arxiv.org/abs/2312.09894

51. Li H, Yang F, Xing X, Zhao Y, Zhang J, Liu Y, et al. Multi-modal Multi-instance Learning Using Weakly Correlated Histopathological Images and Tabular Clinical Information. In: de Bruijne M, Cattin PC, Cotin S, Padoy N, Speidel S, Zheng Y, et al., editors. Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. Cham: Springer International Publishing; 2021. p. 529–39. (Lecture Notes in Computer Science).

52. Lucas MW, Lijnsvelt J, Pulleman S, Scolyer RA, Menzies AM, Van Akkooi ACJ, et al. The NADINA trial: A multicenter, randomised, phase 3 trial comparing the efficacy of neoadjuvant ipilimumab plus nivolumab with standard adjuvant nivolumab in macroscopic resectable stage III melanoma. J Clin Oncol. 2022 Jun;40(16_suppl):TPS9605–TPS9605.

53. Versluis JM, Menzies AM, Sikorska K, Rozeman EA, Saw RPM, van Houdt WJ, et al. Survival update of neoadjuvant ipilimumab plus nivolumab in macroscopic stage III melanoma in the OpACIN and OpACIN-neo trials. Ann Oncol Off J Eur Soc Med Oncol. 2023 Apr;34(4):420–30.

54. Reijers ILM, Menzies AM, van Akkooi ACJ, Versluis JM, van den Heuvel NMJ, Saw RPM, et al. Personalized response-directed surgery and adjuvant therapy after neoadjuvant ipilimumab and nivolumab in high-risk stage III melanoma: the PRADO trial. Nat Med. 2022 Jun;28(6):1178–88.

55. van Dijk N, Gil-Jimenez A, Silina K, Hendricksen K, Smit LA, de Feijter JM, et al. Preoperative ipilimumab plus nivolumab in locoregionally advanced urothelial cancer: the NABUCCO trial. Nat Med. 2020 Dec;26(12):1839–44.

56. Vos JL, Elbers JBW, Krijgsman O, Traets JJH, Qiao X, van der Leun AM, et al. Neoadjuvant immunotherapy with nivolumab and ipilimumab induces major pathological responses in patients with head and neck squamous cell carcinoma. Nat Commun. 2021 Dec 22;12(1):7348.

57. Schirris Y, Gavves E, Nederlof I, Horlings HM, Teuwen J. DeepSMILE: Contrastive self-supervised pre-training benefits MSI and HRD classification directly from H&E whole-slide images in colorectal and breast cancer. Med Image Anal. 2022 Jul;79:102464.

58. Vorontsov E, Bozkurt A, Casson A, Shaikovski G, Zelechowski M, Liu S, et al. Virchow: A Million-Slide Digital Pathology Foundation Model [Internet]. arXiv; 2023 [cited 2024 Jan 15]. Available from: http://arxiv.org/abs/2309.07778