PathLDM: Text conditioned Latent Diffusion Model for Histopathology

Srikanth Yellapragada * Alexandros Graikos * Prateek Prasanna Tahsin Kurc Joel Saltz

Stony Brook University

Abstract

To achieve high-quality results, diffusion models must be trained on large datasets. This can be notably prohibitive for models in specialized domains, such as computational pathology. Conditioning on labeled data is known to help in data-efficient model training. Therefore, histopathology reports, which are rich in valuable clinical information, are an ideal choice as guidance for a histopathology generative model. In this paper, we introduce PathLDM, the first text-conditioned Latent Diffusion Model tailored for generating high-quality histopathology images. Leveraging the rich contextual information provided by pathology text reports, our approach fuses image and textual data to enhance the generation process. By utilizing GPT’s capabilities to distill and summarize complex text reports, we establish an effective conditioning mechanism. Through strategic conditioning and necessary architectural enhancements, we achieved a SoTA FID score of 7.64 for text-to-image generation on the TCGA-BRCA dataset, significantly outperforming the closest text-conditioned competitor with FID 30.1. 1

1. Introduction

Diffusion models have achieved impressive performance across diverse applications such as image generation, text-to-video synthesis, and audio generation [13,14,22,42]. Recently, Latent Diffusion Models (LDM) [37] have paved the way for higher-resolution text-conditioned image generation by employing a two-step process involving image encoding and diffusion within the latent space.

Unsurprisingly, diffusion models in histopathology image generation are an active area of research. Existing works such as [28,29,48] are constrained by limited dataset size and conditioning capabilities. Our model performs a critical step toward the building of a foundational model, which is to accommodate multiple modalities, including images, text, class labels, etc.

Conditioning on labeled data improves the diffusion model’s generation quality by enabling data-efficient model training. As shown in [31], conditioning can be particularly effective in refining diffusion models, especially when dealing with complex data distributions. Existing histopathology diffusion models condition on class labels, devoid of textual input. Due to the recent advances in Vision Language models [16,27,35], embeddings produced by VLMs are often more expressive than manually picked labels.

We introduce PathLDM, the first Latent Diffusion Model for text-conditioned histopathology image generation. Our novel approach hinges on employing pathology text reports as the conditioning signal. Histopathology data is multi-modal, with pathology reports often accompanying whole slide images. These reports often include detailed descriptions of cell types, disease classification, and other domain-specific insights pathologists provide. They capture critical contextual information that is complementary to the diffusion model’s generation ability.

We consider the extensively studied Cancer Genome Atlas Breast Invasive Carcinoma (TCGA-BRCA) dataset [1,24]. Each entry within this dataset includes Whole Slide Images (WSI) and accompanying text reports provided by pathologists. Given that these reports can be lengthy (with a median length of 2800 characters) and unstructured, we adopt GPT’s [32] capability to distill them into coherent summaries succinctly. These summaries serve as the guiding signal for PathLDM, effectively combining image and textual information. PathLDM produces visually coherent histopathology images aligned with the semantic details embedded within the text summaries through this synthesis.

Starting from the architecture of a Stable Diffusion [37] model, we propose a set of architectural modifications and develop PathLDM. Our technical modifications cover all three components of the LDM: the text encoder, the U-Net [38] denoiser, and the Variational Autoencoder (VAE).

For the conditioning to work, we use a cyclical positional embedding in the text encoder. We pick a VAE that preserves the intricate details, such as cell structures and spa-
tial layout of primitives, found in pathology images. We choose the appropriate U-Net, which acts as the right initialization for our VAE. We demonstrate that our effective conditioning, coupled with the architectural refinements, enable PathLDM to achieve a state-of-the-art Fréchet Inception Distance (FID) score of 7.64 for text-to-image generation on the TCGA-BRCA dataset.

Comparing previous pathology diffusion models with Stable Diffusion, we observed that fine-tuned Stable Diffusion outperforms others. This can be attributed to Stable Diffusion being trained on a larger dataset. By conditioning with the appropriate pathology report text and the necessary architectural changes, we are the first to show that a pathology diffusion model can outperform fine-tuned Stable Diffusion.

Our contributions are as follows:

- We develop the first text-conditioned Latent diffusion model for histopathology image generation.
- We leverage GPT’s capabilities to standardize and summarize pathology text reports for effective conditioning.
- We propose a framework that leverages conditioning with the appropriate architectural elements (cyclical positional embedding, VAE configuration, U-Net initialization) to improve the SoTA FID score dramatically.
- We are the first to demonstrate that a bespoke Pathology diffusion model can perform significantly better than Stable Diffusion.

2. Related Work


Classifier-free diffusion guidance [15] combines unconditional and conditional score estimates to achieve performance akin to Classifier guidance without training an auxiliary classifier. The acceleration of diffusion sampling by 10 to 50 times is proposed in [43] through the Denoising Diffusion Implicit Models (DDIM). Latent Diffusion Models, introduced in [37], bifurcate the diffusion model training into two phases - an autoencoder for low-dimensional representation and a latent space diffusion model, thereby enhancing training efficiency. This approach surpasses previous methods in class conditional ImageNet generation and text conditional MS-COCO [26] synthesis [30]. Our paper is the first to develop a text-conditioned Latent Diffusion Model for histopathology images.

The use of pretrained diffusion models for synthetic data creation has gained attention recently. In [3], the authors employ a text conditional model on a large scale to create synthetic data for ImageNet. Similarly, [10] leverage features extracted from an unconditional diffusion model’s U-Net to provide guidance and enhance synthetic data quality. Both adopt the Classification Accuracy Score [36] as a crucial metric to evaluate the quality of synthetic data.

The inception of CLIP: Contrastive Language Image Pretraining [35] marked a significant advancement in Computer Vision pre-training. Their approach involved jointly training image and text encoders to develop a multi-modal embedding space. For text conditional diffusion models such as Stable Diffusion [37], CLIP models were employed to encode text captions.

There have been recent advancements in Vision Language Models (VLMs) in the field of pathology. For instance, MI-Zero [27] leverages a dataset comprising over 33k image-text pairs to train their VLM. Similarly, PLIP [16] gathers pathology data from Twitter to train a CLIP model. The authors of BioMedCLIP [49] introduce domain-specific adaptations tailored to biomedical VLP to outperform prior VLP approaches. However, none of these methods have utilized the pathology image-text pairs to train diffusion models.

The application of diffusion models in histopathology remains limited. The authors of [28] train a pixel-level diffusion model for histopathology images, targeting 33,777 patches from low-grade gliomas within the TCGA dataset [11]. Similarly, [48] explore a Diffusion Autoencoder [34] trained on the NCT-CRC-HE [19] and PCam [45] datasets. Medfusion, introduced in [29], is a Latent diffusion trained on 19,558 colorectal cancer images from CRCDX [18].

Both [28] and [29] focus on training class conditional models. Meanwhile, [48] opts for separate unconditional models per class, which is less efficient than a shared class conditional model. It’s worth noting that these approaches employ relatively small datasets and lack a text conditioning mechanism. Our PathLDM cures both of these deficiencies.

3. Methodology

In this section, we describe our approach to conditioning and the necessary modifications applied to Latent diffusion models for the histopathology setting.

3.1. GPT Summarization

We leverage the TCGA-BRCA dataset for our model training. Each entry within this dataset includes a Whole
The breast pathology report indicates the following information:

1. Histological classification: The patient has Luminal A subtype breast cancer. The biopsy of the left axilla sentinel lymph node shows metastatic carcinoma following breast carcinoma.

2. Subtype: The invasive carcinoma is of the lobular type, poorly differentiated and multifocal, with extensive lymphatic permeation. Lobular neoplasia, type B (pleomorphic), is widespread. The report also mentions fibrocystic disease, proliferative, with atypia.

3. Necrosis: No mention of necrosis.

4. Tumor infiltrating lymphocytes: The report mentions extensive lymphatic permeation but does not mention tumor infiltrating lymphocytes.

Please generate a succinct report in 75 words from the above information. Exclude any filler words or sentences if something is not mentioned or specified, exclude it from the report. Low verbosity.

Figure 1. Example of our GPT-3.5 usage where we employ the chain-of-thought prompting technique to generate summaries. This approach leverages GPT’s capability to progressively develop a coherent line of reasoning, yielding a more refined and informative summary.

The core of our methodology revolves around the development of PathLDM, a Latent Diffusion Model tailored for histopathology image generation with text conditioning. PathLDM’s architecture consists of the Variational Autoencoder (VAE), the U-Net denoiser and the text encoder. We provide an overview of our method in Figure 2.
Figure 2. Overview of PathLDM. We start with a WSI and an accompanying Pathology report. We crop the WSI into 256 × 256 patches at 10x resolution. Leveraging GPT, we summarize the pathology report. For each patch, we compute Tumor and TIL probabilities, fuse them with the slide-level summary, and condition the LDM with CLIP embeddings of the fused summary. The VAE and text encoder remain frozen, and only the U-Net is trained.

3.2.1 Variational Autoencoder

The VAE learns a latent representation from the input images. A significant criterion for evaluating the VAE’s performance is its ability to reconstruct the original image through the decoder. Accurate reconstructions are crucial, as the subsequent diffusion process operates on these latent vectors and can, therefore, be constrained by the VAE’s capabilities.

Reconstruction quality becomes paramount when dealing with histopathology images that contain fine details such as cells. The authors of Medfusion [29] observed that using off-the-shelf Stable Diffusion’s VAE can result in imperfect reconstructions. They modified the VAE architecture to overcome this issue by adjusting the number of channels and Convolution layer parameters. This refinement led to an increase in the Structural Similarity Index (SSIM) [46] of the reconstructed images.

However, both Stable Diffusion and Medfusion’s VAE employ a downsampling factor of 8, effectively compressing 256 × 256 images down to a 32 × 32 × 4 (or 8 channels for Medfusion) latent. Opting for an overly small downsampling factor is undesirable, as it leads to a considerable increase in memory consumption and substantially hampers the training speed of the LDM. Thus, finding the right balance in compression levels is crucial for accurate reconstructions.

In PathLDM, we instead choose a 3-channel latent VAE with a downsampling factor of 4. In Section 4.3, we delve into the impact of the downsampling factor on VAE’s reconstruction quality when fine-tuned on Pathology images. We experimentally validate the advantages of our latent size choice by showing how we can maintain the fine structures found in large images.

3.2.2 U-Net Conditioning

Leveraging GPT, we procure a single summary for each WSI. We then incorporate patch-level statistics corresponding to Tumor and Tumor-Infiltrating Lymphocyte (TIL) data to enrich these summaries. TILs have emerged as a promising biomarker in solid tumors, as a prognostic marker in triple-negative (TNBC) and HER2-positive breast cancer [2, 7, 40]. We use the trained models from [1] and [24] to obtain a Tumor and TIL probability value for each patch. These probability values were integrated into the text captions by first transforming them into an ordinal scale of Low - High and prepending them to the text description of each
patch. This process created captions like “High tumor; low til; {GPT summary}”, effectively combining quantitative details with the summarization outcomes.

Our approach effectively fuses conditioning information derived from three distinct sources. By integrating the GPT summary, which encapsulates Whole-slide level conditioning, with Tumor and TIL statistics that represent patch-level information, we create an advantageous synthesis of global and local details. In contrast, existing Pathology diffusion models [28, 29, 48] exclusively rely on class conditioning mechanisms, devoid of textual input. We delve into the impact of our conditioning mechanism in Section 4.5.

3.2.3 Encoding large prompts

We use a CLIP [35] text encoder to process the text summaries obtained through GPT fused with Tumor and TIL presence. Its role is to transform these summaries into embeddings, compact yet rich textual context representations. Existing CLIP [16, 35] models are limited by a maximum context window length of 77 tokens. In our approach, we accommodate GPT summaries with a length of up to 154 tokens. This choice is deliberate, as it is twice the context length of the CLIP encoder. Opting for a length of 77 tokens would have resulted in an overly concise summary lacking in informative context.

To embed the full token sequence, we individually embed each of the halves and concatenate. This is equivalent to using a cyclical positional embedding that is repeated every 77 tokens, which we found to be sufficient for embedding the longer text sequences. We apply a causal mask and direct the full embedding sequence through the trained transformer encoder for processing. As an ablation, we replace the standard OpenAI CLIP [35] encoder with the more specialized PLIP encoder, discussed further in Section 4.6.

4. Experiments

In this section, we delve into the experimentation conducted to validate the effectiveness of our proposed PathLDM. We introduce architectural improvements to the Variational Autoencoder, U-Net denoiser, and text encoder components and explain how they affect the Fréchet Inception Distance (FID) metric.

We train all our diffusion models on 3 NVIDIA RTX 8000 GPUs with a batch size of 48 per GPU. We use the training code and checkpoints provided by [37], implemented in Pytorch [33]. Our training employs a learning rate of $2 \times 10^{-5}$ with 10,000 warmup steps. For FID computation, we use the implementation from [41].

We adopt DDIM sampling [43] with 50 steps and incorporate classifier-free guidance [15] at a scale of 1.75 to generate synthetic images. We randomly sample classes for the class conditional models, and for the text conditional models, we randomly select summaries from the training set for conditioning.

4.1. Dataset

We selected 1136 Whole slide images from the TCGA-BRCA dataset and partitioned them into 977 for training and 199 for testing. Using the code from [25], we extract $256 \times 256$ patches at 10x magnification, culminating in 3.2 million training patches. Leveraging GPT, we procure a single summary for each WSI.

We use the trained models from [1] and [24] to obtain a Tumor and TIL probability value for each patch and incorporate them to enrich these summaries. This process created captions like “High tumor; low til; {GPT summary}”, effectively combining quantitative details with the summarization outcomes.

4.2. Baseline

To assess the impact of different architectural changes on the FID, we begin with a class conditional LDM baseline. We exclusively utilize patch-level Tumor and TIL probabilities in this baseline without incorporating GPT’s text summary. We first convert the probability values into an ordinal scale using a threshold of 0.5, denoted as “Low” and “High.” Subsequently, each patch is assigned a class label from a range of values: 0 representing “Low Tumor + Low TIL”, 1 for “Low Tumor + High TIL”, 2 for “High Tumor + Low TIL”, and 3 for “High Tumor + High TIL.”

4.3. VAE

To quantify the influence of VAE on the diffusion model’s generation quality, we train two class-conditional LDMs, leveraging patch-level Tumor and TIL data. One model uses a VAE with a downsampling factor of 8, denoted as “vq-f8”, while the other employs a VAE with a downsampling factor of 4, denoted as “vq-f4”. As exhibited in Table 1, FID shows noteworthy enhancement, improving from 48.14 to 39.72. This improvement underlines the substantial impact of VAE selection on the diffusion model’s generative performance.

From the pretrained VAEs provided by [37], we start with VAE VQ-f4 with a downsampling factor of 4. We then fine-tune it on our image patches. In Table 2, we highlight the significance of the downsampling factor or latent vector dimension in the quality of VAE reconstructions. Our choice of VAE yielded a significant boost in the SSIM of the reconstructed images, suggesting that overly aggressive compression may hinder the retrieval of intricate details, particularly in images featuring small cellular structures.

Our observations underscore the vital role of VAE selection, even in other large-image settings such as satellite imagery. The Deepglobe Landcover dataset [6] comprises
4.4. U-Net

Moving forward, our next focus is the U-Net denoiser. To gauge the advantage of fine-tuning, we train two class conditional LDMs. In the first model, we initialize the U-Net denoiser weights from scratch. Second, we start with the ImageNet [39] class conditioned U-Net denoiser weights from [37]. The outcome presented in Table 1 shows that employing a fine-tuned U-Net enhances the FID from 39.72 to 29.45. Intriguingly, even though the ImageNet weights are tailored to Natural images - distinct from the complexity of histopathology images - the learned denoising mechanism within the U-Net contributes significantly to the efficacy of our PathLDM training.

4.5. Text conditioning

To measure the concrete effects of our VAE and U-Net changes, we initially established a class conditional LDM baseline. Now, we aim to assess the benefit of introducing text summaries. Leveraging GPT, we obtain a summary for each Whole slide image (WSI). These summaries are combined with patch-level Tumor and TIL classes and employed to condition the diffusion model.

The results, as outlined in Table 1, are impressive: the combination of text reports and patch-level statistics significantly reduces the FID score from 29.45 to 10.64. This improvement outperforms the gains achieved solely through architectural enhancements. Incorporating context from text summaries introduces a powerful interplay between visual and textual data, substantially enhancing model performance.

4.6. Text encoder

Stable Diffusion uses the OpenAI CLIP [35] encoder, trained on Natural image caption pairs. As an ablation experiment, we introduce PLIP [16], a CLIP model trained on pairs of pathology images and their corresponding text. As shown in Table 1, adopting the PLIP encoder improved the FID from 10.64 to 7.64. This improvement is attributed to the PLIP model’s more adept grasp of the nuances present in pathology-specific text summaries.

4.7. Comparison with other methods

The cumulative architectural enhancements detailed in the preceding subsections result in the creation of our most potent Latent Diffusion Model - PathLDM. We compare our method against three alternative methods - Moghadam et al. [28], Medfusion [29] and Stable Diffusion [37].

For Moghadam et al. and Medfusion, both class conditional diffusion models, we train them with Tumor and TIL classes. Similarly, we train Stable Diffusion using text summaries, akin to our approach for PathLDM. We initialize the U-Net denoiser of Stable Diffusion and PathLDM.
Figure 3. We choose a single text report and produce synthetic samples using Medfusion [29], Stable Diffusion [37], and PathLDM. Samples generated by Medfusion and Stable Diffusion show artifacts (indicated by green boxes) that are not present in our outputs. Moghadam et al. [28] produces images in lower resolution ($128 \times 128$) and exhibits artifacts and blurriness (green boxes in the last row) in the output images. The first row contains the real samples from the corresponding WSI.

<table>
<thead>
<tr>
<th>Method</th>
<th>Conditioning type</th>
<th>FID ↓</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moghadam et.al [28]</td>
<td>Tumor + TIL</td>
<td>105.81</td>
</tr>
<tr>
<td>Medfusion [29]</td>
<td>Tumor + TIL</td>
<td>46.49</td>
</tr>
<tr>
<td></td>
<td>Report + Tumor + TIL</td>
<td>39.56</td>
</tr>
<tr>
<td>Stable Diffusion [37]</td>
<td>Report + Tumor + TIL</td>
<td>30.1</td>
</tr>
<tr>
<td>PathLDM</td>
<td>Tumor + TIL</td>
<td>29.45</td>
</tr>
<tr>
<td></td>
<td>Report + Tumor + TIL</td>
<td>7.64</td>
</tr>
</tbody>
</table>

Table 4. Comparison of FID using 10k synthetic images. PathLDM outperforms other methods by a huge margin. Training Medfusion with text-conditioning improved the FID, highlighting the efficacy of our text summaries.

Using pretrained weights from [37]. However, Medfusion employs an 8-channel VAE, and no existing pretrained U-Net for Stable Diffusion has done more than four channels, which makes them unsuitable for initialization. Hence, we train it from scratch.

The results, as shown in Table 4, demonstrate the superiority of our text conditional PathLDM over other methods, establishing a significant FID gap. Furthermore, we also conducted text-conditional training for the Medfusion architecture, leading to an FID improvement from 46.49 to 39.56, highlighting the efficacy of our text summaries. Notably, even our class conditional model, trained without text reports, outperforms the class conditional variants of both Moghadam et al. and Medfusion, owing to the potency of the architectural enhancements we introduced.

Pixel-level diffusion models are computationally demanding, and no straightforward method exists to condition them on text captions. Recent works like GLIDE [30] and Imagen [13] are typically constrained to smaller image sizes ($64 \times 64$) and rely on upsampling models to achieve high-resolution image synthesis. Since Moghadam et al. utilizes a pixel-level diffusion model, we haven’t trained a text-conditional variant.

For visualization, we select a text report and generate synthetic samples by employing Medfusion, Stable Diffusion, and PathLDM. Four samples each, categorized as “low tumor” and “high tumor,” are generated for evaluation. The figures in Figure 3 highlight the diversity and authenticity of the samples produced by our approach. Notably, our method generates samples that closely adhere to the distribution of actual samples. On the other hand, samples generated by Medfusion and Stable Diffusion show artifacts, such as blue stains, that are absent in our outputs.
Table 5. Classification Accuracy scores of a ResNet-34 on synthetic data. Higher CAS proves the effectiveness of PathLDM-generated synthetic samples in capturing significant features.

<table>
<thead>
<tr>
<th>Data type</th>
<th>Data source</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real</td>
<td>TCGA-BRCA</td>
<td>90.31</td>
</tr>
<tr>
<td>Only Synthetic</td>
<td>Medfusion</td>
<td>79.12</td>
</tr>
<tr>
<td>Only Synthetic</td>
<td>PathLDM</td>
<td>83.32</td>
</tr>
<tr>
<td>Real + Synthetic</td>
<td>PathLDM</td>
<td>90.74</td>
</tr>
</tbody>
</table>

Table 6. Ablation to highlight the benefit of text summaries.

<table>
<thead>
<tr>
<th>PathLDM Variant</th>
<th>FID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline - True summary</td>
<td>7.64</td>
</tr>
<tr>
<td>Random summary</td>
<td>22.16</td>
</tr>
<tr>
<td>5 summaries per WSI</td>
<td>7.34</td>
</tr>
</tbody>
</table>

4.8. Classification Accuracy Score

In addition to the widely used FID for evaluating generative models, an alternative evaluation metric is the Classification Accuracy Score (CAS) [36]. To compute CAS, a classifier is trained on synthetic data and evaluated on real validation data. To train PathLDM, we use a dataset of 3 million patches, each associated with a Tumor label of either “Low” or “High.” We create synthetic datasets using PathLDM and Medfusion, generating 3 million patches. We assign the Tumor label used in generating each patch as its ground truth.

We then train ImageNet pretrained ResNet-34 [12] classifiers on these synthetic datasets for 40 epochs with a batch size 256 and the Adam optimizer [21]. We use a learning rate of $10^{-3}$ for the first 20 epochs, then a reduction to $\frac{1}{10}$ at epochs 20 and 30. As presented in Table 5, the classifier trained on our synthetic data achieves a validation accuracy of 83.32%, significantly higher than a classifier trained on Medfusion’s data (79.12%). When the classifier is trained using a mix of real and synthetic data, there is a modest improvement in accuracy, increasing from 90.31% to 90.74%. This marginal enhancement suggests that the task of tumor classification is inherently less complex.

5. Ablation studies

5.1. Sampling parameters

We adopt DDIM sampling with 50 steps and integrate classifier-free guidance. We conduct a parameter sweep across guidance values ranging from 0.1 to 3, resulting in the optimal FID outcome at 1.75. This value is then employed consistently across all our experimental scenarios.

5.2. Benefit of text reports

We perform two ablation experiments to underscore the significance of text report conditioning. In the first ablation, we train an LDM where, for each patch, we substitute its Whole Slide Image (WSI) summary with a randomly chosen summary from a different WSI. In the second ablation, we extend the utility of GPT by generating five distinct summaries from each text report obtained through five separate GPT API queries. Each of these queries operates independently and adheres to the same token limitations. Subsequently, for each patch, we randomly select only one summary from the five provided by GPT.

The outcomes, as shown in Table 6, provide valuable insights - substituting a random summary leads to the deterioration of the diffusion model, increasing the FID from 7.64 to 22.16. This outcome demonstrates that using an inappropriate summary or the absence of one altogether worsens the generation quality. Conversely, deploying five summaries per WSI improved the FID from 7.64 to 7.34. These ablations underline the crucial role that accurate text conditioning plays in enhancing the performance of our model.

6. Limitation

A limitation of our approach is the reliance on text prompts from the training set for generating synthetic images. We did not employ held-out reports for image generation, which could be a potential avenue for future research.

7. Conclusion

Our paper presents PathLDM, a groundbreaking Latent Diffusion Model designed for generating histopathology images conditioned on text. By effectively summarizing complex pathology text reports using GPT, we establish a robust conditioning mechanism that bridges the gap between textual and visual content. Our strategic conditioning, coupled with architectural enhancements, led to a SoTA FID score of 7.76 on the TCGA-BRCA dataset, significantly outperforming the closest competitor with FID 30.1. This study performs a critical step toward the building of a foundational model, which is accommodating multiple modalities, including images, text, class labels, etc. We anticipate that our work will encourage future exploration of domain-specific pathology diffusion models.

Acknowledgements

The reported research was partially supported by NCI awards 5U24CA215109, 1R21CA258493-01A1, NSF grants IIS-2123920, IIS-2212046 and Stony Brook Profund 2022 seed funding.
References


[19] Jakob Nikolas Kather, Niels Halama, and Alexander Marx. 100,000 histological images of human colorectal cancer and healthy tissue, Apr 2018. 2


