His-MMDM: Multi-domain and Multi-omics Translation of Histopathology Images with Diffusion Models

Zhongxiao Li1,2,#, Tianqi Su3,#, Bin Zhang1,2, Wenkai Han1,2, Sibin Zhang3, Guiyin Sun3, Yuwei Cong4, Xin Chen5, Jiping Qi4, Yujie Wang4, Shiguang Zhao5,*, Hongxue Meng6,*, Peng Liang3,*, and Xin Gao1,2,*

1Computer Science Program, Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

2Center of Excellence on Smart Health, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

3Department of Neurosurgery, Harbin Medical University Cancer Hospital, Harbin, China

4Department of Pathology, The First Affiliated Hospital of Harbin Medical University, Harbin, China

5Department of Neurosurgery, The First Affiliated Hospital of Harbin Medical University, Harbin, China

6Department of Pathology, Harbin Medical University Cancer Hospital, Harbin, China

#Equal Contribution

*All correspondence should be addressed to Shiguang Zhao (E-mail: guangsz@hotmail.com), Hongxue Meng (E-mail: menghongxue@hrbmu.edu.cn), Peng Liang (E-mail: liangpeng@hrbmu.edu.cn) and Xin Gao (E-mail: xin.gao@kaust.edu.sa, Tel: +966-12-8080323)

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract
Generative AI (GenAI) has advanced computational pathology through various image translation models. These models synthesize histopathological images from existing ones, facilitating tasks such as color normalization and virtual staining. Current models, while effective, are mostly dedicated to specific source-target domain pairs and lack scalability for multi-domain translations. Here we introduce His-MMDM, a diffusion model-based framework enabling multi-domain and multi-omics histopathological image translation. His-MMDM can translate images across an unlimited number of categorical domains, enabling new applications like the translation of tumor images across various tumor types, while performing comparably to dedicated models on previous tasks such as transforming cryosectioned images to formalin-fixed paraffin-embedded (FFPE) ones. Additionally, it can perform genomics- and/or transcriptomics-guided editing of histopathological images, illustrating the impact of driver mutations and oncogenic pathway alterations on tissue histopathology. These versatile capabilities position His-MMDM as a versatile tool in the GenAI toolkit for future pathologists.

Keywords
Image Generative Models, Image Translation, Diffusion Models, Histopathology
Introduction

Generative artificial intelligence (GenAI) in imagery has marked a groundbreaking advance in the field of artificial intelligence (AI), offering transformative capabilities across a wide range of applications. At the forefront, the state-of-the-art text-to-image models such as Stable Diffusion, Midjourney, and DALL-E, transform text descriptions into vivid visual representations. Conversely, image-to-text models, such as GPT-4V, interpret and convert visual data into descriptive text, enabling open-ended dialogs about the image content with users. There are also image-to-image translation models that use existing images as templates to generate new ones, enabling applications such as image editing, resolution enhancement, and sketch-to-image synthesis. Collectively, these models are revolutionizing the way we create, interpret, and interact with visual contents.

Such a revolution is also gradually taking place in the field of computational pathology. Early explorations have demonstrated the effectiveness of using generative adversarial networks (GANs) or diffusion models (DMs) to generate synthetic histopathological images, and have been mainly used as a data augmentation strategy. Recently, a lot of efforts have been devoted to improving the quality and enlarging the resolution of the generated images. Concurrently, visual language foundation models and copilot systems are also being developed in computational pathology to automate medical visual question answering. As for image-to-image translation, dedicated models have been developed for various common tasks such as color normalization, stain transfer, virtual staining, and transforming cryo-sectioned or stimulated Raman images to formalin-fixed paraffin-embedded (FFPE) ones.

In general, there are two types of image-to-image translation tasks, including the paired and unpaired ones. The paired image translation task requires matched source and target domain images during training, which is usually not readily available. In contrast, the unpaired image translation task requires the algorithm to learn the mapping between a source and target domain on itself, making it more useful in real-world applications. It is thus not surprising that almost all the existing image translation models in computational pathology are un-paired. Although the abovementioned image translation models serve as important computational pathology applications, their full potential still has not yet been unleashed. Firstly, current applications are mostly limited to GANs, whereas recent developments of DMs show their superior performance on conditional generation as compared to GANs. Secondly, the current GAN-based histopathological image translation models are limited to the translation between a couple of specific source-target domain pairs and their extension to the translation between multiple...
domains is not straightforward. Hence, simultaneous translation across multiple domains has a wider range of applications in computational pathology. For example, one can edit one real histopathological image into multiple synthesized versions corresponding to altered genomic or transcriptomic profiles, which can potentially model the impact of mutations and altered gene expression on the histopathological appearance of the tissue. Furthermore, multiple-domain translations may also alleviate the inefficiency of pairwise translation, reducing the number of $n(n-1)/2$ individual pairwise models to one multi-domain model. In addition, compared to pairwise models, multi-domain models may learn invariant knowledge shared across all the domains. Thirdly, current GAN-based models have specifically designed components or loss functions that may hinder their extensibility to new applications.

Due to the above limitations, in this study, we propose a novel versatile framework, Histopathological image Multi-domain Multi-omics translation with Diffusion Models (His-MMDM). His-MMDM achieves image translation by diffusing the source domain images into noisy images distributed as Gaussian distribution and then denoising the noisy images back into target domain images. Compared with previous image translation models in histopathology, His-MMDM stands out with two unique capabilities: (1) it can be efficiently trained to translate images between an unlimited number of categorial domains (through class-conditional inputs to the model) (2) it can perform genomics- or transcriptomics- guided editing of histopathological images (through multi-omics inputs to the model). For categorical domain translation, we demonstrated His-MMDM’s performance through four comprehensive tasks using four independent datasets, including primary tumor type translation, tumor organ site translation, translation of cryo-sectioned slide images to FFPE ones, and virtual immunohistochemical (IHC) staining of hematoxylin and eosin (H&E) images. We found that, without additional architectural enhancement, His-MMDM is on par with state-of-the-art dedicated models over all four tasks. Using the histopathological images in The Cancer Genome Atlas (TCGA) that are equipped with the matched genomic and transcriptomic profiles, we demonstrated His-MMDM’s effectiveness in performing genomics- and transcriptomics- guided editing of the images. We then systematically discussed and illustrated through notable examples the effect of high-frequency somatic mutations and alterations on the expression level of genes/pathways on the histopathology of the tissue from a generative model’s perspective. Overall, we believe that His-MMDM showcases the possibility of generic multi-domain and multi-omic translation and generation in histopathology, provides a powerful tool for pathologists, and could serve as an integral component of GenAI copilot systems in the future.
Results

The design of His-MMDM model

His-MMDM is trained based on the principle of diffusion models (DMs)\(^7\) and is trained with histopathological images. A DM is comprised of a forward diffusion and a backward denoising process (Fig. 1). In the forward diffusion procedure, a histopathological image \(X_T\) is gradually diffused into a noisy image \(X_0\) following a Gaussian distribution using a pre-determined schedule and in the backward denoising procedure, the noisy image is gradually denoised back into a histopathological image (Methods). At each iteration of the denoising procedure, a U-net-based\(^7,30\) network is used to predict noise from the image at the current iteration step, which is utilized in the subsequent denoising step derived from the inversion of the forward procedure (Methods)\(^3^1\). Conditional generation is achieved by the addition of condition information to the U-net through embedding vectors added to each layer of the U-net blocks (Fig. 1, Table S1, Methods). The DM used in His-MMDM can handle three types of condition information: (1) categorical labels, such as tumor types, primary sites, and IHC stains, (2) genomics, which include the mutation status of the corresponding sample, and (3) transcriptomics, which include the transcriptomic profile of the sample. To train the U-net for noise prediction, a forward diffusion procedure is performed on the images from the training dataset of histopathological images and executed to various intermediate stages (Methods). The parameters of the U-net are updated to predict the noise added in the forward diffusion procedure when the corresponding condition information of the sample is provided to the U-net. After the conditional U-net is trained, His-MMDM achieves image translation by (1) executing the forward diffusion procedure of the trained DM by conditioning the U-net with the source domain condition information and then (2) executing the backward denoising procedure by conditioning the U-net with the target domain condition information. The justification and theoretical guarantees of bridging together the forward and backward processes of DMs are briefly discussed in Methods and by Su et al.\(^2^9\).

We demonstrate in the following sections that such a strategy is beneficial in achieving multi-domain and multi-omics image translation and has notable GenAI applications in computational pathology.

Translation of histopathological images across primary tumor types

We first tested His-MMDM’s ability to translate across categorical image domains. The categorical domains for the first task are 19 primary tumor types of The Cancer Genome Atlas (TCGA) (Methods). These 19 tumor types can be classified into five categories, including five gastrointestinal (GI) (ESCA, STAD, COAD, READ, and PAAD), two lung (LUAD and LUSC), three kidney (KIRP, KICH, and KIRC), three pan-gynecological (GYN) (OV, BRCA, and
UCEC), and five other tumor types (LIHC, HNSC, SARC, THCA, BLCA, and PRAD) (Methods, statistics shown in Table S2, abbreviations defined in Table S3). Using discriminative models, previous studies demonstrated high cross-classification performance on a lot of these tumor type pairs, i.e., the performance of a tumor classifier trained on one tumor type and tested on another, which suggests conserved tumor image features across different types of tumors.

We tasked His-MMDM to translate the histopathology images between each two across these 19 primary tumor types (Fig. 2a). After training the model on 1.7 million image patches from the 19 tumor types, we translated the 19,456 images (1,024 in each tumor type) in the test set between each pair of the tumor types. As a comparison, we also trained two GAN-based models, CycleGAN and CUT. The Frechet Inception Distance (FID) of the translated images to a particular tumor type and the real images of that tumor type were computed as the evaluation of image fidelity (the lower FID scores imply higher fidelity) (Methods). Overall, His-MMDM achieved similar or even better performance (lower FID scores) compared to its GAN-based counterparts, especially for tumor types such as BLCA, KIRP, and LUSC (Fig. 2b). We further computed the FID scores of the real un-translated images between tumor types (Fig. S1a) and computed the reduction of this metric (ΔFID = FID_{before} − FID_{after}) before and after translation by His-MMDM (Fig. 2c, Methods). The largest ΔFIDs are observed between very distant tumor types, such as between the GYN tumors and kidney tumors. In particular, due to the distinctiveness of kidney and liver tumors, their ΔFID values are the highest when other tumors are either translated to or from them (Fig. 2c). Previous cross-classification studies using discriminative models reported high performance between related tumor types, such as within the GI tumors or the lung tumors. Consistently, we also observed that the ΔFID values tend to be the lowest within the GI tumors and relatively low within the GYN tumors, kidney, and lung tumors.

The translated images generally keep the location and arrangement of the tissue contents in the original images, but they are usually modified to have specific characteristics of the target tumor type. As an example, colorectal adenocarcinoma (COAD) usually features the glandular structures formed by the epithelial cells (Fig. 2d). When the colorectal images are translated to other adenocarcinomas (such as lung (LUAD), gastric (STAD), and pancreatic adenocarcinoma (PAAD)), those structures are kept intact (Fig. 2d). However, when they are translated to non-adenocarcinomas, those structures are attenuated as much as possible and replaced with some target type-specific contents (Fig. 2d). From another perspective, the effect of translation by His-MMDM is also evident from the space of image features. Specifically, the Inception V3 features of the translated images of the five tumor types from each tumor category (LUAD,
ESCA, OV, LIHC, and KICH) are distributed much more evenly compared to their un-translated counterparts (Fig. 2e and Fig. S1b-S1e).

Translation of images from one tumor type to another enables several applications. For example, suppose we have a well-trained discriminative model for tumor type A and we would like to apply it to another tumor type B where such a model is unavailable. Instead of directly applying it to the images from tumor type B, which is the cross-classification strategy and a few/zero-shot learning task as in Noorbakhsh et al. 32, we could first adapt the images into type A and then use the model to classify or further fine-tune the model on the adapted images. We compared the performance (in terms of F1 score) of binary tumor classification models trained on one tumor type (the target of translation) and cross-classifying images from another tumor type (the source of translation/the goal of classification) using either the original images or the translated images. The classification performance in four out of five tumor families corresponding to 12 out of 19 tumor types on average showed higher performance using the translated images (Fig. 2f and Fig. S2a, ‘ratio for fine-tuning = 0’). Additionally, to improve the classification models’ performance, we fine-tuned the classification models using 30%, 60%, and 90% of the translated images (Fig. 2f and Fig. S2a, ‘ratio for fine-tuning = 0.3, 0.6, 0.9’). The number of tumor types with higher performance using the translated images increases to 13 at 60% and 15 at 90%. The performance improvement is most evident when translating from the GI tumors to the lung tumors, and the improvement is more in recall than in precision (Fig. S2b). This improvement in recall could be due to the adaptation of the features of the unknown-type image into known ones. Therefore, translating images from unseen tumor types to known tumor types is indeed beneficial for a discriminative model’s inference and further fine-tuning.

Inference of tumor of unknown primary (TUP) has recently gained a lot of attention due to the strong performance of machine learning models on such tasks 37,38. We demonstrated that the translation between different primary tumor types enabled by His-MMDM is useful for the interpretation of such decisions from a generative model’s perspective and delegate the discussions to the Supplementary Materials (Supplementary Notes Section S1, Fig. S3a-S3b).

Genomics and transcriptomics-guided editing of histopathological images

Somatic mutations in oncogenes and tumor suppressors are the driver factors in tumorigenesis. Previous studies have reported strong associations of driver mutations with histopathological image features, which were demonstrated through the predictive power of discriminative models 39-45. More recently, pan-cancer studies have shown the identifiability of common driver mutations (e.g. TP53 and PTEN) shared in multiple tumor types using deep transfer learning 46.
Apart from genomic mutations, several other works have been dedicated to the prediction of transcriptomic expression levels from histopathological images, achieving various levels of success.\cite{46,47}

We explored the capability of His-MMDM to decipher these associations but from a generative model’s perspective. To this end, we trained another His-MMDM model from the TCGA histopathological images conditioned simultaneously on the genomic mutation profiles and the transcriptomic profiles of the corresponding samples (Methods). In total, the genomic embeddings of His-MMDM take into account the 334 genes from the ten common oncological signaling pathways,\cite{48} and 188 other genes with high pan-cancer mutation rate or discriminative performance\cite{46} in the TCGA cohort (Table S4). Similarly, the transcriptomic embeddings include 4,193 genes that were included in the collection of 50 hallmark pathways of MSigDB (v2023.2)\cite{49,50}, as well as an additional 268 genes that either displayed the greatest variability between tumor and normal samples or have previously reported discriminative performance\cite{47} (Table S4).

We first performed in silico genomic modification experiments using the trained His-MMDM model. For a given histopathological image, the forward diffusion process is executed upon it conditioned on its original genomic mutation profile, e.g., gene X is wide-type (WT) and gene Y is mutated, and the backward denoising process is then executed conditioned on a modified genomic profile, e.g., gene X is mutated and gene Y is WT (Fig. 3a). In this way, we are essentially asking the model to edit the given histopathological image into a version if its gene X and gene Y’s mutation statuses were as designated.

We first investigated the effect of single mutations among the top 30 genes with the highest mutation rates across the aforementioned 19 tumor types on the appearance of histopathological images. For each image, we used His-MMDM to produce one version of it when all genes are WT. Then we produced a series of other images when each one of the genes is mutated. The mutation of different genes resulted in modifications in different aspects of the histopathological image. For example, in COAD, the mutation of \textit{APC} had the most significant effect on the nuclear shape of the cell, \textit{TP53} on the cellular density, and \textit{SMAD4} on the background and microenvironment (Fig. S4a). To measure the visual effect of the mutation, we computed the cosine distance between the WT image and each ‘mutated image’. Specifically, for each gene in a specific tumor type, the cosine distances of the Inception V3 features of the two versions are averaged and normalized both within- and cross-tumor (Fig. 3b). Some tumor types show the greatest overall effect of single mutations on their visual appearance, such as the kidney tumors (KICH, KIRC, and KIRP). While some other tumor types showed greater effect within their
tumor categories, such as COAD and READ among the GI tumors and OV in GYN tumors (Fig.
3b). Mutations in HIPPO, NOTCH, and RTK-RAS pathways tend to have the greatest pan-cancer
effect on histopathology. Interestingly, these genes in NOTCH and RTK-RAS are NCOR2 and
ERBB4, but are not NOTCH1 and KRAS themselves (Fig. 3b). The mutations of genes have a
greater effect on histopathology in tumor types where there were previously reported associations
between them based on discriminative performance reported by Fu et al. 46 (all gene-tumor pairs:
p = 6.7e-05; p = 0.03 only high-frequency mutations (>10%). We applied the cell detection
algorithm, Hover-Net 51, on the generated images of the genes which have high mutation rates
across multiple tumors. Most notably, the generated images corresponding to the mutated version
of APC, PTEN, TP53, and CDKN2A have a higher number of malignant cells compared to their
WT counterparts in COAD, UCEC, OV, and ESCA, respectively (Fig. 3c and 3d).

Apart from single mutations, we investigated the effect of accumulating mutations on the
generated histopathological images by His-MMDM. For a total of 331 genes in nine oncological
pathways (one pathway excluded due to its low number of genes), we sorted the genes in each of
them according to their mutation rate (from high to low, Table S4) and edited the images by
sequentially performing the mutation of 25%, 50%, 75%, and 100% genes in each pathway. We
then compared the cosine similarity of Inception V3 features of each of them to the WT version
of the image (Fig. S4b). Some pathways, such as Cell Cycle and WNT signaling pathways in
COAD, displayed continuous changes as the mutation accumulates (Fig. S4b, S4c). But more
frequently, a saturation effect was observed after the most significant change had already resulted
from the initial 25% of the mutations (Fig. S4b).

Using a CRISPR/Cas9 mouse model, a recent study 52 has confirmed the sufficiency of loss-of-
function mutations of five tumor suppressor genes (PTEN, TRP53 (mouse homolog of the human
TP53), RB1, STK11, and RNaseL) to induce prostate tumor progression and three additional
epigenetic factors (KMT2C, KMT2D, and ZBTB16) to initiate tumor metastasis (Fig. 3e). Among
them, only RNaseL and ZBTB16 are not included in the His-MMDM genomic embeddings. We
therefore attempted to edit the PRAD images in the TCGA cohort according to the mutation of
the remaining 7 genes. For each image, we used His-MMDM to produce one version free of
mutations in all tumor suppressors and epigenetic factors (WT), and several mutant versions for
(1) individual mutations in the tumor suppressors and epigenetic factors, (2) combined mutations
in all tumor suppressors, (3) combined mutations in all epigenetic factors, and (4) combined
mutations in all tumor suppressors + epigenetic factors. We then computed the cosine distances of
the Inception V3 features of the mutated images versus the WT images (Fig. 3f). Among the
tumor suppressors, \textit{STK11} induces the greatest histopathological changes in the tissue histopathology (Fig. 3f). The tumor suppressors generally had higher impact than the epigenetic factors, which is consistent the previous study that it is these factors that mainly contributed to the \textit{in situ} tumorigenesis of prostate tumor. The advantages of His-MMDM enabled \textit{in silico} interpolation of histopathological images corresponding to different mutation statuses of the tumor. As for images from normal tissue, His-MMDM can infer its appearance when one or more of the tumor suppressors/epigenetic factors are mutated (Fig. 3g). As for images from tumor tissue with some somatic mutations, His-MMDM can traceback to its WT appearance or infer later stages of the tumor with more accumulating mutations. For instance, for an image from the prostate tumor with existing mutations in \textit{TP53} and \textit{KMT2D}, His-MMDM can edit it into more or fewer mutations in the tumor suppressors/epigenetic factors and can freely modify it into versions when the mutations of tumor suppressors and epigenetic factors are applied either together or alone (Fig. 3h).

Besides genomic mutations, we investigated the effect of modifying transcriptomic profiles on the generated histopathological images. As transcriptomic profiles are shaped by regulations of pathways, we first investigated the effect of the top 10 MSigDB transcriptomic pathways that are most dysregulated in TCGA tumor samples (Methods). To exaggerate the effect of transcriptomics on the generated images, for each histopathological image in the TCGA cohort, we used His-MMDM to produce one version of it corresponding to when a pathway was knocked up and another one when it was knocked down. Similarly, we computed the cosine distance between the two versions of the image and visualized it accordingly (Fig. 4a). Most notably, the manipulation of the ‘Mitotic Spindle’ pathway resulted in the greatest change in histopathology in almost all tumor types (Fig. 4a and Fig. S5a). The result is not surprising as tumor proliferation through mitosis usually has a high correlation with histopathological grading. Targets of the transcription factor Myc (‘Myc Targets V1’) are influential in the tumorigenesis of multiple GI tumors, and therefore, they display a high impact on the histopathology of COAD, READ, and STAD (Fig. 4a and Fig. S5a). The manipulation of transcriptomic pathways also had a greater histopathological effect in tumor types when their genes had previously reported associations with histopathology based on discriminative performance reported by Fu et al. \textsuperscript{46} (p=5.2e-03). Using Hover-Net, we observed a significant reduction in the number of malignant cells when either the ‘G2M checkpoint’ (in 9/19 tumor types) or the ‘DNA repair’ (in 7/19 tumor types) pathways was knocked up compared to when it was knocked down (Fig. 4b, 4c, Fig. S5b and S5c), suggesting that His-MMDM can capture the effect of the activation of such pathways in limiting tumor growth potential.
Aside from pathway-level alterations, changes of individual genes could also be critical in tumorigenesis. Most notably, the amplification of genomic segments can directly alter the transcriptomic expression level of genes through somatic copy number variations (SCNVs). We investigated the five genes with high pan-cancer SCNV rates reported by Aaltonen et al. that are also available in the transcriptomic embeddings of His-MMDM. These include the prominent oncogenic transcription factor, MYC, a core component of the PI3K signaling pathway, PIK3CA, two genes from the cyclin family, CCNE1 and CCND1, and one gene from the Bcl-2 family, MCL1. We used His-MMDM to produce one image with an average expression level of the genes in the normal samples and compared it against another version when each of their expression levels were elevated (3 x s.d. above the average). Most notably, we observed a greater effect of the cyclin genes, CCNE1 and CCND1, on the tumor histopathology, suggesting the SCNVs that promote the progression of the cell cycle may have a universal effect across tumor types (Fig. 4d, Fig. S5d-S5e). In line with the results of Myc target genes in the pathway analysis, the gene MYC itself also induced strong histopathological alterations in multiple GI tumor types (Fig. 4d, Fig. S5d-S5e). As before, for these SCNV genes, there is also an association between their greater histopathological effect and previously reported discriminative performance on their transcriptomic levels (p=0.04 overall and p=0.02 if only tumor types where the genes have high amplification rates are considered).

The immune microenvironment of tumors plays an essential role in tumor development, subtyping, and prognosis. Thorsson et al. systematically characterized the immunogenomic signatures in TCGA and classified each tumor sample into six immune subtypes (C1-C6) based on the clustering of transcriptomic profiles of key immune pathways. As it is reported that tumor samples in such immune subtypes display distinct characteristics such as tumor proliferation and the types of infiltrated immune cells, we were interested in investigating whether His-MMDM can reproduce such observations. We selected the most abundant three (C1, C2, and C3) out of the six immune subtypes studied by Thorsson et al. and dropped the three kidney tumor types due to their low numbers of C1 and C2 immune subtypes (see Fig. 1D of Thorsson et al.). We then used His-MMDM to edit the images from one immune subtype into another one (e.g., editing C1 images into C2 or C3) by manipulating the five underlying pathways that define the immune subtypes (Methods, Table S4). The detection of cell types using Hover-Net on the images before and after His-MMDM’s edit revealed an increase (with a few exceptions) in the number of necrotic cells when translating images from C1 to C3 (in five tumor types) and C2 to C3 (in five tumor types) (Fig. 4e, 4f, Fig. S5f) which is consistent with C3’s lower tumor cell proliferation and high inflammatory response. As a negative control, we used
randomly selected sets of genes (while keeping the sizes of the gene sets the same) \((\text{Table S4})\) for
the underlying immune pathways and did not observe statistically significant increments in most
cases \((\text{Fig. S5g})\). We finally demonstrate through an example when His-MMDM is guided by the
combination of genomic mutations and transcriptomic profiles to edit histopathological images
between BRAF-like and RAS-like thyroid tumors. However, we delegate such discussions to the
Supplementary Materials \((\text{Supplementary Notes Section S2, Fig. S5h-S5i})\).

**Translation of histopathological images between primary and metastatic organ sites**

Metastasis of tumor is one of the leading factors that reduces a patient’s life quality and survival.
Therefore, the prediction of metastasis of tumors has important application values in clinical
diagnosis. From a histopathological perspective, tumors displaying higher levels of malignancy
tend to have a higher tendency to metastasize \(^5^5\). Several previous studies have already attempted
to build discriminative models that can infer the metastatic potential of tumors from such images
\(^5^6\)-\(^5^8^\).

Using His-MMDM, we investigated our model’s potential application on this task but from a
generative model’s perspective. We trained His-MMDM on 475 lung tumor histopathological
WSIs, including 315 from primary lung tumors, 76 from lymph node metastases, and 84 from
brain metastases from the Harbin Medical University Cancer Hospital \((\text{the ‘HMU-C dataset’,}
\text{Methods})\) \((\text{Methods, Table S2})\). During training, His-MMDM was conditioned on the site \((\text{lun}g,
lymph, \text{and brain})\) of the tumor tissue, and during inference, His-MMDM was asked to translate
the images from one site to another \((\text{Fig. S6a})\). As before, we computed \(\Delta \text{FID}\) between each pair
of translations. The \(\Delta \text{FID}\) scores are much higher when translating from a non-solid organ site
(lymph node) to a solid organ site \((\text{lung and brain})\) and the translation within solid organ sites
(lung and brain) tends to result in lower \(\Delta \text{FID}\)s \((\text{Fig. 5a})\). Inspecting the original and translated
histopathological images, His-MMDM keeps the cellular arrangements and characteristics of the
tumor subtypes during the translation processes, but the tumor growth environment is modified to
match the designated organ sites \((\text{Fig. 5b})\).

We systematically visualized the images translated from the primary organ site \((\text{lung})\) to the
metastatic organ sites, including the lymph node \((\text{Fig. 5c})\) and brain \((\text{Fig. S6b})\) according to their
Inception V3 features in the 2D tSNE space. The images located at the interior of the distribution
are more likely to be properly translated from the primary site characterized by the properly
arranged tumor cell nuclei and the more realistic background tumor environment when assessed
by pathologists \((\text{Fig. 5c, Fig. S6b and Figure S6c})\). On the contrary, the translated images that
look eccentric \((\text{characterized by incorrect cell shape, infiltration, and unrealistic backgrounds})\)
tend to be located at the edge of the distribution whose appearance indicates less successful image
translation which could be attributed to a lack of compatibility of the primary site image with the
metastatic site (Fig. 5c). We trained an outlier detection model, Isolation Forest $^{59}$, on the
Inception V3 features of the translated images and produced an ‘inlier score’ (in the range [-1,1],
-1 for the most out-of-distribution and 1 for the most in-distribution) for each lung image to
measure its compatibility with the metastatic organ site (Fig. 5c and Fig. S6b). We speculated
tumors that are easy to translate to a particular metastatic site image will have a higher ‘inlier
score’ (compatibility) with that particular organ site. To verify this, we translated primary lung
tumor histopathological images (LUAD and LUSC) from TCGA to the metastatic organ sites
(brain and lymph node) using the trained His-MMDM model and computed their inlier scores
using the outlier detection model. Interestingly, we found that the inlier scores indeed showed
some associations with the staging of the patients. Specifically, a patient’s high brain inlier score
is associated with M1 staging and an overall staging of III-IV, and a patient’s high lymph node
inlier score is associated with N1-N3 staging and an overall staging of II-IV (Fig. 5d). Moreover,
a high lymph node inlier score is associated with poorer survival of the patients (Fig. 5e, p=5.6e-04). This trend also holds for the brain inlier score (Fig. 5f, p=0.039).

**Cryosection to FFPE conversion and virtual IHC staining of histopathological images**

We finally test the effectiveness of His-MMDM for general image translation tasks such as
converting cryosectioned slide images to formalin-fixed and paraffin-embedded (FFPE) ones$^{25}$
and virtual IHC staining $^{23,24,60}$, without needing further task-specific adaptation. This task has
important clinical applications because cryosectioned slides are most often done during surgeries
and rapid diagnosis is needed for such intra-operative processes.

To convert cryosectioned slide images into FFPE with His-MMDM, we trained it to conditionally
generate TCGA slide images of their respective slide type. We ran the forward diffusion process
on a cryosectioned slide image under the condition ‘cryosection’ followed by the backward
denoising process under the condition ‘FFPE’ (Methods, Fig. S7a). We translated the test image
patches from all 19 tumor types in TCGA using our pre-trained model and used the $\Delta FID$ as the
metric for evaluating the performance of translation (Fig. 6a). Although His-MMDM is not a
dedicated model on this task, it still achieved comparable performance to the dedicated method
AI-FFPE $^{25}$ in most tumor types and even better in five out of 19 tumor types. Through more
detailed visual examination, we found that His-MMDM can alleviate the problem of tissue
dehydration and nuclear deformation that are common in cryosectioned slides and generate a
better presentation of contrasting regions and arrangement of the histopathological contents (Fig. 6b).

For IHC virtual staining of histopathological images, we collected patch images from 2,753 H&E slides and 3,447 IHC slides of 13 common markers used in brain tumor diagnosis spanning five glioma subtypes and five meningioma subtypes from The First Affiliated Hospital of Harbin Medical University (hereafter referred to as the ‘HMU-1st dataset’) (Methods, Table S2). Of the 13 IHC markers, five are routinely used in glioma diagnosis, four are used in meningioma diagnosis, and four are commonly used for both two categories of brain tumors (Table S5). In this dataset, each patient has both the H&E slides and the IHC slides, but they are not from the same tissue slices thereby needing an unpaired image translation model for this task. His-MMDM is then trained on this dataset to perform conditional generation of both H&E images as well as each of the 13 IHC markers. Eventually, it achieves virtual staining by running in serial the forward diffusion process conditioned on the label ‘H&E’ and the backward denoising process conditioned on the label of the desired IHC marker (Fig. S7b).

We evaluated the ‘inverted normalized FID scores’ (Methods) of the translated images in each subtype of glioma and meningioma (Fig. 6d). The IHC markers into which the H&E images are translated tend to have higher performance when the marker is more compatible with the brain tumor type/subtype. For instance, the glioma markers (GFAP, ATRX, IDH-1, MGMT, and Oligo-2) tend to have higher performance in gliomas and so do the meningioma markers (EMA, PR and SSTR2) in meningiomas (Fig. 6d). The positive regions of the virtually stained images (as indicated by the brown-colored diaminobenzidine (DAB) stain) is also in agreement with the marker’s expected location of expression (in the nucleus or the cytoplasm and extracellular matrix) (Fig. 6e and 6f, Fig. S7c and S7d). Using color deconvolution, we computed the intensity of the DAB stain on the virtually stained images in each glioma and meningioma subtype (Fig. 6g-6i). Most notably, positivity of IDH-1 (suggesting IDH-1 mutation) is higher in lower-grade glioma subtypes (AA, AO, DA, and O) rather than glioblastomas (GBM), which is consistent with previous findings that IDH-1 mutations are more frequently found in the lower-grade gliomas or secondary GBMs that stem from them, rather than the primary GBMs (Fig. 6g). The intensity of subtype-nonspecific markers tends to correlate with the proliferation activity and cell density of the specific subtypes. For example, Oligo-2, as a transcription factor necessary in the development of nearly all glioma subtypes, tends to have a higher intensity in the more proliferative GBM subtype (Fig. 6g). Similarly, EMA, PR, and SSTR2, being common meningeal markers, tend to have higher intensity in the more proliferative atypical or the more
dense fibrous subtype (Fig. 6h). As for the common markers used in both glioma and meningioma, CD34, an endothelium marker, reasonably exhibits a higher intensity in angiomatous meningioma and in GBM where angiogenesis is frequent (Fig. 6i) 66. Additionally, GBM and the higher-grade atypical meningioma also show higher intensity of the cell proliferation marker Ki67 (suggesting higher expression) and tumor suppressor p53 (suggesting p53 mutation). Collectively, these analyses demonstrate the effectiveness of His-MMDM in virtual staining and the consistency of the synthesized images with prior knowledge.

Discussion

In this study, we have developed His-MMDM, a histopathological image translation model that can translate histopathological images between multiple categorial domains or edit them guided by genomic or transcriptomic profiles. Based on the principle of DMs 7, His-MMDM connects two diffusion models in which one of them diffuses the given source image into Gaussian noise and the other denoises the Gaussian noise into the target image. In both translation settings, His-MMDM uses the given source image as a template and aims to generate an image under the target condition while preserving the semantics in the original image.

A lot of previous evidence has suggested the feasibility of building such general image translation models across histopathological domains. Firstly, several previous studies have pointed out the existence of conserved histopathological image features across different tumor types 32,33. Secondly, pan-cancer studies have demonstrated through discriminative models the high predictive performance of the tumor origin from histopathological images 37, multiple genomic mutations 46, and multiple genes’ expression levels 47. We demonstrated, for the first time, that such relationships can be manifested through a generative model, by explicitly generating histopathological images corresponding to in silico altered tumor types or genomic and transcriptomic profiles. Thirdly, several dedicated histopathological image translation models have already been proposed in specific application domains such as cryosection to FFPE translation and in silico virtual staining, which proves the applicability of image translation in histopathology 24,25. We have shown that the general His-MMDM model can achieve performance comparable to these dedicated models.

Very recently, several new generative models to synthesize histopathological images based on genomics or transcriptomics have begun to emerge. For instance, Dolezal et al. proposed a conditional GAN (cGAN) based model 67 that can synthesize images of different subtypes in lung, breast, head & neck, and thyroid tumors. In particular, the model can synthesize thyroid tumor images from BRAF^{V600E}-like ones to RAS-like ones. Carrillo-Perez et al. sequentially
developed a GAN-based \(^68\) and a diffusion-based generative model \(^69\) for the synthesis of histopathological images from bulk RNA-seq profiles. It is worth noting that these previous works are generative models conditioned only on a genomic/transcriptomic profile that synthesize new images \textit{from scratch}. His-MMDM, however, distinguishes itself as an image translation model that can edit existing images into new ones. This property is indispensable in our analyses of the effect of genomic/transcriptomic manipulations on real histopathological images. This also makes His-MMDM more useful in real-world applications in that it achieves multi-omics-guided editing of the existing images rather than the un-constrained generation of new ones.

In recent years, GenAI-based copilot systems have shown great potential to revolutionize the way we interact with computers, whether it is about the autonomous AI agents powered by large language models (LLMs) \(^70\), GenAI-based multi-media content creation \(^71\), or LLM-based program coding \(^72\). In parallel, preliminary attempts have also been taking place in biomedicine and such GenAI systems are showing great potential in helping biologists and healthcare practitioners \(^18,19,73,74\). Specifically in histopathology, existing software systems, such as QuPath \(^75\), already showing excellent extensibility in scripting and plugin development, could be further updated by enabling GenAI-powered functionalities. GenAI models such as His-MMDM could fit naturally into such ecosystems and become useful for a pathologist’s day-to-day workflow.

One limitation of His-MMDM concerns its high computational demand. Due to the iterative nature of the diffusing and denoising processes of diffusion models, it still takes minutes to translate a single batch of histopathological image patches on a typical machine equipped with eight NVIDIA V100 GPUs. Strategies for developing more efficient diffusion models \(^76\) can be adopted to streamline the process of image translation. Experiments with more efficient diffusion model solvers \(^77,78\) can also be an interesting exploration for future works. Currently, for simplicity, we trained His-MMDM with fixed resolution (128×128) under a unified magnification level in each cohort. Extending His-MMDM to simultaneously work at different magnification levels will further unleash its application potential.

**Methods**

**Datasets**

In this research, three cohorts of histopathological image datasets were used to train and test the His-MMDM model under different image translation settings.

\textit{The Cancer Genome Atlas cohort (TCGA)} We downloaded 22,596 whole-side histopathological images from The Cancer Genome Atlas (TCGA) web portal (https://portal.gdc.cancer.gov/) of the 19 cancer types following the practice of \(^32\). The 19 cancer types contain five gastrointestinal...
(ESCA, STAD, COAD, READ, and PAAD), two lung (LUAD and LUSC), three kidney (KIRP, KICH, and KIRC), three pan-gynecological (OV, BRCA, and UCEC), and five other tumor types (LIHC, HNSC, SARC, THCA, BLCA, and PRAD) (statistics shown in Table S2). These tumor types in TCGA were specially chosen because they contain enough numbers of tumor and normal slides, as well as enough numbers of FFPE and cryosectioned slides for analyses. The patient’s metadata and clinical information were downloaded along with the WSIs. The processed genomic mutations (TCGA Unified Ensemble "MC3" gene-level mutation calls) and the bulk transcriptomic profiles (Illumina Hi-Seq) of the corresponding samples were downloaded from the UCSC Xena website (https://tcga.xenahubs.net) 79. This cohort is used for training and testing His-MMDM to translate across different primary tumor types and different genomic and transcriptomic profiles. For genomic profiles, the 522 genes we considered in total include the 334 genes from the ten oncological signaling pathways defined in 48, and 188 other genes with high pan-cancer mutation rate or discriminative performance 46 in the TCGA cohort. We strictly followed the "MC3" gene-level mutation calls’ definition and only non-synonymous mutation statuses were considered for each gene. For transcriptomic profiles, we considered a total of 4,461 genes that are available in the TCGA transcriptomic profiles, including 4,193 genes that are included in the collection of 50 hallmark pathways of MSigDB (v2023.2) 49,50, as well as an additional 268 genes that either displayed the greatest variability between tumor and normal samples (in terms of the t-statistic) or previously reported discriminative performance 46. During the evaluation of the His-MMDM model, we prioritized the top 30 genes in terms of pan-cancer mutation rate from nine oncological pathways for genomic-guided editing and the top 10 pathways in terms of pan-cancer t-statistic from MSigDB for transcriptomic manipulation.

The Harbin Medical University Cancer Hospital Cohort (HMU-C) This cohort contains 475 primary and metastatic lung tumor histopathological WSIs of 400 patients from the Harbin Medical University Cancer Hospital. This HMU-C cohort consists of 315 WSIs from primary lung tumor tissues, 76 WSIs from lymph node metastases, and 84 WSIs from the brain metastases of the same set of patients. The primary and metastatic tumor tissues are from the most common lung tumor subtypes: adenocarcinoma, squamous cell carcinoma, and small cell carcinoma (statistics in Table S2). All slides in this cohort are FFPE slides.

The First Affiliated Hospital of Harbin Medical University Cohort (HMU-1st) This cohort contains 6,200 brain tumor histopathological WSIs of 557 patients from the First Harbin Medical University Hospital. Among the 6,200 WSIs, 2,660 are glioma slides, and 3,540 are meningioma slides; 2,753 were H&E-stained and 3,447 were IHC-stained with 14 different markers (Table...
Some of these markers are useful for the confirmation of the source of the tumor tissue, such as ATRX; some are used to display specific cellular processes, such as Ki67; while others are useful in tumor subtyping and forecasting prognoses, such as IDH-1 and MGMT (Table S5). Among them six markers are commonly used in both meningiomas and gliomas, four are used exclusively in gliomas and the other four are used exclusively in meningiomas (Table S5).

The protocols for the human studies comply with all relevant ethical regulations and are approved by the Ethics Committee of The First Harbin Medical University Hospital and The Harbin Medical University Cancer Hospital (KY2021-42). The consent forms of the patients were waived before this research was carried out under the retrospective research protocol of the institutions.

**Patch selection and patch feature extraction**

The WSIs from each cohort were segmented for tissue regions from the empty slide background. The slide image was then converted to a binary mask using Otsu’s thresholding method on the Gaussian blurred version of the saturation channel. Subsequently, using a sliding window-based approach, the carved-out tissue area is completely covered with image patches that are 256×256 (if the objective magnification is 20×) or 512×512 (if the objective magnification is 40×) in size. To select patches from the WSIs that contain representative tumor tissues, we extracted the image patches’ features using a pretrained (on ImageNet 80) ViT-L-16 81 model. From each cohort’s extracted patches, we asked the pathologists to select a small set of positive patches (which contains representative tumor cells and microenvironment) and a small set of negative patches (in which the tumor tissue content is less than 10% of the area covered). We then classified the patches as positive/negative from each WSI based on their similarity to the positive/negative patch sets (as measured by the cosine similarity of their ViT-L-16 feature vectors). From each WSI, we randomly sampled a certain number of patches from the ‘positive patches’ determined above and then incorporated them in the train/test set of the His-MMDM model. The patches were resized to 128×128 before being processed by the generative models. The statistics of the extracted patches in each cohort are shown in Table S2.

**The Multi-domain multi-omics image translation model**

His-MMDM is an image translation model that learns a mapping $F: X^{(src)} \rightarrow X^{(trg)}$ that maps a source domain image $X^{(src)}$ to a target domain image $X^{(trg)}$. If the images in the source domain are distributed as $p_{src}(X)$ and the images in the target domain are distributed as $p_{trg}(X)$, a well-trained image translation model is expected to have the property that $F(X^{(src)})$ distributes as $p_{trg}(X)$. Although it is possible to translate images between different dimensions 29, we are
assuming that $X^{(\text{src})}$ and $X^{(\text{trg})}$ to have the same dimensions in this paper for simplicity. This assumption is valid since we have resized images to represent the same physical dimension when they come from slides with different magnification levels.

Su et al. have demonstrated theoretically that such mapping $\mathcal{F}$ corresponds to a deterministic solution of the probability flow ordinary differential equations (PF-ODEs) of the Schrödinger Bridge Problem (SBP) between the two distributions $p_{\text{src}}$ and $p_{\text{trg}}$. The SBP from $p_{\text{src}}$ to $p_{\text{trg}}$ aims to establish the most likely evolutionary path of a distribution from $p_{\text{src}}$ to $p_{\text{trg}}$. They also established the connection between the SBP and the score-based generative modeling (SGM) of DMs and proved that SGM is a special case of the SBP when the evolution is from a particular data distribution (could be $p_{\text{src}}$ or $p_{\text{trg}}$) to the multivariate Gaussian distribution. In this way, such mapping $\mathcal{F}$ can be found by composing the forward (from $p_{\text{src}} \to N(0, I)$) and backward diffusion processes (from $N(0, I) \to p_{\text{trg}}$) of the SGM. Concretely, during the forward diffusion processes, we obtain the latent noisy image $X^{(\text{lat})}$ from a source domain image $X^{(\text{src})}$ with

$$X^{(\text{lat})} = \text{ODESolve}(X^{(\text{src})}, f^{(\text{src})}, t_1 = 0, t_2 = T)$$

where $f^{(\text{src})}$ is the denoising diffusion model trained in the SGM of the source domain, ODESolve($\cdot; t_1 = 0, t_2 = T$) represents the numerical approximation of the forward PF-ODE that maps the source domain image $X^{(\text{src})}$ to the latent image $X^{(\text{lat})}$, and $T$ is the preset number of discretization steps of the solver. If we assume that the SGM models perfectly the source domain score function and there is no discretization error in solving the PF-ODE, $X^{(\text{lat})}$ should distribute as $N(0, I)$. Symmetrically, we can then transform the latent image $X^{(\text{lat})}$ to the target domain with

$$X^{(\text{trg})} = \text{ODESolve}(X^{(\text{lat})}, f^{(\text{trg})}, t_1 = T, t_2 = 0)$$

where $X^{(\text{trg})}$ is the generated target domain image, ODESolve($\cdot; t_1 = T, t_2 = 0$) represents the numerical approximation of the backward PF-ODE that maps the latent image to the target domain image, and $f^{(\text{trg})}$ is the denoising diffusion model of the target domain SGM.

His-MMDM focuses on the translation of histopathological images belonging to different domains. Such domains include categorical domains such as different organ sites of the tumor and different types of stains, as well as multi-omic domains that correspond to different genomic or transcriptomic profiles. Therefore, the denoising diffusion model used in His-MMDM needs to simultaneously take care of the three types of condition information. For clarity, we first write the categorical condition as $c \in \mathcal{C}$, where $\mathcal{C}$ is a finite set containing all the possible categorical
conditions that are handled by this model. Secondly, we use \( G = \{(g_i, m_i)\}_{i=1}^{|G|} \) to hold the genomic mutation profiles of all the genes considered by the model (denoted by gene set \( G \)) for a particular train/test example, in which \( g_i \) is the name of the \( i \)th gene and \( m_i \) is the Boolean value of its mutation status. Thirdly, we use \( T = \{(g_i, t_i)\}_{i=1}^{|T|} \) to hold the transcriptomic expression levels of all the genes considered by the model, in which \( g_i \) is the name of the \( i \)th gene and \( t_i \) is the quantile-normalized value of its expression level. In this way, the source domain model \( f^{(\text{src})} \) and the target domain model \( f^{(\text{trg})} \) can be obtained by providing one model parameterized by \( \theta \) \( (f_{\theta}) \) with different condition information, i.e., \( f_{\theta}(\cdot; c^{(\text{src})}, G^{(\text{src})}, T^{(\text{src})}) \) and \( f_{\theta}(\cdot; c^{(\text{trg})}, G^{(\text{trg})}, T^{(\text{trg})}) \), respectively.

As the denoising diffusion model is usually implemented as a U-Net architecture, the two types of condition information are added to each layer of the U-Net model via the categorical embedding and the multi-omics embedding, respectively:

\[
\text{Layer}[l+1] = \text{Layer}[l] + e_{\text{cat}} + e_{\text{genomic}} + e_{\text{transcriptomic}}
\]

The categorical embeddings are initialized randomly and trained together with the parameters of the U-Net. As for the genomic embeddings, for each gene \( g \), His-MMDM utilizes one embedding for the wide-type (WT) version of the gene (\( e_g \)) and the other for the mutated version of the gene (\( \tilde{e}_g \)) and use a multi-layer perceptron (MLP) to transform them, in other words:

\[
e_{\text{genomic}} = \sum_{i=1}^{|G|} \text{MLP}_{\text{genomic}}(e_{g_i} \cdot \mathbb{I}[m_i \text{ is false}] + \tilde{e}_{g_i} \cdot \mathbb{I}[m_i \text{ is true}])
\]

To account for the quantitative effect of transcriptomics, \( e^{(i)}_{\text{transcriptomic}} \) is implemented as two different networks (both as MLPs) that deal with genes and expression levels separately:

\[
e_{\text{transcriptomic}} = \sum_{i=1}^{|G|} \text{MLP}_{\text{transcriptomic}}(e_{g_i}) \times \text{MLP}_{\exp}(e_{g_i}, t_i)
\]

We first pre-train the denoising diffusion model \( f_{\theta} \) using the standard classifier-guided diffusion training procedure on the domains of images between which we wish to translate. We used the 

**Algorithm 1** in \(^7\) for the training procedure. After \( f_{\theta} \) is successfully trained, we use the denoising diffusion implicit models (DDIMs) as ODESolve for image translation. The process of it can be summarized as **Algorithm 1**. The details of the model architecture, hyperparameters and other configurations regarding training and translation are specified in Table S1.
For comparison, we compared the performance of His-MMDM with two GAN-based models, CycleGAN \(^{27}\) and CUT \(^{34}\). To translate images from domain X to domain Y, CycleGAN simultaneously establishes two GANs where one translates images from X to Y and the other from Y to X. To ensure the consistency of the translated images, a cycle-consistency loss is imposed to ensure the invariance when an image is translated from X to Y and then back from Y to X. The other GAN-based model, CUT, resorts to a different approach. CUT establishes only one GAN model to achieve one-way translation from X to Y. It ensures consistency of translation using patchwise contrastive learning \(^{83,84}\) to encourage image patches of the source and translated images of the same location to be similar.

### Image translation tasks and the evaluation of the synthesized images

We summarize the specific settings of image translation tasks, including the specifications of the model checkpoints, the dataset used the condition information of the translation tasks in Table S1. The Frechet Inception Distance (FID) \(^{35}\) was used to evaluate the quality of a set of synthesized images against a set of real images. For baseline models (CycleGAN, CUT, and AI-FFPE), we trained and tested them on the same train/test split as His-MMDM. The image features after the last pooling layer of an Inception V3 model were used for the computation. Let \((\mu, \Sigma)\) and \((\mu_0, \Sigma_0)\) denote the mean and covariance of the Inception V3 features of the synthesized images and real images respectively. The FID score is defined as:

\[
\text{FID}(\mu, \Sigma, \mu_0, \Sigma_0) = \|\mu - \mu_0\|^2 + \text{Tr}(\Sigma + \Sigma_0 - 2(\Sigma\Sigma_0)^{1/2}).
\]

To quantify the effect of image translation across primary tumor types and tumor organ sites, we computed the difference of FID scores before and after image translation, i.e.,

\[
\Delta \text{FID} = d_{\text{FID}}(\mu_{\text{before}}, \Sigma_{\text{before}}, \mu_0, \Sigma_0) - d_{\text{FID}}(\mu_{\text{after}}, \Sigma_{\text{after}}, \mu_0, \Sigma_0).
\]

Although both CycleGAN and CUT are GAN-based algorithms, CycleGAN utilizes cycle-consistency for image translation while CUT leverages patch-wise contrastive learning \(^{83,84}\). Based on the reported performance on TCGA, although patch-wise contrastive learning is an effective objective to train image translation models between different photography domains, it seems to be less effective in histopathological domains, possibly due to the repetitive nature of the contents of the histopathological images which can confuse this contrastive objective. To compute the magnitude of the effect of genomic- and transcriptomic-guided editing, we used cosine distance rather than FID as we were interested in the editing effect on a single image rather than a collection of images. Due to the vast difference between H&E and IHC-stained images, we evaluated the quality of IHC virtual staining in a different way than the previous two tasks. Instead, we computed the ratio of the FID score between two different sets of real IHC images.
and the FID score between synthesized and real IHC images, i.e., \[ \frac{d_{\text{FID}}((\mu_{\text{syn}}, \Sigma_{\text{syn}}), (\mu_{\text{real}}, \Sigma_{\text{real}}))}{d_{\text{FID}}((\mu_{\text{syn}}, \Sigma_{\text{syn}}), (\mu_{\text{syn}}, \Sigma_{\text{syn}}))} \]. We refer to this as the ‘inverted normalized FID score’ in the main text. In this way, both the effect of image translation and the inherent variability of images of an IHC-stain can be taken into account, and the inversion makes the metric greater if the synthetic IHC images are more similar to the real IHC images.

**Outlier detection of translated histopathological images**

When His-MMDM translates an image in the source domain that is unlikely to have a corresponding image in the target domain, the translated image tends to look very eccentric (Fig. 4c). To quantify the eccentricity of the translated images, we fitted an outlier detection model (Isolation Forest \(^{59}\), implemented by the ‘scikit-learn’ package \(^{85}\) on the ViT-L-16 features of them. Using the fitted model, we were able to obtain an ‘inlier score’ in the range of \([-1,1]\) of each sample, where a score close to 1 indicates that the image is a perfect inlier and a score close to -1 indicates that the image is likely to be an outlier. Superimposing this score on the tSNE plot of the ViT-L-16 features of the translated patch images also shows that the patches with low scores tend to be located on the edge of the tSNE structure. Using the trained His-MMDM model and the fitted outlier detection model, we applied them to the TCGA LUAD and LUSC cohorts. To obtain an inlier score of a WSI, we sampled 50 image patches from it and aggregated their scores by averaging them. Similarly, we obtained the scores of each patient by averaging the inlier scores of all their available WSIs. Finally, we analyzed the relationship between a patient’s inlier score and the patient’s staging and survival information. The Kaplan-Meier method was used to estimate the survival function of the patients under each condition. P values of the fitted Cox proportional hazard models were reported as statistical significance values.

**Stain quantification**

To compute the positivity of the IHC stains in the translated images, we applied the color deconvolution algorithm \(^{86}\) (implemented by the ‘scikit-image’ package \(^{87}\)) to the images and obtained the intensity of the two stains components used in our HMU-1st cohort, i.e., hematoxylin (as background) and 3, 3’-diaminobenzidine (DAB, as positive signal). The averaged intensity of the DAB stain is used as the metric of the positivity of the patch.

**Statistics**

Wilcoxon rank-sum tests and Wilcoxon signed-rank tests were performed using the statistical functions from the Scipy package (scipy.stats) \(^{88}\). Statistical significance (p values) was reported in the respective figures.
Algorithm 1 Image-translation with His-MMDM

function TRANSLATE-IMAGE \((X^{\text{src}}, f_0, \{c^{\text{src}}, G^{\text{src}}, T^{\text{src}}\}, \{c^{\text{trg}}, G^{\text{trg}}, T^{\text{trg}}\})\)

inputs:
\(X^{\text{src}}\): the source image to translate
\(f_0\): the trained denoising diffusion model to use
\(\{c^{\text{src}}, G^{\text{src}}, T^{\text{src}}\}\): the three types of condition information of the source image
\(\{c^{\text{trg}}, G^{\text{trg}}, T^{\text{trg}}\}\): the three types of condition information of the target image

constants:
\(T\): number of steps of the DDIM solver

outputs:
\(X^{(t)}\): the translated image

\[X_0^{\text{src}} \leftarrow X^{\text{src}}\]

for \(t = 0 \ldots T - 1\) do // Implementation of ODESolve\(\left(X^{\text{src}}, f^{\text{src}}, t_1 = 0, t_2 = T\right)\)
\[X_{t+1}^{\text{src}} \leftarrow \sqrt{\frac{\Delta t}{4}} X_t^{\text{src}} \left(f_0(\cdot | c^{\text{src}}, G^{\text{src}}, T^{\text{src}}) + \sqrt{1 - \frac{\Delta t}{4}} \frac{1}{\partial_{x^r}} \hat{e} \left(X_t^{\text{src}}, f_0(\cdot | c^{\text{src}}, G^{\text{src}}, T^{\text{src}})\right)\right)\]
end

\[X_T^{\text{trg}} \leftarrow X_0^{\text{trg}}\]

for \(t = T \ldots 1\) do // Implementation of ODESolve\(\left(X^{(t)}, f^{(t)}, t_1 = T, t_2 = 0\right)\)
\[X_{t-1}^{(t)} \leftarrow \sqrt{\frac{\Delta t}{4}} X_t^{(t)} \left(f_0(\cdot | c^{(t)}, G^{(t)}, T^{(t)}) + \sqrt{1 - \frac{\Delta t}{4}} \frac{1}{\partial_{x^r}} \hat{e} \left(X_t^{(t)}, f_0(\cdot | c^{(t)}, G^{(t)}, T^{(t)})\right)\right)\]
end

\[X_0^{\text{trg}} \leftarrow X_0^{\text{trg}}\]

return \(X^{\text{trg}}\)

Acknowledgments

This publication is based upon work supported by the King Abdullah University of Science and Technology (KAUST) Office of Research Administration (ORA) under Award No REI/1/5234-01-01, REI/1/5414-01-01, RGC/3/4816-01-01, REI/1/5289-01-01, REI/1/5404-01-01, REI/1/5992-01-01, and URF/1/4663-01-01. The results shown here are in whole or part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

Authors’ Contributions

Z.L., T.S., P.L., and X.G. conceived the study. Z.L. and X.G. developed the His-MMDM model and performed the computational analyses, with algorithm development advised by W.H. and computational analysis advised by B.Z. T.S., H.M., S.Z., Y.W., and P.L. prepared and scanned the histopathological images for the HMU-T dataset, while Y.C., X.C., and S.Z. prepared and scanned the histopathological images for the HMU-1st dataset. Pathologists T.S., Y.C., H.M., and G.S. participated in the inspection and interpretation of the experimental results. Z.L., T.S., and X.G. drafted the majority of the manuscript. All authors reviewed and approved the final manuscript.
Data and Code Availability

The code and trained checkpoints of His-MMDM are available at https://github.com/lzx325/His-MMDM. Histopathology Data from TCGA used in this study are available from the Genomic Data Commons Portal of the National Cancer Institute (https://gdc.cancer.gov/). Example data from The First Harbin Medical University Hospital and The Harbin Medical University Cancer Hospital are deposited at Zenodo (10.5281/zenodo.12636449) and are available from the lead contact upon reasonable request.

Declaration of Interests

The authors have declared no competing interests.

Figures

Figure 1 Architecture of His-MMDM

The architecture of Histopathological image Multi-domain Multi-omic translation with Diffusion Models (His-MMDM). Based on the principle of DMs, the forward diffusion procedure gradually diffuses a histopathological image into a noisy image following a Gaussian distribution. The backward procedure takes in the noisy image and gradually denoises it back into a real histopathological image. A U-net-based network makes noise prediction for a given input and takes in condition information in the format of categorical labels, genomic profiles, and transcriptomic profiles in its embedding layers.

Figure 2 Translation of histopathological images across primary tumor types

(a) Schematic that illustrates how His-MMDM translates across primary tumor types.

(b) Evaluation of His-MMDM (in terms of FID scores) in performing cross-tumor type translation and comparison with other image translation models.

(c) Detailed pairwise heatmap showing the improvement of FID score ($\Delta$FID = FID$_{before}$ - FID$_{after}$) between each pair of TCGA classes. Bar plots show the averaged $\Delta$FID for a particular translation source tumor type (rows) or target tumor type (columns).

(d) Image examples showing the effect of translation across tumor types. The COAD patches were translated across three adenocarcinomas (READ, STAD, and PAAD) and four non-adenocarcinomas (LUSC, LIHC, KICH, and SARC). The typical glandular structures of adenocarcinomas were retained in the former group but attenuated in the latter group.

(e) The effect of image translation visualized in the tSNE space. We selected five tumor types, LUAD, ESCA, OV, LIHC, and KICH, each from a distinct family. One tumor type was set as the
translation target (LUAD in this panel) and images from the other four types were translated to it. tSNE dimensionality reduction is performed on the Inception V3 features of the untranslated and translated images. The translated images indeed distribute more similarly to the target of translation (LUAD) compared to their untranslated counterparts.

(f) The classification performance (F1 score) of binary tumor classifiers that are trained on other tumor types (as targets of translation) using either translated/untranslated images when there is no additional fine-tuning (‘Ratio for fine-tuning = 0’) or with additional fine-tuning (‘ratio for fine-tuning = 0.3, 0.6, 0.9’) in each tumor type (as the source of translation and the goal of classification). Data for each source of translation are aggregated for a particular tumor family and through macro-averaging.

Figure 3 The genomics-guided editing of histopathological images
(a) Schematic that illustrates how His-MMDM achieves histopathological image editing guided by genomic and transcriptomic profiles.
(b) The effect of genomic mutations on histopathological images by each tumor type, measured by cosine distance between the WT version image and mutated image. The top 30 genes with the highest pan-cancer mutation rate are selected. The cosine distances aggregated by each gene-tumor type pair through averaging are normalized either by each tumor type (circle color) or cross-tumor (circle size). Bar plots aggregate per each gene (row) or tumor type (row). The mutations of genes have a greater effect on histopathology in the gene-tumor type pairs where there were previously reported associations by Fu et al. (Wilcoxon rank-sum test; all gene-tumor pairs: p = 6.7e-05; only high-frequency (>10%) mutations: p = 0.03).
(c) The mutation of genes APC, PTEN, TP53, and CDKN2A increases the numbers of malignant cells detected by Hover-Net in COAD, UCEC, OV, and ESCA, respectively. P values are from the Wilcoxon signed-rank test. Each dot represents a particular image without (blue) or with (red) a particular mutation.
(d) Histopathological image examples illustrating the effect of mutations of APC, PTEN, TP53, and CDKN2A in COAD, UCEC, OV, and ESCA, respectively.
(e) Summary of the effect of five tumor suppressors and three epigenetic factors on the development and metastasis of prostate tumors reported by Cai et al.
(f) The effect of genomic mutations in the tumor suppressors and epigenetic factors on the histopathology images.
(g) An example illustrating His-MMDM’s ability to infer images at different tumor
developmental stages (mutation in one tumor suppressor gene, mutation in all tumor suppressor
genes, and mutation in all tumor suppressor and epigenetic factor genes) from a normal image.
Cosine distances are computed w.r.t. the initial normal image (w/o mutation).

(h) An example illustrating His-MMDM’s ability to infer images when there is more mutation or
less mutation than the observed image (mutation in $TP53$ and $KMT2D$). Such inference can be
made along different paths, i.e., the solid line and the dashed line, representing mutations
accumulating in different orders. Cosine distances are computed w.r.t. the initial inferred image
(w/o mutation).

Figure 4 The Transcriptomics-guided editing of histopathological images

(a) The effect of transcriptomic pathway manipulations on the histopathological images by each
tumor type, measured by the cosine distances between a knocked-down version and a knocked-up
version. The top 10 pathways with the highest t-statistics between TCGA tumor and normal
samples are selected for the experiment. The cosine distances aggregated by each pathway-tumor
type pair through averaging are normalized either by each tumor type (circle color) or cross-
tumor (circle size). Bar plots aggregate per each pathway (row) or tumor type (row). The
manipulation of pathways would have a greater effect on histopathology if the pathways have
genes with previously reported histopathological associations by Fu et al. (Wilcoxon rank-sum
test; $p = 5.2e-03$).

(b) The knock-up of the pathway ‘G2M checkpoint’ reduces the number of malignant cells
detected by Hover-Net in OV and UCEC respectively. The knock-up of the pathway ‘DNA repair’
reduces the number of malignant cells detected by Hover-Net in UCEC and COAD respectively.
Each dot represents a particular image when a particular pathway is knocked-down (blue) or
knocked-up (red).

(c) Histopathological image examples illustrating the effect of knock-up and knock-down of the
‘G2M checkpoint’ and the ‘DNA repair’ pathways in OV, UCEC, and COAD.

(d) The effect of transcriptomic manipulations of high-frequency SCNV genes reported by
Aaltonen et al., measured by the cosine distances between a knocked-down version and a
knocked-up version. Bar plots aggregate per each gene (row) or tumor type (row). The
manipulation of genes would have a greater effect on histopathology if the genes have previously
reported histopathological associations by Fu et al. (Wilcoxon rank-sum test; all gene-tumor pairs:
$p = 0.04$; only genes with high amplification rates ($>10\%$): $p = 0.02$).
(e) Editing of histopathological images based on transcriptomic signatures of immune profiles that result in a significantly higher number of necrotic cells detected by Hover-Net. An arrow is drawn from one immune subtype to another if the increase of such cells is statistically significant (Wilcoxon signed-rank test, p < 0.001).

(f) Histopathological image examples illustrating the effect of editing images into different immune subtypes from their original ones.

**Figure 5 Translation of histopathological images between primary and metastatic organ sites**
(a) The improvement of FID score before and after translation between each pair of organ sites (lung, lymph node, and brain). Bar plots show the averaged ΔFID for a particular translation source site (rows) or target site (columns).

(b) Histopathological image examples of three major lung tumor subtypes (adenocarcinoma, squamous cell carcinoma, and small cell carcinoma) illustrating the effect of translating images from the primary site (lung) to the common metastatic sites of lung tumor (brain and lymph node).

(c) Schematic and visualization of running the outlier detection model (Isolation Forest) on the translated images from the lung to the lymph nodes. The images are visualized according to their structure in the tSNE space and the ‘inlier scores’ are superimposed. The images with low ‘inlier scores’ (out of distribution) tend to be eccentric-looking and are located on the edge.

(d) The relationship between the brain inlier score and the stagings of cases in the TCGA lung tumor cohort. The stagings include the M (distant metastasis) and N (extent of regional lymph node spread) stagings in the TNM staging system.

(e-f) The relationship between the lymph node (e) and brain (f) inlier score to TCGA patients’ overall survival. The inlier scores of the images are aggregated per patient through averaging.

**Figure 6 Cryosection to FFPE conversion and virtual IHC staining of histopathological images**
(a) Evaluation of His-MMDM (in terms of ΔFID scores) in performing cryosection to FFPE conversion and comparison with other dedicated (AI-FFPE) and non-dedicated (CycleGAN and CUT) image translation models.

(b-c) Examples of cryosection to FFPE conversion in LUAD (b) and UCEC (c)

(d) The ‘inverted normalized FID scores’ of the virtually stained images for each IHC marker in each glioma and meningioma subtype
(e-f) Examples of virtual staining of H&E histopathological images in glioblastoma (e) and atypical meningioma (f). His-MMDM can perform virtual staining of the glioma-specific markers, the meningioma-specific markers, as well as the common markers used by both primary brain tumor types. The positive regions (indicated by the brown-colored DAB stain) are also in agreement with the respective markers’ expected location of expression (nucleus, cytoplasm, and extracellular matrix).

(g-i) The violin plots showing the intensity of the positive (DAB) stain on the virtually-stained images of glioma-specific markers (g) meningioma-specific markers (h) and common markers (i) among the primary brain tumor subtypes.

References


Backward Denoising Process

Forward Diffusion Process

Multi-omic Denoising U-Net

Translate Tumor Types

Genomics-guided Editing

Transcriptomics-guided Editing

Cryosection to FFPE conversion

Virtual Staining

\[ X_T \rightarrow X_{T-1} \rightarrow \ldots \rightarrow X_1 \rightarrow X_0 \]

\[ X_0' \rightarrow X_{T-1}' \rightarrow X_T' \]

Categorical Embedding

Primary Site

Genomics Embedding

Gene Embedding

Mutation Status

Transcriptomics Embedding

Gene Embedding

Expression Level

Virtual Staining
In Distribution Out-of Distribution

Incorrect cell shape and nucleus
Incorrect infiltration of blood cells and unrealistic backgrounds
Dense tumor cells and arrangement

Lymph Node Inlier Score

** OS lymph node inlier score
** OS brain inlier score

p = 5.6 \times 10^{-4} (Cox)
p = 0.039 (Cox)
ESCA (Source of translation, goal of classification)

STAD

COAD

READ

PAAD

LUAD

LUSC

KIRP

KICH

KIRC

OV

BRCA

UCEC

LIHC

HNSC

SARC

THCA

BLCA

PRAD

ΔF1

ΔPrecision

ΔRecall

0.0 0.1 0.2 0.3 0.4

0.0 0.1 0.2 0.3 0.4

0.0 0.1 0.2 0.3 0.4

F1 score
Significant changes observed at the initial stage
Lack of cellular content

Lack of microenvironment and with unrealistic components

Well-aranged tumor cells with realistic background

Brain Inlier Score

-0.10 -0.05 0.00 0.05 0.10

Tumor site 1

Tumor site 2

Lymph node

Brain
a

Cryosection

$X_T$, $X_{T-1}$, $X_1$, $X_0$, $X_0'$, $X_1'$, $X_{T-1}'$, $X_T'$

b

H&E

$X_T$, $X_{T-1}$, $X_1$, $X_0$, $X_0'$, $X_1'$, $X_{T-1}'$, $X_T'$

c

<table>
<thead>
<tr>
<th>Glioma-specific Markers</th>
<th>Common Markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATRX, IDH-1, MGMT, Oligo-2, GFAP</td>
<td>Ki67, PS3, Vimentin</td>
</tr>
</tbody>
</table>

- Nucleus
- Cytoplasm and/or extracellular
- Both

d

Meningioma-specific Markers

<table>
<thead>
<tr>
<th>CK, PR, EMA, SSTR2</th>
</tr>
</thead>
</table>

- Nucleus
- Cytoplasm and/or extracellular
- Both