# Predicting Primary Site of Secondary Liver Cancer with a Neural Estimator of Metastatic Origin (NEMO)

**Geoffrey F. Schau**[a,b]**, Erik A. Burlingame**[a,b]**, Guillaume Thibault**[b]**, Tauangtham Anekpuritanang**[c,d]**, Ying Wang**[c]**, Joe W. Gray**[b,e]**, Christopher Corless**[c,e,f]**, Young Hwan Chang**[a,b,e,*]

[a]Oregon Health & Science University, Computational Biology Program, Biomedical Engineering Department, Portland, OR, USA
[b]Oregon Health & Science University, OHSU Center for Spatial Systems Biomedicine, Biomedical Engineering Department, Portland, OR, USA
[c]Oregon Health & Science University, Knight Diagnostic Laboratories, Portland, OR, USA
[d]Mahidol University, Department of Pathology, Faculty of Medicine Siriraj Hospital, Bangkok, Thailand
[e]Oregon Health & Science University, Knight Cancer Institute, Portland, OR, USA
[f]Oregon Health & Science University, Department of Pathology, Portland, OR, USA

**Abstract.**

Pathologists rely on clinical information, tissue morphology, and sophisticated molecular diagnostics to accurately infer the metastatic origin of secondary liver cancer. In this paper, we introduce a deep learning approach to identify spatially localized regions of cancerous tumor within hematoxylin and eosin stained tissue sections of liver cancer and to generate predictions of the cancer's metastatic origin. Our approach achieves an accuracy of 90.2% when classifying metastatic origin of whole slide images into three distinct classes, which compares favorably to an established clinical benchmark by three board-certified pathologists whose accuracies ranged from 90.2% to 94.1% on the same prediction task. This approach illustrates the potential impact of deep learning systems to leverage morphological and structural features of H&E stained tissue sections to guide pathological and clinical determination of the metastatic origin of secondary liver cancers.

**Keywords:** Deep Learning, Digital Pathology, Cancer, Liver, Metastasis.

*Young Hwan Chang, chanyo@ohsu.edu

## 1 Introduction

Metastatic liver cancer accounts for 25% of all metastases to solid organs, yet because liver metastases can arise from almost anywhere in the body, accurately determining the origin of metastatic liver cancer is of paramount importance for guiding effective treatment.[1,2] In clinical practice, pathologists commonly rely on clinical information, tissue examination, and molecular assays to determine the metastatic origin of a patient's secondary liver tumor. Although clinically effective, this approach requires significant expertise, experience, and time to perform properly.

Deep learning methods have rapidly accelerated the automation of key processes in identifying and quantifying clinically meaningful features in biomedical images and continues to drive modern advancements in digital pathology.[3,4] Furthermore, deep learning systems have been applied to settings where their performance matches and even exceeds the ability of clinical human practitioners in tasks related to image analysis, including in clinical instances that rely on inspection of hematoxylin and eosin (H&E) stained tissue.[5–9] The emerging power and success of many deep learning approaches applied to image content analysis stem from their ability to learn and leverage meaningful features from large data data collections that cannot be explicitly mathematically modeled.[6,10–12] For example, these approaches can provide robust and reproducible solutions for automated detection and analysis of tumor lesions within whole slide images containing both normal and cancerous tumor tissue segments.[13–15]

Our key contribution in this paper is a deep learning approach to identify metastatic tissue within whole slide section and classify these tumors by their metastatic origin. We evaluate model performance with respect to a clinical benchmark established by three board-certified pathologists charged with the same classification task as our model in which each pathologist was tasked to infer the metastatic origin of liver cancer directly from H&E stained tissue sections without the use of molecular immunohistochemistry assays or clinical data. Through this work, we demonstrate feasibility of deep learning systems to automatically characterize the biological origin of metastatic cancers by their morphological features presented in H&E tissue sections.

## 2 Method

Our approach is composed of two deep neural networks that operate in series. The first stage model is trained to filter tiles containing normal or stromal tissue from whole slide images (WSIs),

2

as these tiles are not expected to have predictive value in estimating the metastatic origin of cancerous tissue. A second stage model is then trained to predict a single label of metastatic origin for each tile in the dataset and aggregate tile predictions within independent WSIs to generate a single whole-slide prediction of metastatic origin. A diagram illustrating the basic workflow of our approach is shown in Fig. 1.

Our approach first leverages pathologist annotation of tumor region to train a binary classifier designed to predict whether a given tile of H&E image is either tumor or non-tumor tissue. A second stage classification model is then trained on the tumor portions of images to correctly predict their metastatic origin with respect to clinically-determined labels. In all cases, model output is reassembled into probabilistic heatmaps over the WSI, enabling a rapid assessment of spatial characteristics driving predictive reasoning. Both first and second stage models utilize the Inception v4 deep learning architecture,[16] which is optimized to capture morphological and architectural features on varying scales with high efficiency and has been shown to achieve human-level prediction capability on the ImageNet dataset. Models were developed in Keras with Tensorflow backend[17] and trained undergoing cyclic learning rates[18] using the Adam optimizer[19] on NVIDIA V100 GPUs made available through the Exacloud HPC resource at Oregon Health & Science University.

Raw H&E images were acquired from the OHSU Knight BioLibrary, uploaded to a secure instance of an OMERO server,[20] programmatically accessed through the OpenSlide python API,[21] normalized with established methods to overcome known inconsistencies in the H&E staining process,[22] and tiled into non-overlapping patches of $299 \times 299$ pixels necessary to accommodate the Inception v4 architecture. Tiles whose mean three-channel, 8-bit intensities were greater than $240$ were filtered out as white non-informative background. Pathological annotations of tumor regions within 28 whole slide H&E images from each of the fourteen metastatic sub-

3

65 types were collected using PathViewer (https://glencoesoftware.com/products/

66 pathviewer/), an interactive utility for the collection and storage of pathological annotation.

67 Tumor region annotations defined the target label for each tile in the WSI dataset as either belong-

68 ing to tumor or non-tumor tissue. We randomly assigned 30% of whole slide images to a held-out

69 test set used for model validation. The code used to generate the results and figures is available in

70 a dedicated Github repository.
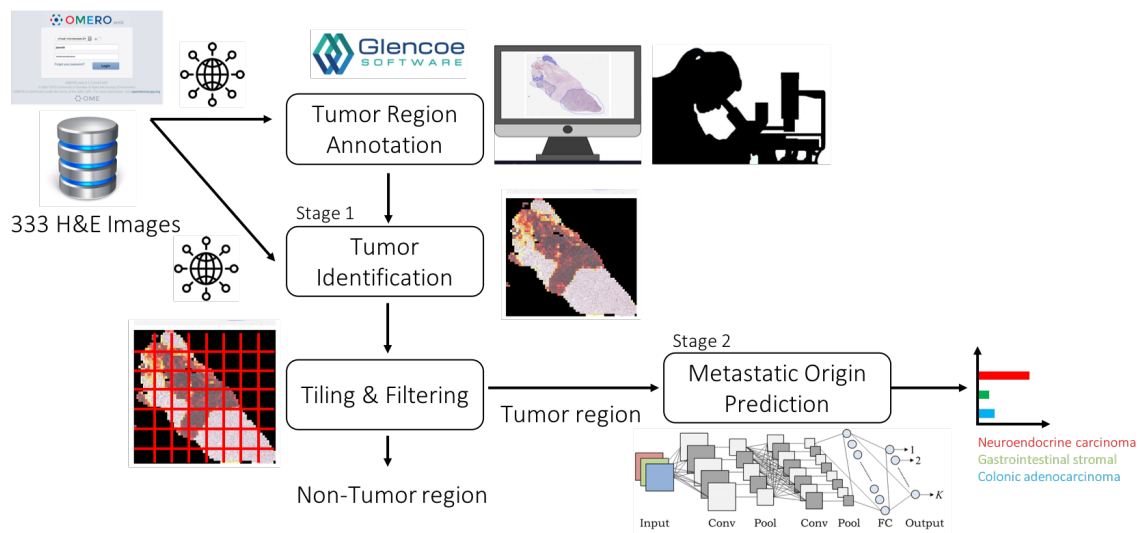


**Fig 1** Deep learning based approach to leverage pathological annotation of tumor region to isolate and localize tumor tissue from a WSI and generate predictions of metastatic origin.

71 ## 3  Results

72 *3.1  Quantitative Localization of Liver Cancer in Whole Slide Images*

73 The first-stage model is a tumor tile binary classifier that generates a prediction between 0 and

74 1 for each tile in the dataset in which a 1 corresponded to perfect confidence that a tile was of

75 tumor tissue and in which a 0 corresponded to perfect confidence that the tile was of normal or

76 stomal tissue. This model achieved an AUC of $0.74$ under the receiver operator characteristics

77 curve which was sufficient to establish good correlation ($R^2 = 0.96$) between clinical estimation

4

78    and our model's estimates of tumor purity in whole slide H&E images as shown in Fig. 2. Further,

79    visual comparisons between the pathological tumor annotation and our model's predictions illus-

80    trate spatial concordance between the drawn tumor-bounding mask and our model's predictions.

81    Once trained, the tumor-region identifying model was deployed on the entire remaining dataset to

82    include only tiles containing cancerous tissue.
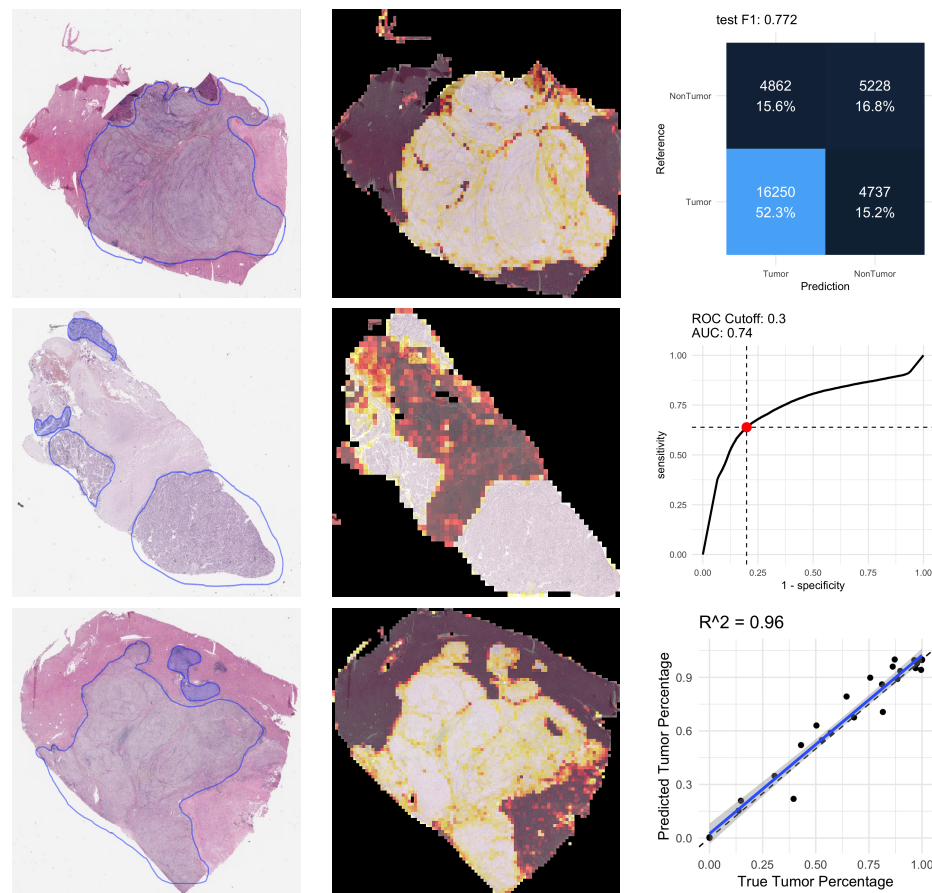


**Fig 2** (left column) Three examples from the held-out testing set with pathological annotation of tumor regions outlined in blue. (center column) Corresponding model predictions estimating regions of whole slide images that contain tumor tissue show concordance between the pathological annotation of tumor region with the outcome of our model. In these illustrations, a brighter color intensity corresponds to higher probability that the underlying tile was labeled as being of tumor by the trained model. (right column, top) Confusion matrix from the held-out testing set for a tumor/non-tumor predictive model illustrating F1 score of 0.772 in the classification task. (right column, center) Receiver Operating Characteristics (ROC) curve illustrating area under the curve of 0.74. (right column, bottom) Comparison between the true tumor purity in the sample inferred from the pathological annotation (x-axis) versus the inferred tumor purity from the model's output (y-axis) with strong correlation ($R^2 = 0.96$)

### 3.2 Quantitative Whole Slide Image Classification of Metastatic Origin

Each image in the dataset is annotated with clinically-determined metastatic origin labels informed by clinical information, pathological inspection of tissue sample, and IHC profiling. These clinical annotations were summarized into 14 distinct subgroups by a clinical practitioner, which are shown in Fig. 3. Although exploratory data analysis considered all fourteen classes, due to class imbalances in this dataset, this work is predicated on the three primarily represented metastatic origins in the dataset (colonic adenocarcinomas, neuroendocrine carcinomas, and gastrointestinal stromal tumors) which collectively represent over 80% of the relevant data.
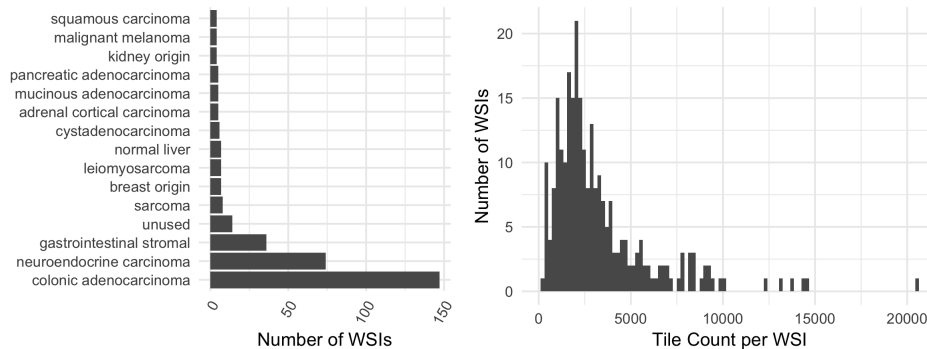


**Fig 3** (left) Summary of the acquired dataset, composed of 333 WSIs each containing metastatic tissue originating from one of 14 sites. (right) Distribution of non-overlapping tile counts in each WSI with mean count 3302 tiles per WSI.

### 3.2.1 Whole Slide Image Classification

After the first stage identifies regions of the H&E images that are tumor, the second stage model learns to classify those tiles according to their metastatic origin. A second Inception v4 deep neural network was designed to generate a three-class prediction for each tile in the training set as belonging to either a colonic adenocarcinoma, gastrointestinal stromal, or neuroendocrine carcinoma. Whole slide image predictions predictions aggregated across all corresponding tumor tiles achieved an F1 score of $0.875$ on the held-out testing set of WSIs, having failed to correctly clas-

6

sify 5 out of the 51 held-out testing samples. Class-specific statistics shown in Table 1 quantify classification performance metrics for the metastatic origin prediction model. Confusion matrices of both WSI and per-tile predictions are shown in Fig. 4.

Several technical factors were associated with incorrect predictions, including slide blurring, tissue folding, and low tumor purity. Our model's confidence was lower for samples that it incorrectly classified, as shown in Fig. 4, though one sample was incorrectly classified with 86% confidence which was driven by misclassified stromal tissue present in teh H&E slide. Individual tiles associated with highly confident predictions for each class are shown in Fig. 5. Pathological inspection of these tiles suggests that tiles associated with highly confident class predictions presentpathological features that guide diagnoses, as the first row contains tiles presenting features associated with primarily spindle-type gastrointestinal stromal tumors and the third row presenting typical well-differentiated neuroendocrine carcinomas. The first two images in the second row represent dirty necrotic tissue which, among the three diseases under consideration, tends to be associated with colonic adenocarcinomas. However, this type of feature is not explicitly associated with cancer, and so should be interpreted with caution. Importantly, this approach obviates the need for pathological region annotation beyond what was required to train the first stage model.

**Table 1** Class-specific statistics of both the tumor identification and three-way origin classification task

| | Sensitivity | Specificity | Pos Pred Value | Neg Pred Value | Precision | Recall | F1 |
|---|---|---|---|---|---|---|---|
| Tumor Identification | 0.77 | 0.52 | 0.77 | 0.53 | 0.77 | 0.77 | 0.77 |
| Colonic Adenocarcinoma | 1.00 | 0.83 | 0.88 | 1.00 | 0.88 | 1.00 | 0.93 |
| Neuroendocrine Carcinoma | 0.79 | 1.00 | 1.00 | 0.92 | 1.00 | 0.79 | 0.88 |
| Gastrointestinal Stromal | 0.78 | 0.98 | 0.87 | 0.95 | 0.88 | 0.78 | 0.82 |
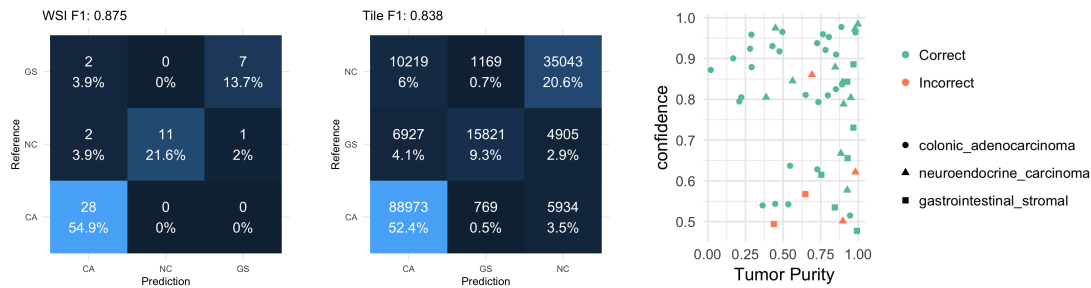
**Fig 4** (left) Confusion matrix of WSI prediction on a held-out test set. (center) Confusion matrix of tile-based predictions. (right) failure cases with respect to the inferred tumor purity in the sample on the x-axis (fraction of tiles predicted to be tumor) and the model's output confidence in its prediction on the y-axis.
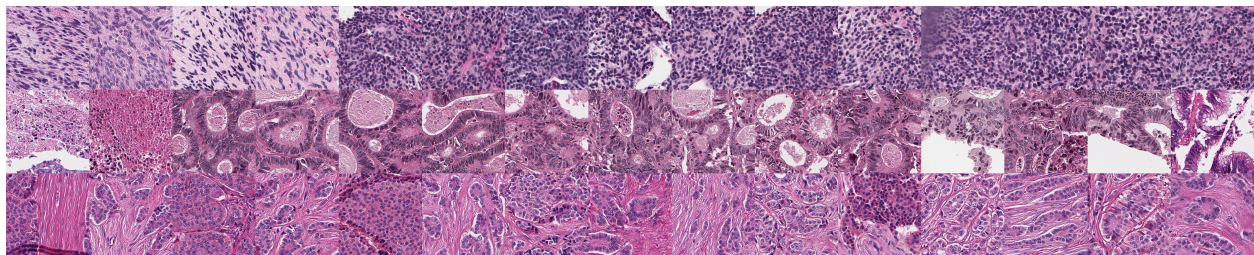


**Fig 5** Example tiles correctly classified by the model with high confidence in which each row is a distinct class (gastrointestinal stromal, colonic adenocarcinoma, and neuroendocrine carcinoma in rows 1, 2, and 3, respectively).

## 3.3 Clinical Benchmark Comparison Study

A study was developed to benchmark our approach to clinical practitioners. This study recruited three board-certified pathologists to independently classify each of the 51 whole slide image samples in the held-out test set according to their metastatic origin. Each participant independently incorrectly classified 3, 4, and 5 samples each, while our neural network model missed 5 samples from the held-out test set. Table 2 summarizes the eleven samples that were missed by either the model or by at least one pathologist and their respective predictions. Interestingly, only two of the mis-classified samples by the model were correctly classified by all three pathologists. Figure 6 illustrates a selected sample classified correctly by the model and all three pathologists, a sample missed by the model that the pathologists all got correct, a sample missed by both the model and at least one pathologist, and a sample for which the model was correct but at least one pathologist made an incorrect classification. All examples illustrate the raw H&E image and three heatmaps

8

126 generated by the model for each of the three-way predictions in which a brighter color corresponds

127 to a higher confidence in the model's prediction for each class. Importantly, predictions are only

128 available for tiles that the first-stage of our model classified as tumor tissue, as non-tumor tiles were

129 filtered out of the metastatic origin prediction task. Although the failure cases are diverse, proba-

130 bilistic overlays of metastatic origin prediction may facilitate faster and more efficient examination

131 of these tissue sections in clinical decision making processes.

**Table 2** Slides mis-classified by either the model or at least one pathologist (GS: Gastrointestinal stromal; CA: Colonic adenocarcinoma; NC: neuroendocrine carcinoma)

| Slide Alias | Ground Truth | Model | Path1 | Path2 | Path3 |
|---|---|---|---|---|---|
| 101 | CA | CA | CA | CA | NC |
| 102 | CA | CA | CA | CA | NC |
| 103 | CA | CA | CA | NC | GS |
| 104 | CA | CA | NC | NC | CA |
| 105 | GS | CA | GS | GS | GS |
| 106 | GS | GS | GS | GS | NC |
| 107 | GS | CA | GS | GS | GS |
| 108 | NC | GS | GS | NC | GS |
| 109 | NC | CA | NC | NC | NC |
| 110 | NC | NC | NC | CA | NC |
| 111 | NC | CA | CA | CA | NC |

132 *3.3.1 Complementary Study: Unsupervised Discovery of Origin-Related Features*

133 As part of our exploratory data analysis, we sought to evaluate whether an unsupervised learning

134 model can learn to identify spatial features of H&E stained tissue tiles that correspond to their

135 metastatic origin. To guide this approach, we employed variational autoencoders (VAE) to learn

136 latent feature representations of metastatic cancers of diverse origins.

137 VAEs are composed of complementary encoder and decoder networks in which the encoder

138 network $\theta$ learns to compress an input data sample into a learned latent representation while the

139 decoder network $\phi$ learns to reconstruct the original input from the latent code. VAEs learn unsu-

9

| Original H&E Alias | CA | GS | NC |
|---|---|---|---|
| 200 (CA) | 97.7 % | 0.7% | 1.5% |
| | | | |
| 109 (NC) | 62.2 % | 1.5% | 36.3% |
| | | | |
| 108 (NC) | 20.3% | 50.2% | 29.5% |
| | | | |
| 106 (GS) | 4.3% | 53.5% | 42.2% |
| | | | |

**Fig 6** Example mis-classified H&E slides with associated annotations from the second stage model illustrating localized predictions of metastatic origin in which brighter colors are associated with more confident class-specific predictions. First row: sample correctly predicted by the model and all three pathologists. Second row: sample missed by the model that all three pathologists got correct. Third row: Example missed by both the model and at least one pathologist. Fourth row: example missed by at least one pathologist that the model got correct. (GS: Gastrointestinal stromal; CA: Colonic Adenocarcinoma; NC: Neuroendocrine Carcinoma)

140   pervised features by minimizing a composite loss function composed of two terms during training.

141   Equation 1 illustrates that VAEs learn to minimize the difference between input sample $x$ and its

142   reconstruction as well as minimize the Kullback-Leibler (KL) divergence of the latent variable $z$

143   distribution with respect to a known prior $p(z)$, which in this case is the standard normal distri-

144   bution. By penalizing divergence between an learned feature distribution and an expected prior,

145   VAEs learn optimal encodings that conform to expected distributions across the dataset and enable

146　a probabilistic interpretation of the embedded feature space.

$$\mathcal{L}_i(x_i, \theta, \phi) = -\mathbb{E}_{z \sim q_\theta(z|x_i)}[\log p_\phi(x_i|z)] + \mathrm{KL}(q_\theta(z|x_i)\|p(z)) \tag{1}$$

147　　A single VAE model was trained to learn latent feature sets of individual tiles composed of four

148　thousand tiles sampled randomly from the fourteen classes presented in the dataset. The learned

149　features space of individual tiles are projected into two-dimensional space with the t-SNE algo-

150　rithm[23] and visualized by their respective metastatic origin in Fig. 7. This unsupervised approach

151　suggests separability of image tiles based on features associated with metastatic origin, suggest-

152　ing feasibility of a supervised classification model to correctly classify the origin of whole slide

153　H&E images of liver metastases. Further, the learned feature space map generates a compelling

154　arrangement of individual tiles grouped by their local feature space similarity. A complementary

155　figure that illustrates an arrangement of tile images onto their respective coordinate projection is

156　available online through FigShare at 10.6084/m9.figshare.8340581.

## 4  Discussion

158　This work presents a deep learning based approach designed to predict the origin of metastatic liver

159　cancer using a two-stage serial model composed of a first model trained to identify tumor from non-

160　tumor within H&E sections of metastatic liver tissue based on pathologists annotation and a second

161　stage model that learns to predict the metastatic origin of individual patches of tumor tissue and

162　aggregates those results into predictions over WSIs. We illustrate through a clinical benchmark

163　comparison that our approach is within performance criteria of board-certified pathologists, sug-

164　gesting that these types of systems may be capable of generating rapid, first-pass assessments of
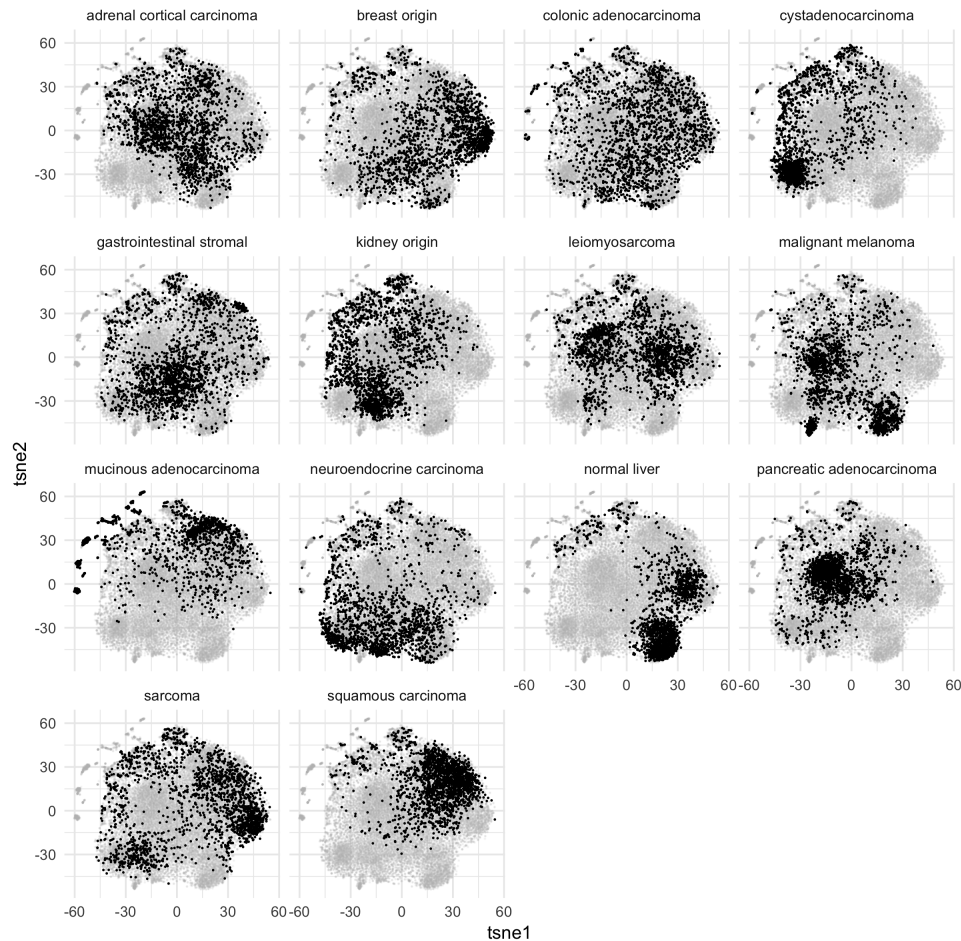
**Fig 7** Learned unsupervised features of H&E tile images from a VAE model projected into two dimensional space and colored by the tiles' metastatic origin illustrate local feature similarity with respect to metastatic origin. Each facet image contains the t-SNE projection of the latent space encoding of each of the tiles in the dataset. That tiles of similar metastatic origin tend to cluster together suggests that morphological features contained within the tiles are informative for estimating where the metastases originated within the body. A higher-resolution figure illustrating input tiles on their respective coordinates is hosted on FigShare (`10.6084/m9.figshare.8340581`).

metastatic origin in the absence of detailed clinical information or comprehensive molecular profiling assay. We believe this type of data-driven visualization augmentation provides an additional layer of information that may facilitate the speed and ease of generating final decisions by clinical care providers.

Although these results illustrate feasibility of our approach, several significant limitations remain. Principally, this analysis was data-limited to only three most-prevalent sources of metastatic

12

origin when in practice metastases can and do originate from a broad variety of biological sources.

A future direction will seek to leverage H&E stained tissue sections of primary disease site and impose a transfer learning approach to predict the primary site of liver cancer *in situ*, without relying on training data drawn exclusively from liver metastases. Secondly, we observe that the first stage model may be inflexible to alternative sites of metastatic tissue. Instead of training a model to identify tiles containing cancer tissue in liver, a more generalizable model may be trained on a broad diversity of primary cancers and regularized appropriately to identify cancer independently of the host tissue. Third, although our model was shown to perform similarly to board-certified pathologists, we have not thoroughly considered the manner by which these types of deep learning models might optimally improve current workflows of practicing pathologists. We believe that robust translation of deep learning systems such as the one presented in this paper may continue to supplement and augment clinical decision-making processes dependent on medical image analysis.

*Disclosures*

No conflicts of interest, financial or otherwise, are declared by the authors.

*Acknowledgments*

192    support of the OHSU Advanced Computing Center.

193    *References*

194    1 A. Ananthakrishnan, V. Gogineni, K. Saeian, *et al.*, "Epidemiology of Primary and Secondary

195      Liver Cancers," *Seminars in Interventional Radiology* **23**(1), 47–63 (2006).

196    2 M. Mohammadian, N. Mahdavifar, and H. Salehiniya, "Liver Cancer in the World: Epidemi-

197      ology, Incidence, Mortality and Risk Factors," *World Cancer Research Journal* **5**(2) (2018).

198    3 U. Djuric, G. Zadeh, K. Aldape, *et al.*, "Precision histology : how deep learning is poised

199      to revitalize histomorphology for personalized cancer care," *npj Precision Oncology* (April),

200      1–4 (2017).

201    4 G. Litjens, C. I. Sánchez, N. Timofeeva, *et al.*, "Deep learning as a tool for increased accu-

202      racy and efficiency of histopathological diagnosis," *Nature Publishing Group* (January), 1–11

203      (2016).

204    5 J. N. Kather, "Predicting survival from colorectal cancer histology slides using deep learning

205      : A retrospective multicenter study," *Plos Medicine* , 1–14 (2019).

206    6 A. Cruz-roa, H. Gilmore, A. Basavanhally, *et al.*, "Accurate and reproducible invasive breast

207      cancer detection in whole- slide images : A Deep Learning approach for quantifying tumor

208      extent," *Nature Publishing Group* (April), 1–14 (2017).

209    7 N. Coudray, P. S. Ocampo, T. Sakellaropoulos, *et al.*, "Classification and mutation prediction

210      from nonsmall cell lung cancer histopathology images using deep learning," *Nature Medicine*

211      , 1 (2018).

212    8 D. Wang and A. Khosla, "Deep Learning for Identifying Metastatic Breast Cancer," *arXiv* ,

213      1–6 (2016).

14

9   J. J. Nirschl, A. Janowczyk, E. G. Peyster, *et al.*, "A deep-learning classifier identifies patients with clinical heart failure using whole-slide images of H & E tissue," *PLoS ONE* , 1–16 (2018).

10  B. Korbar, A. M. Olofson, A. P. Miraflor, *et al.*, "Deep Learning for Classification of Colorectal Polyps on Whole-slide Images," *J Pathol Inform.* **8**(30) (2017).

11  Z. Al-milaji, I. Ersoy, A. Hafiane, *et al.*, "Integrating segmentation with deep learning for enhanced classification of epithelial and stromal tissues in H & E images," *Pattern Recognition Letters* **119**, 214–221 (2019).

12  Y. Zheng, Z. Jiang, F. Xie, *et al.*, "Feature extraction from histopathological images based on nucleus-guided convolutional neural network for breast lesion classification," *Pattern Recognition* **71**, 14–25 (2017).

13  D. A. Dahab, S. S. A. Ghoniemy, and G. M. Selim, "Automated Brain Tumor Detection and Identification Using Image Processing and Probabilistic Neural Network Techniques," *International Journal of Image Processing and Visual Communication* **1**(2), 1–8 (2012).

14  M. Havaei, A. Davy, D. Warde-farley, *et al.*, "Brain tumor segmentation with Deep Neural Networks," *Medical Image Analysis* **35**, 18–31 (2017).

15  T. Ching, D. S. Himmelstein, B. K. Beaulieu-jones, *et al.*, "Opportunities and obstacles for deep learning in biology and medicine," *bioRxiv* (2017).

16  C. Szegedy, S. Ioffe, V. Vanhoucke, *et al.*, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," *arXiv* (2016).

17  M. Abadi, P. Barham, J. Chen, *et al.*, "TensorFlow: A System for Large-Scale Machine Learning," (2016).

18  L. N. Smith, "Cyclical Learning Rates for Training Neural Networks," *arXiv* (April) (2015).

19  D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *ICLR* , 1–15 (2015).

20  C. Allan, J.-M. Burel, J. Moore, *et al.*, "OMERO: flexible, model-driven data management for experimental biology," *Nature Methods* **9**, 245 (2012).

21  A. Goode, B. Gilbert, J. Harkes, *et al.*, "OpenSlide: A vendor-neutral software foundation for digital pathology," *Journal of Pathology Informatics* **4**(1) (2013).

22  M. Macenko, M. Niethammer, J. S. Marron, *et al.*, "A method for normalizing histology slides for quantitative analysis," *ISBI* , 1107–1110 (2009).

23  D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," *arXiv preprint* (Ml), 1–14 (2013).

Biographies and photographs of the authors are not available.