

Generalizing Few-Shot Classification of Whole-Genome Doubling Across Cancer Types*

Sherry Chao

*Department of Biomedical Informatics, Harvard University
Boston, MA, United States
E-mail: schao@g.harvard.edu*

David Belanger

*Google Brain
Cambridge, MA, United States
E-mail: dbelanger@google.com*

The study and treatment of cancer is traditionally specialized to the cancer's site of origin. However, certain phenotypes are shared across cancer types and have important implications for clinical care. To date, automating the identification of these characteristics from routine clinical data - irrespective of the type of cancer - is impaired by tissue-specific variability and limited labeled data. Whole-genome doubling is one such phenotype; whole-genome doubling events occur in nearly every type of cancer and have significant prognostic implications. Using digitized histopathology slide images of primary tumor biopsies, we train a deep neural network end-to-end to accurately generalize few-shot classification of whole-genome doubling across 17 cancer types. By taking a meta-learning approach, cancer types are treated as separate but jointly-learned tasks. This approach outperforms a traditional neural network classifier and quickly generalizes to both held-out cancer types and batch effects. These results demonstrate the unrealized potential for meta-learning to not only account for between-cancer type variability but also remedy technical variability, enabling real-time identification of cancer phenotypes that are too often costly and inefficient to obtain.

Keywords: Histopathology; Cancer genomics; Meta-learning; Few-shot classification.

1. Introduction

Genomic characteristics of a patient's cancer, such as gene mutations and aneuploidy, are increasingly used to improve the course of care.^{1,2} Despite their clinical benefit, however, these characteristics are difficult to measure from routinely-collected patient data. Moreover, identifying the same characteristic across cancer types is made difficult by cancer's inherently heterogeneous nature and the limited size of patient cohorts. There is an unmet need to build tools that automate fast identification of cancer phenotypes from routinely-collected patient data, irrespective of cancer type, particularly when the phenotype (i) has prognostic

*This work is supported by the Simons Foundation and the Google Cloud Research Credits program with the award GCP19980904.

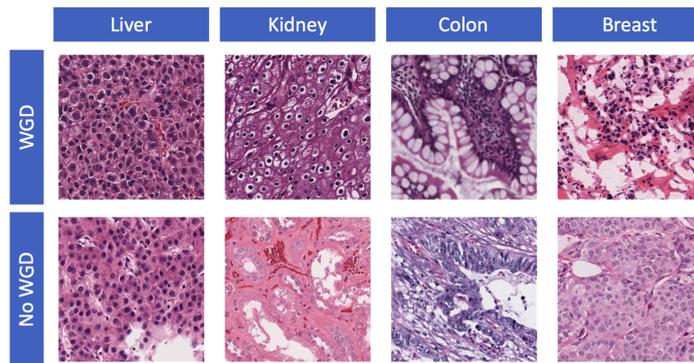


Fig. 1. Representative histopathology image tiles from four cancer types, depicting tumor biopsy samples with (**top**) and without (**bottom**) whole-genome doubling.

or therapeutic implications and (ii) is expensive or slow to measure under traditional means.

Whole-genome doubling (WGD) is one such phenotype. WGD is a genome-wide aberration characterized by the presence of at least twice the normal number of chromosomes and is associated with advanced metastasis and overall poor prognosis.³ Patients with WGD events are more prone to aneuploidy, which lends itself to more aggressive treatment regimens for multiple cancer types.^{4,5} Moreover, WGD itself confers unique vulnerabilities that can be therapeutically targetable.⁶⁻⁹ The prevalence and prognostics of WGD merits knowledge of WGD status in determining the course of care; however, measuring WGD is inefficient. Karyotyping costs \$11k/diagnosis and DNA sequencing costs \$10k/genome, both of which take several weeks to complete.^{10,11} The medical oncology community would significantly benefit from automating WGD identification via more time- and cost-efficient means.

We propose inferring WGD from digitized histopathology images of tumor biopsies, a routinely-collected source of patient data (Figure 1). Across cancer types, the tissue morphology is a manifestation of the genomic characteristics of the tumor. However, histopathology images from different cancer types exhibit tissue-specific characteristics (e.g., colon, lung, skin) even if they share the same WGD status. Traditionally, good performance on cancer-related classification tasks has been achieved via training separate models for each cancer type.¹² This approach has several shortcomings: (i) it necessitates acquiring many training examples from all cancer types, as each model learns from a single cancer type, and (ii) the models are not interchangeable, i.e., a model trained to classify WGD for lung cancer is unable to classify WGD for breast cancer. Successfully integrating machine learning into the clinic necessitates a model that can sufficiently handle inter-cancer diversity.¹³

Recent work by Ref. 14 to classify WGD from histopathology images across cancer types shows good performance on only seven out of 27 cancer types. We propose using meta-learning to automate the classification of WGD across cancer types (Figure 2). In the meta-learning regime, models are learning to learn from few examples. Let us consider a toy example of the standard meta-learning framework. We are given three small datasets: Dataset A contains images of cats/non-cats, Dataset B contains images of dogs/non-dogs, and Dataset C contains images of horses/non-horses. Each dataset has been curated to train a binary classifier on its respective label (i.e., cat, dog, horse). Instead of training on each task individually, however,

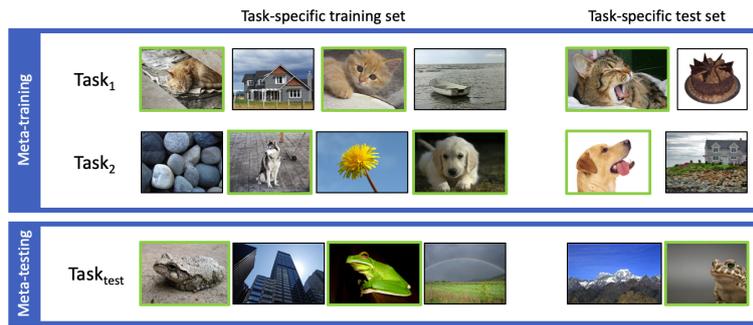


Fig. 2. Overview of meta-learning. The model is trained on multiple tasks (“meta-training”), and at deployment time, the model is presented with a small set of labeled examples and quickly bootstraps a task-specific classifier (“meta-testing”). In contrast, under standard supervised learning, the model remains static at deployment time. Positive-labeled examples are highlighted in green.

we instead train on how to *learn* to learn the tasks from only a few training examples. At meta-training, the model is presented with eight labeled images from each dataset before being asked to correctly classify eight new unlabeled images from each dataset. Therefore, at meta-test time, when the model is given a fourth dataset, Dataset D, which contains images of frogs/non-frogs, it will ideally have learned to learn a new task. Namely, after the model is presented with eight labeled images from Dataset D, it will be able to accurately classify subsequent images from Dataset D as either a frog or a non-frog.

In this work, we adapt the model-agnostic meta-learning (MAML) framework to the problem of WGD classification from histopathology images across multiple cancer types.¹⁵ We take a multi-task view by treating WGD classification for each cancer type as a separate, learnable task. By optimizing for fast learning on each task, the MAML approach is able to outperform a traditional neural network classifier and generalize well to unseen cancer types in the held-out meta-test set. We subsequently extend this approach to accounting for batch effects, or distributional shifts across histopathology images due to technical variation in data collection. Batch effects are pervasive in biomedical datasets. Whereas under standard practices (e.g., fitting a simple model such as linear regression), we mitigate batch effects by incorporating a batch effect-specific term, the complexity of deep neural network classifiers invalidates this solution because the interactions between variables entangles the batch effect with the effect of interest. We show that meta-learning is able to address this problem by treating each batch as its own task-specific dataset such that the model is focused on learning to learn the label (e.g., WGD status) instead of learning specifics about the batch (e.g., image brightness).

Ultimately, we extend the application of meta-learning beyond classifying different labels (one **label** per task) to two novel use cases in medical imaging: classifying the same label across (i) different cancer types (one **cancer type** per task) and (ii) different batches (one **batch** per task). Thus, in a real-world scenario where a clinician would like to quickly classify the WGD status of a particular cancer type or batch, he/she need only label a small handful of histopathology images for the meta-learner to automate labeling of subsequent images.

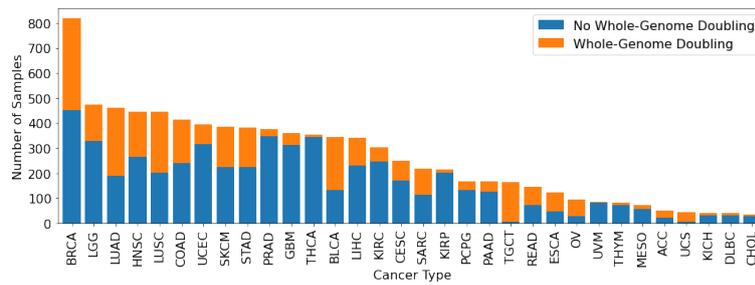


Fig. 3. Summary of the distribution of WGD status by cancer type for samples from The Cancer Genome Atlas. Refer to Appendix A in the Supplementary Materials for acronym descriptions.

2. Related Work

Much effort has been devoted to automating cancer diagnosis by training neural networks to discern tumor from normal tissue in histopathology images.^{16,17} Cancer diagnosis efforts have further delved into cancer subtyping of individual cancer types.^{18,19} Such work has progressed in tandem with detailed tissue segmentation approaches.^{20,21} More recently, increasing efforts are being made in multi-class classification, namely aggregating cancer types in an attempt to accurately diagnose the correct cancer type from all possible cancer types.^{22,23}

Applying machine learning to histopathology images in order to infer characteristics of a patient’s cancer is a growing research area.²⁴ One avenue of applications has focused on predicting survival and prognosis.^{25,26} Another avenue has focused on using morphological features to infer molecular features about a patient’s cancer.^{12,27} Moreover, studies demonstrate we are now able to predict phenotypes such as microsatellite instability and tumor mutational burden from histopathology images of certain cancer types.^{28,29} While these advancements have important therapeutic implications, their applicability is cancer type-specific. The application of meta-learning to medical imaging is relatively nascent. While meta-learning algorithms have been used to classify diabetic retinopathy and rare diseases, meta-learning has unrealized potential for applications in cancer.^{30–33} In this work, we attempt to fill the gap by marrying a meta-learning training regime to a generalizable cancer classification task.

3. Cohort

3.1. Cohort Selection

We obtained diagnostic hematoxylin and eosin (H&E) stained histopathology slides of primary tumor and corresponding WGD status labels for 3,596 samples across 17 cancer types from The Cancer Genome Atlas (TCGA) (Figure 3). We included all samples with available images provided by the National Cancer Institute (NCI) Genomic Data Commons (GDC) Data Portal and slide-level WGD labels provided by Ref. 34. Cancer types were chosen based on the number of available images, selecting for cancer types whose number of images was within one standard deviation of the median number of images (BLCA, COAD, ESCA, HNSC, KIRC, LIHC, READ, STAD, UCEC). To include common cancer types with more than 450 images (BRCA, LUAD, LUSC) while remaining within our storage constraints, we randomly subsampled 25% (BRCA) or 50% (LUAD, LUSC) of the images to yield 200-250 images per common cancer

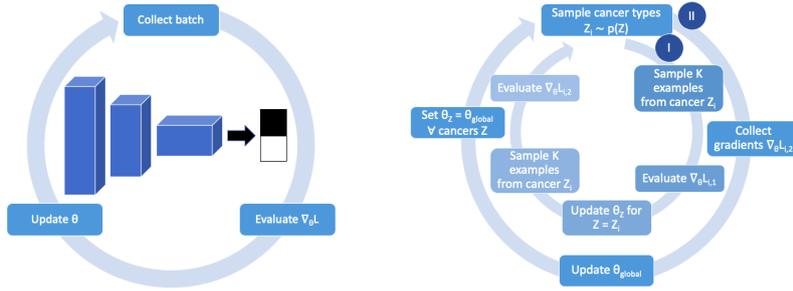


Fig. 4. Comparison of training regimes. While traditional approaches to model optimization optimize solely the global parameters (**left**), meta-learning also optimizes parameters that are local to the particular task (e.g., cancer type or batch) (**right**).

type. Rare cancer types with less than 100 images (ACC, CHOL, KICH, OV, UCS) were included to reflect real-world scenarios with limited labeled data. In total, 42% of slide images, or 1,522 images, were positive for WGD, constituting 19%-82% of images by cancer type. Refer to Appendices A and B in the Supplementary Materials for acronym descriptions and a detailed description of cohort clinical features, respectively.

3.2. Feature Extraction

Since slides are digitized at multiple magnifications, it was important to determine which magnification would be most useful for WGD classification. Preliminary training of a ResNet18-based model on WGD classification of colorectal cancer slides at different magnifications (5x, 10x, 20x) showed the best performance at 10x magnification, which yielded the highest slide-level accuracy and AUC. Thus, we extracted images at 10x magnification. Each histopathology slide was segmented into adjacent, non-overlapping tiles of dimension $3 \times 256 \times 256$ ($C \times H \times W$), wherein only tiles with less than 50% whitespace were retained. On average, each histopathology image was segmented into 3,155 tiles for a cohort-level total of 10.9 million tiles.

4. Methods

4.1. Model

Let data set \mathcal{D} consist of Z cancer types, each of which is comprised of N_z slide images $s_{z,i}$ and binary labels $y_{z,i}$, $i \in \{1, \dots, N_z\}$. As described in Section 3.2, each slide image $s_{z,i}$ is segmented into $T_{z,i}$ non-overlapping tiles $x_{z,i,t}$, $t \in \{1, \dots, T_{z,i}\}$, with C color channels, H pixel height, and W pixel width. Formally, we train the network with the objective: $\theta^* = \underset{\theta}{\operatorname{argmax}} p(Y | S, \theta) = \underset{\theta}{\operatorname{argmax}} \prod_z^Z \prod_i^{N_z} p(y_{z,i} | s_{z,i}, \theta)$ The predictive distribution $p(y | s)$ is parameterized by a deep neural network comprised of a ResNet18 and two fully-connected layers.³⁵ We initialize training with a ResNet18 pre-trained on the ImageNet dataset.³⁶ The pre-trained ResNet18 undergoes additional pre-training on histopathology images sampled from the meta-train set. The final fully-connected layers are trained from scratch on the histopathology images, while the lower layers are fine-tuned to these images. The fully-connected layers have a hidden size of 512 with dropout and tanh activation function. We employ LogSumExp pooling across tiles as a

smooth approximation to max pooling, enabling learning across multiple tiles. LogSumExp pooling of the tile-level prediction scores yields the final slide-level prediction score.

4.2. Training

Model training optimizes the model parameters in order to maximize the log likelihood, yielding the maximum likelihood estimation. Each model is trained on the meta-train set (BLCA, BRCA, COAD, HNSC, LUAD, LUSC, READ, STAD), tuned on the meta-validation set (ESCA, LIHC), and evaluated on the meta-test set (ACC, CHOL, KICH, KIRC, OV, UCS, UCEC). Each cancer type is subsequently split into train and test sets, e.g., for ACC, eight samples are randomly selected for the meta-test train set; the remaining samples are selected for the meta-test test set. Prior to training, tile images are normalized by channel to the mean and standard deviation of the meta-train set. In our baseline scenario (“CNN”), the model is pre-trained on the meta-train set before undergoing meta-validation/-test. In our meta-learning scenario (“MAML”), the model is pre-trained and meta-trained on the meta-train set before undergoing meta-validation/-test. Thus, the models differ solely in training regime while being subject to the same evaluation scheme (Figure 4). Models are trained with – and performance is averaged over – 40 random initializations.

4.2.1. Pre-Training

During pre-training, the model is trained on WGD classification using all cancer types in the meta-train set. Slides and tiles in the train set are shuffled for every epoch and augmented via random vertical/horizontal flips and color jitter. To prevent overfitting, we apply 50% random dropout at each fully-connected layer. We train for up to 200 epochs with a minibatch size of 24 slides (50 randomly sampled tiles per slide) and a learning rate of 0.0001 with an Adam optimizer.³⁷ We reduce the learning rate by a factor of 0.1 upon validation loss plateau with a patience of five epochs. To encourage regularization, model parameters are saved when the binary cross entropy loss on the validation set improves upon that of the previous epoch.

4.2.2. Meta-Training

Meta-training proceeds according to Algorithm 1 using the pre-trained embeddings described in Section 4.2.1. At every iteration, the parameters of the local (cancer type-specific) models are set to the parameters of the global model. We sample a batch of meta-train cancer types \mathcal{Z}_i and a batch of K examples per cancer type \mathcal{Z}_i . Using these examples, we perform one gradient update of the local parameters. Next, we sample a second batch of K examples from each cancer type \mathcal{Z}_i . Using these examples, we perform one forward pass with their respective local models and store the gradient of the loss with respect to the parameters. Once one batch of cancers is complete, the global parameters are updated using the stored gradients.

In all experiments, we meta-train for up to 50 epochs with a learning rate of 0.0001 for both the local and global parameter updates. For each iteration, we sample a batch of 16 slides (50 randomly sampled tiles per slide) from five out of the eight cancer types in the train set, with uniform sampling of the cancer types. Eight slides are used for the local update, and the remaining eight slides are used for the global update. For the local and global parameter

Algorithm 1 Meta-Training

Require: $p(\mathcal{Z})$: distribution over cancer types**Require:** α, η : step size hyperparametersInitialize θ randomly**repeat** Sample batch of cancer types $\mathcal{Z}_i \sim p(\mathcal{Z})$ **for all** \mathcal{Z}_i **do** Sample K examples from $\mathcal{D}_{\mathcal{Z}_i}$ Evaluate $\nabla_{\theta} \mathcal{L}_{\mathcal{Z}_i}(f_{\theta})$ with respect to the K examples Compute adapted parameters using gradient descent: $\theta_i = \theta - \alpha \nabla_{\theta} \mathcal{L}_{\mathcal{Z}_i}(f_{\theta})$ Sample K additional examples from $\mathcal{D}_{\mathcal{Z}_i}$ for the global update **end for** Update $\theta \leftarrow \theta - \eta \nabla_{\theta} \sum_{\mathcal{Z}_i \sim p(\mathcal{Z})} (\mathcal{L}_{\mathcal{Z}_i}(f_{\theta}))$ **until** forever

updates, we use an Adam optimizer and stochastic gradient descent, respectively.³⁷

4.3. Meta-Validation and Meta-Test

Following pre-training and meta-training, we assess the model’s few-shot classification performance using a train set size of eight slides per cancer type. The remaining slides are allocated to the test set. The model takes a fixed number of gradient steps on the meta-validation/-test train set before being evaluated on the meta-validation/-test test set. We employ a learning rate of 0.0001 with an Adam optimizer.³⁷ The amount of dropout d and number of gradient steps g is tuned based on the average meta-validation test set binary cross entropy loss from taking g gradient steps and applying d dropout on the meta-validation train set. Once the optimal hyperparameters are determined, we evaluate the model by taking g gradient steps and applying d dropout on the meta-test train set and measuring WGD classification AUC on the meta-test test set for each cancer type in the meta-test set. In all experiments, the baseline CNN classifier performed optimally with five gradient steps and 0% dropout, and the MAML classifier performed optimally with 20 gradient steps and 25% dropout.

4.4. Experiments4.4.1. *Cancer Types*

To assess the utility of meta-learning for generalizing few-shot WGD classification across cancer types, we applied the meta-training approach by treating WGD classification for each cancer type in the meta-train set as separate tasks. During meta-training, the model is optimized to learn to learn WGD classification for different cancer types that exhibit different tissue morphology. The meta-test test set classification performance of the MAML classifier trained under the meta-learning regime is subsequently compared to that of the baseline CNN classifier trained via standard fine-tuning of a pre-trained deep neural network.

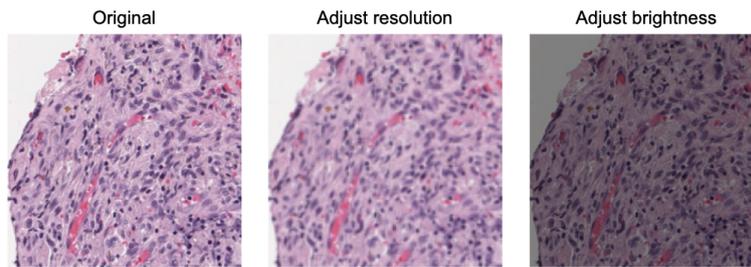


Fig. 5. Given the original unperturbed images (**left**), we assess the ability of the meta-learning training regime to generalize few-shot classification across cancer types when faced with two batch effects: lower resolution (**center**) and lower brightness (**right**).

4.4.2. *Batch Effects*

To assess the utility of meta-learning for generalizing few-shot WGD classification across batches, we applied two transformations to the images in the meta-test set, reflective of real-world technical variations in image capture (Figure 5):

- (1) **Resolution.** Reduction of the effective pixel width and height by 50% to mimic a systematic distributional shift to a lower-resolution input distribution.
- (2) **Brightness.** Reduction of the pixel intensity by 50% to mimic a systematic distributional shift to a dimmer input distribution.

Analogous to the experiment in Section 4.4.1, we applied the meta-training approach by treating WGD classification for each cancer type in the meta-train set as separate tasks. During meta-test, we assess the performance of the MAML classifier and the CNN classifier on the *batch-adjusted* images of the meta-test set.

5. Results

5.1. *Cancer Types*

We evaluate model performance based on the meta-test test set AUC, which compares the model prediction scores to the ground truth binary WGD labels. Table 1 depicts the classification performance on the seven cancer types in the meta-test set. On average, the baseline CNN classifier achieves an AUC of 0.6888, ranging from 0.6611 to 0.7000. In contrast, the MAML classifier achieves an AUC of 0.6944, ranging from 0.6843 to 0.7022, achieving better performance on average than the CNN classifier on five of the meta-test cancer types. Notably, the MAML approach outperforms the baseline CNN approach on four out of the five rare cancer types. Taken together, the MAML approach outperforms the baseline CNN approach on the meta-test set (Wilcoxon signed-rank one-sided p-value=0.0411).

5.2. *Batch Effects*

5.2.1. *Image Resolution*

Table 2 depicts the classification performance on the resolution-adjusted images of the seven cancer types in the meta-test set. On average, the baseline CNN classifier achieves an AUC

Table 1. Results comparing the WGD classification AUC (average \pm 1 std dev) on the held-out meta-test set with baseline standard (CNN) and meta-learning (MAML) training.

	CNN	MAML
ACC	0.6873 \pm 0.0540	0.6988 \pm 0.0581
CHOL	0.6890 \pm 0.0532	0.6845 \pm 0.0643
KICH	0.6928 \pm 0.0312	0.7022 \pm 0.0303
KIRC	0.6611 \pm 0.0609	0.6843 \pm 0.1018
OV	0.6950 \pm 0.0393	0.7020 \pm 0.0435
UCEC	0.7000 \pm 0.0602	0.6859 \pm 0.0816
UCS	0.6846 \pm 0.0387	0.6908 \pm 0.0667
Total	0.6888 \pm 0.0506	0.6944 \pm 0.0773

Table 2. Results comparing the WGD classification AUC (average \pm 1 std dev) on the resolution-adjusted meta-test set with baseline standard (CNN) and meta-learning (MAML) training.

	CNN	MAML
ACC	0.6316 \pm 0.1011	0.6930 \pm 0.0689
CHOL	0.6733 \pm 0.0570	0.6730 \pm 0.0548
KICH	0.7081 \pm 0.0261	0.6941 \pm 0.0285
KIRC	0.6269 \pm 0.0805	0.6401 \pm 0.0855
OV	0.7097 \pm 0.0728	0.7136 \pm 0.0751
UCEC	0.6794 \pm 0.0496	0.6877 \pm 0.0569
UCS	0.6697 \pm 0.0330	0.6649 \pm 0.0886
Total	0.6713 \pm 0.0716	0.6809 \pm 0.0717

of 0.6713, ranging from 0.6269 to 0.7097. In contrast, the MAML classifier achieves an AUC of 0.6809, ranging from 0.6401 to 0.7136, achieving better performance on average than the CNN classifier on four of the meta-test cancer types. Taken together, the MAML approach outperforms the baseline CNN approach on the meta-test set (Wilcoxon signed-rank one-sided p-value=0.0312). Due to the coarse-grained nature of the resolution-adjusted images and associated loss of pixel information, the MAML classifier learns better on the original unperturbed meta-test set images than the resolution-adjusted meta-test set images (Wilcoxon signed-rank one-sided p-value=0.0254). This result is consistent with our feature extraction analysis to identify the optimal magnification, which showed superior performance on 10x magnification images compared to 5x magnification images.

5.2.2. Image Brightness

Table 3 depicts the classification performance on the brightness-adjusted images of the seven cancer types in the meta-test set. On average, the baseline CNN classifier achieves an AUC of 0.6884, ranging from 0.6670 to 0.7101. In contrast, the MAML classifier achieves an AUC of 0.6973, ranging from 0.6837 to 0.7147, achieving better performance on average than the CNN classifier on five of the meta-test cancer types. Notably, the MAML approach outperforms the baseline CNN approach on three out of the five rare cancer types. Taken together, the MAML approach outperforms the baseline CNN approach on the meta-test set (Wilcoxon signed-rank one-sided p-value=0.0370). Moreover, the MAML classifier is able to learn equally well on the original unperturbed and brightness-adjusted meta-test set images, with no significant difference in performance (Wilcoxon signed-rank two-sided p-value=0.9967).

6. Discussion

In this work, we demonstrate that machine learning enables signal extraction from medical imaging data mired in tissue site-specific idiosyncrasies. Unlabeled data is often abundant in healthcare settings because label acquisition is expensive. The meta-learning training regime

Table 3. Results comparing the WGD classification AUC (average \pm 1 std dev) on the brightness-adjusted meta-test set with baseline standard (CNN) and meta-learning (MAML) training.

	CNN	MAML
ACC	0.6742 \pm 0.0801	0.7062 \pm 0.0655
CHOL	0.6819 \pm 0.0475	0.7147 \pm 0.0388
KICH	0.7101 \pm 0.0216	0.7079 \pm 0.0335
KIRC	0.6670 \pm 0.0795	0.7060 \pm 0.0825
OV	0.6989 \pm 0.0774	0.7000 \pm 0.0952
UCEC	0.6899 \pm 0.0312	0.6961 \pm 0.0379
UCS	0.6900 \pm 0.0495	0.6837 \pm 0.0874
Total	0.6884 \pm 0.0620	0.6973 \pm 0.0704

enables fast learning with only a handful of examples. In the case of WGD classification, the MAML classifier outperforms the baseline CNN classifier on the meta-test set when it is trained on only eight training examples per cancer type. In addition, we introduce two batch effects; for every image in the meta-test set, we perturb the image by (i) reducing the brightness by 50%, or (ii) reducing the resolution by 50%. In both cases, the MAML classifier outperforms the baseline CNN classifier on WGD classification.

Ultimately, accounting for variations *between* cancer types is made possible by fast learning on only a handful of labeled images, which successfully extended to accounting for technical variations *within* cancer types. From a clinical perspective, the ability to accurately and cost-effectively stratify patients enables a more fine-grained study of and tailored approach to treatment. From a machine learning perspective, fast adaptation to new tasks is key to mitigating heterogeneity in high-dimensional data that is nonspecific to the signal of interest. As natural extensions of this work, we will expand this analysis to include all 33 cancer types from TCGA and multiple labels beyond WGD status. We will also devise methods that can learn from multiple slices and magnifications to capture intra- and inter-cellular patterns.³⁸ By facilitating a more complete picture of the tumor, we envision these technologies can be seamlessly integrated into the clinic for real-time histology assessment and decision support.³⁹

Software and Data

The code to reproduce all results is publicly available: <https://github.com/chsher/CAML>. TCGA images can be accessed via the NCI GDC Data Portal: <https://portal.gdc.cancer.gov>. WGD labels can be obtained from Ref. 34. Supplementary materials: <https://drive.google.com/file/d/19ps0Vdf891Ay-efZuhDoCei9f2mvDujn/view?usp=sharing>.

References

1. L. Sun, M. E. Marmarelis and C. J. Langer, Systemic therapy for mutation-driven nscl, in *Seminars in Radiation Oncology*, (2)2021.

2. J. Xie, A. Nachabe *et al.*, The prognostic implications of tetraploidy/near-tetraploidy in acute myeloid leukemia, *Leukemia & Lymphoma* **62**, 203 (2021).
3. C. M. Bielski, A. Zehir *et al.*, Genome doubling shapes the evolution and prognosis of advanced cancers, *Nature Genetics* **50**, 1189 (2018).
4. S. M. Dewhurst, N. McGranahan *et al.*, Tolerance of whole-genome doubling propagates chromosomal instability and accelerates cancer genome evolution, *Cancer discovery* **4**, 175 (2014).
5. A. Y. Kuznetsova, K. Seget *et al.*, Chromosomal instability, tolerance of mitotic errors and multidrug resistance are promoted by tetraploidization in human cells, *Cell cycle* **14**, 2810 (2015).
6. R. J. Quinton, A. DiDomizio, M. A. Vittoria, K. Kotýnková, C. J. Ticas, S. Patel, Y. Koga, J. Vakhshoorzadeh, N. Hermance, T. S. Kuroda *et al.*, Whole-genome doubling confers unique genetic vulnerabilities on tumour cells, *Nature* , 1 (2021).
7. A. Vasudevan, K. M. Schukken, E. L. Sausville, V. Girish, O. A. Adebambo and J. M. Sheltzer, Aneuploidy as a promoter and suppressor of malignant growth, *Nature Reviews Cancer* **21**, 89 (2021).
8. Y. Cohen-Sharir, J. M. McFarland, M. Abdusamad, C. Marquis, S. V. Bernhard, M. Kazachkova, H. Tang, M. R. Ippolito, K. Laue, J. Zerbib *et al.*, Aneuploidy renders cancer cells vulnerable to mitotic checkpoint inhibition, *Nature* **590**, 486 (2021).
9. R. W. Sabnis, Novel kif18a inhibitors for treating cancer (2020).
10. Y. Li, L. A. Anderson, E. I. Ginns and J. J. Devlin, Cost effectiveness of karyotyping, chromosomal microarray analysis, and targeted next-generation sequencing of patients with unexplained global developmental delay or intellectual disability, *Molecular diagnosis & therapy* **22**, 129 (2018).
11. K. Schwarze, J. Buchanan *et al.*, The complete costs of genome sequencing: a microcosting study in cancer and rare diseases from a single center in the united kingdom, *Genetics in Medicine* **22**, 85 (2020).
12. J. N. Kather, L. R. Heij, H. I. Grabsch, C. Loeffler, A. Echle, H. S. Muti, J. Krause, J. M. Niehues, K. A. Sommer, P. Bankhead *et al.*, Pan-cancer image-based detection of clinically actionable genetic alterations, *Nature Cancer* **1**, 789 (2020).
13. A. Kleppe, O.-J. Skrede, S. De Raedt, K. Liestøl, D. J. Kerr and H. E. Danielsen, Designing deep learning studies in cancer diagnostics, *Nature Reviews Cancer* , 1 (2021).
14. Y. Fu, A. W. Jung, R. V. Torne, S. Gonzalez, H. Vöhringer, A. Shmatko, L. R. Yates, M. Jimenez-Linan, L. Moore and M. Gerstung, Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis, *Nature Cancer* **1**, 800 (2020).
15. C. Finn, P. Abbeel and S. Levine, Model-agnostic meta-learning for fast adaptation of deep networks, *CoRR* **abs/1703.03400** (2017).
16. C. Zhou, Y. Jin, Y. Chen, S. Huang, R. Huang, Y. Wang, Y. Zhao, Y. Chen, L. Guo and J. Liao, Histopathology classification and localization of colorectal cancer using global labels by weakly supervised deep learning, *Computerized Medical Imaging and Graphics* **88**, p. 101861 (2021).
17. C.-L. Chen, C.-C. Chen, W.-H. Yu, S.-H. Chen, Y.-C. Chang, T.-I. Hsu, M. Hsiao, C.-Y. Yeh and C.-Y. Chen, An annotation-free whole-slide training approach to pathological classification of lung cancer types using deep learning, *Nature communications* **12**, 1 (2021).
18. M. Y. Lu, D. F. Williamson, T. Y. Chen, R. J. Chen, M. Barbieri and F. Mahmood, Data-efficient and weakly supervised computational pathology on whole-slide images, *Nature Biomedical Engineering* , 1 (2021).
19. L. Pantanowitz, G. M. Quiroga-Garza *et al.*, An artificial intelligence algorithm for prostate cancer diagnosis in whole slide images of core needle biopsies: a blinded clinical validation and deployment study, *The Lancet Digital Health* **2**, e407 (2020).
20. D. J. Ho, D. V. Yarlagadda, T. M. DAlfonso, M. G. Hanna, A. Grabenstetter, P. Ntiamoah, E. Brogi, L. K. Tan and T. J. Fuchs, Deep multi-magnification networks for multi-class breast

- cancer image segmentation, *Computerized Medical Imaging and Graphics* **88**, p. 101866 (2021).
21. X. Wang, Y. Fang, S. Yang, D. Zhu, M. Wang, J. Zhang, K.-y. Tong and X. Han, A hybrid network for automatic hepatocellular carcinoma segmentation in h&e-stained whole slide images, *Medical Image Analysis* **68**, p. 101914 (2021).
 22. S. Kalra, H. R. Tizhoosh, S. Shah, C. Choi, S. Damaskinos, A. Safarpour, S. Shafiei, M. Babaie, P. Diamandis, C. J. Campbell *et al.*, Pan-cancer diagnostic consensus through searching archival histopathology images using artificial intelligence, *NPJ digital medicine* **3**, 1 (2020).
 23. A. Riasatian, M. Babaie, D. Maleki, S. Kalra, M. Valipour, S. Hemati, M. Zaveri, A. Safarpour, S. Shafiei, M. Afshari *et al.*, Fine-tuning and training of densenet for histopathology image representation using tcga diagnostic slides, *Medical Image Analysis* , p. 102032 (2021).
 24. A. Echle, N. T. Rindtorff, T. J. Brinker, T. Luedde, A. T. Pearson and J. N. Kather, Deep learning in cancer pathology: a new generation of clinical biomarkers, *British Journal of Cancer* , 1 (2020).
 25. W. Bulten, M. Balkenhol, J.-J. A. Belinga, A. Brillhante, A. Çakır, L. Egevad, M. Eklund, X. Farré, K. Geronatsiou, V. Molinié *et al.*, Artificial intelligence assistance significantly improves gleason grading of prostate biopsies by pathologists, *Modern Pathology* , 1 (2020).
 26. A. Saito, H. Toyoda *et al.*, Prediction of early recurrence of hepatocellular carcinoma after resection using digital pathology images assessed by machine learning, *Modern Pathology* **34**, 417 (2021).
 27. D. Bychkov, N. Linder *et al.*, Deep learning identifies morphological features in breast cancer predictive of cancer erbb2 status and trastuzumab treatment efficacy, *Scientific reports* **11**, 1 (2021).
 28. A. Echle, H. I. Grabsch, P. Quirke, P. A. van den Brandt, N. P. West, G. G. Hutchins, L. R. Heij, X. Tan, S. D. Richman, J. Krause *et al.*, Clinical-grade detection of microsatellite instability in colorectal tumors by deep learning, *Gastroenterology* **159**, 1406 (2020).
 29. H. Xu, S. Park, J. R. Clemenceau, N. Radakovich, S. H. Lee and T. H. Hwang, Deep learning approach to predict tumor mutation burden (tmb) and delineate its spatial heterogeneity from whole slide images, *bioRxiv* , p. 554527 (2020).
 30. S. Hu, J. Tomczak and M. Welling, Meta-learning for medical image classification, in *1st Conference on Medical Imaging with Deep Learning*, 2018.
 31. X. Li, L. Yu, Y. Jin, C.-W. Fu, L. Xing and P.-A. Heng, Difficulty-aware meta-learning for rare disease diagnosis, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2020.
 32. O. Gevaert, Meta-learning reduces the amount of data needed to build ai models in oncology, *British Journal of Cancer* , 1 (2021).
 33. S. I. Garcia, Meta-learning for skin cancer detection using deep learning techniques, *arXiv preprint arXiv:2104.10775* (2021).
 34. A. M. Taylor, J. Shih *et al.*, Genomic and functional approaches to understanding cancer aneuploidy, *Cancer cell* **33**, 676 (2018).
 35. K. He, X. Zhang, S. Ren and J. Sun, Deep residual learning for image recognition, *CoRR abs/1512.03385* (2015).
 36. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in *2009 IEEE conference on computer vision and pattern recognition*, 2009.
 37. D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980* (2014).
 38. J. T. Liu, A. K. Glaser, K. Bera, L. D. True, N. P. Reder, K. W. Eliceiri and A. Madabhushi, Harnessing non-destructive 3d pathology, *Nature biomedical engineering* , 1 (2021).
 39. P.-H. C. Chen, K. Gadepalli *et al.*, An augmented reality microscope with real-time artificial intelligence integration for cancer diagnosis, *Nature medicine* **25**, 1453 (2019).