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.\textsuperscript{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.

© 2021 The Authors. Open Access chapter published by World Scientific Publishing Company and distributed under the terms of the Creative Commons Attribution Non-Commercial (CC BY-NC) 4.0 License.
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. Patients with WGD events are more prone to aneuploidy, which lends itself to more aggressive treatment regimens for multiple cancer types. Moreover, WGD itself confers unique vulnerabilities that can be therapeutically targetable. 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. 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. 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.

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,

![Figure 1](image.png)

Fig. 1. Representative histopathology image tiles from four cancer types, depicting tumor biopsy samples with (top) and without (bottom) whole-genome doubling.
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. 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.\textsuperscript{16,17} Cancer diagnosis efforts have further delved into cancer subtyping of individual cancer types.\textsuperscript{18,19} Such work has progressed in tandem with detailed tissue segmentation approaches.\textsuperscript{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.\textsuperscript{22,23}

Applying machine learning to histopathology images in order to infer characteristics of a patient’s cancer is a growing research area.\textsuperscript{24} One avenue of applications has focused on predicting survival and prognosis.\textsuperscript{25,26} Another avenue has focused on using morphological features to infer molecular features about a patient’s cancer.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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).

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 $D$ consist of $Z$ cancer types, each of which is comprised of $N_z$ slide images $s_{z,i}$ and binary labels $y_{z,i}, i \in \{1, \ldots, N_z\}$. As described in Section 3.2, each slide image $s_{z,i}$ is segmented into $T_{z,i}$ non-overlapping tiles $x_{z,i,t}, t \in \{1, \ldots, T_{z,i}\}$, with $C$ color channels, $H$ pixel height, and $W$ pixel width. Formally, we train the network with the objective: 

$$
\theta^* = \arg\max_{\theta} p(Y | S, \theta) = \arg\max_{\theta} \prod_{z}^{Z} \prod_{i}^{N_z} p(y_{z,i} | s_{z,i}, \theta) $$

The predictive distribution $p(y | s)$ is parameterized by a deep neural network comprised of a ResNet18 and two fully-connected layers. We initialize training with a ResNet18 pre-trained on the ImageNet dataset. 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. 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 $Z_i$ and a batch of $K$ examples per cancer type $Z_i$. 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 $Z_i$. 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: \( p(Z) \): distribution over cancer types

Require: \( \alpha, \eta \): step size hyperparameters

Initialize \( \theta \) randomly

repeat
  Sample batch of cancer types \( Z_i \sim p(Z) \)
  for all \( Z_i \) do
    Sample K examples from \( D_{Z_i} \)
    Evaluate \( \nabla_\theta \mathcal{L}_{Z_i}(f_\theta) \) with respect to the K examples
    Compute adapted parameters using gradient descent: \( \theta_i = \theta - \alpha \nabla_\theta \mathcal{L}_{Z_i}(f_\theta) \)
    Sample K additional examples from \( D_{Z_i} \) for the global update
  end for
  Update \( \theta \leftarrow \theta - \eta \nabla_\theta \sum_{Z_i \sim p(Z)} (\mathcal{L}_{Z_i}(f_\theta)) \)
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. Experiments

4.4.1. 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.
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 ± 1 std dev) on the held-out meta-test set with baseline standard (CNN) and meta-learning (MAML) training.

<table>
<thead>
<tr>
<th></th>
<th>CNN</th>
<th>MAML</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC</td>
<td>0.6873 ± 0.0540</td>
<td>0.6988 ± 0.0581</td>
</tr>
<tr>
<td>CHOL</td>
<td>0.6890 ± 0.0532</td>
<td>0.6845 ± 0.0643</td>
</tr>
<tr>
<td>KICH</td>
<td>0.6928 ± 0.0312</td>
<td>0.7022 ± 0.0303</td>
</tr>
<tr>
<td>KIRC</td>
<td>0.6611 ± 0.0609</td>
<td>0.6843 ± 0.1018</td>
</tr>
<tr>
<td>OV</td>
<td>0.6950 ± 0.0393</td>
<td>0.7020 ± 0.0435</td>
</tr>
<tr>
<td>UCEC</td>
<td>0.7000 ± 0.0602</td>
<td>0.6859 ± 0.0816</td>
</tr>
<tr>
<td>UCS</td>
<td>0.6846 ± 0.0387</td>
<td>0.6908 ± 0.0667</td>
</tr>
<tr>
<td>Total</td>
<td>0.6888 ± 0.0506</td>
<td>0.6944 ± 0.0773</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>CNN</th>
<th>MAML</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC</td>
<td>0.6316 ± 0.1011</td>
<td>0.6930 ± 0.0689</td>
</tr>
<tr>
<td>CHOL</td>
<td>0.6733 ± 0.0570</td>
<td>0.6730 ± 0.0548</td>
</tr>
<tr>
<td>KICH</td>
<td>0.7081 ± 0.0261</td>
<td>0.6941 ± 0.0285</td>
</tr>
<tr>
<td>KIRC</td>
<td>0.6269 ± 0.0805</td>
<td>0.6401 ± 0.0855</td>
</tr>
<tr>
<td>OV</td>
<td>0.7097 ± 0.0728</td>
<td>0.7136 ± 0.0751</td>
</tr>
<tr>
<td>UCEC</td>
<td>0.6794 ± 0.0496</td>
<td>0.6877 ± 0.0569</td>
</tr>
<tr>
<td>UCS</td>
<td>0.6697 ± 0.0330</td>
<td>0.6649 ± 0.0886</td>
</tr>
<tr>
<td>Total</td>
<td>0.6713 ± 0.0716</td>
<td>0.6809 ± 0.0717</td>
</tr>
</tbody>
</table>

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 ± 1 std dev) on the brightness-adjusted meta-test set with baseline standard (CNN) and meta-learning (MAML) training.

<table>
<thead>
<tr>
<th></th>
<th>CNN</th>
<th>MAML</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC</td>
<td>0.6742 ± 0.0801</td>
<td>0.7062 ± 0.0655</td>
</tr>
<tr>
<td>CHOL</td>
<td>0.6819 ± 0.0475</td>
<td>0.7147 ± 0.0388</td>
</tr>
<tr>
<td>KICH</td>
<td><strong>0.7101 ± 0.0216</strong></td>
<td>0.7079 ± 0.0335</td>
</tr>
<tr>
<td>KIRC</td>
<td>0.6670 ± 0.0795</td>
<td>0.7060 ± 0.0825</td>
</tr>
<tr>
<td>OV</td>
<td>0.6989 ± 0.0774</td>
<td>0.7000 ± 0.0952</td>
</tr>
<tr>
<td>UCEC</td>
<td>0.6899 ± 0.0312</td>
<td>0.6961 ± 0.0379</td>
</tr>
<tr>
<td>UCS</td>
<td>0.6900 ± 0.0495</td>
<td>0.6837 ± 0.0874</td>
</tr>
<tr>
<td>Total</td>
<td>0.6884 ± 0.0620</td>
<td><strong>0.6973 ± 0.0704</strong></td>
</tr>
</tbody>
</table>

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. 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.

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/19psOVdf89lAy-efZuhDoCe19f2mvDujn/view?usp=sharing.

References
3. C. M. Bielski, A. Zehir et al., Genome doubling shapes the evolution and prognosis of advanced cancers, Nature Genetics 50, 1189 (2018).
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).
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).


