Optimize Deep Learning Models for Prediction of Gene Mutations Using Unsupervised Clustering

Author: Zihan Chen1*, Xingyu Li1*, Miaomiao Yang2, Hong Zhang1†, Xu Steven Xu3*

1 Department of Statistics and Finance, School of Management, University of Science and Technology of China;
2 Clinical Pathology Center, The Fourth Affiliated Hospital of Anhui Medical University, Hefei, Anhui, China
3 Data Science/Translational Research, Genmab Inc., Princeton, New Jersey, USA;

Abstract

Deep learning has become the mainstream methodological choice for analyzing and interpreting whole-slide digital pathology images (WSIs). It is commonly assumed that tumor regions carry most predictive information. In this paper, we proposed an unsupervised clustering-based multiple-instance learning, and apply our method to develop deep-learning models for prediction of gene mutations using WSIs from three cancer types in The Cancer Genome Atlas (TCGA) studies (CRC, LUAD, and HNSCC). We showed that unsupervised clustering of image patches could help identify predictive patches, exclude patches lack of predictive information, and therefore improve prediction on gene mutations in all three different cancer types, compared with the WSI based method without selection of image patches and models based on only tumor regions. Additionally, our proposed algorithm outperformed two recently published baseline algorithms leveraging unsupervised clustering to assist model prediction. The unsupervised-clustering-based approach for mutation prediction allows identification of the spatial regions related to mutation of a specific gene via the resolved probability scores, highlighting the heterogeneity of a predicted genotype in the tumor
microenvironment. Finally, our study also demonstrated that selection of tumor regions of WSIs is not always the best way to identify patches for prediction of gene mutations, and other tissue types in the tumor micro-environment may provide better prediction ability for gene mutations than tumor tissues.
Introduction

The diagnosis of cancer is typically based on histopathological assessment of tissue sections, and supplemented by genetic and other molecular tests [1-6]. The search for molecular biomarkers and gene mutations has made a great impact on development of novel treatment options. For example, KRAS mutations, present in about 30% to 50% of colorectal cancers (CRCs), have been shown to be associated with poor prognosis and advanced disease [7-13]. In lung adenocarcinoma (LUAD), EGFR has been reported to be mutated in about 20% of LUAD patients, and multiple EGFR targeted therapies have been developed and approved by the Food and Drug Administration (FDA) [14, 15]. Due to the high turnaround time, tissue usage and costs in the current oncology workflows [16], there is a growing need to employ new, inexpensive, and scalable approaches in medical oncology.

Deep learning-based computer vision algorithms have been developed to predict gene mutations based on whole-slide images (WSIs) [17-23]. Coudray et al. developed deep convolutional neural network (DeepPATH) to predict gene mutations in LUAD based on WSIs [24]. Kather et al. proposed, optimized, and extensively validated a one-stop-shop workflow based on a lightweight neural network, ShuffleNet [23]. They showed that a wide range of genetic mutations, molecular tumor subtypes, gene expression signatures, and standard pathology biomarkers could be inferred from histopathology tissue slides.

Because a large number of image patches (e.g. from hundreds to thousands) are available for each WSI, and not all areas of a WSI are relevant to gene mutations, direct use of all image
patches of a WSI in a prediction model may not produce optimal prediction performance. It has been postulated that certain image regions/patches within the WSI (e.g., tumor) could carry more predictive value and improve the prediction performance. A common approach is to use pathologists to annotate regions relevant to the diagnostic task (e.g., tumor) and train the model on the labeled regions [23, 25-31]. Scientists also trained tissue classifiers (tumor and nontumor) to select tumor-like tiles to predict mutated genes in different cancers [18, 24, 32, 33].

In the field of digital pathology, unsupervised clustering has been widely used to reduce the dimensionality of patches to facilitate multiple instance learning (e.g., patches from WSIs can be fit on a GPU at once) [34], to derive additional cluster-based features, and to identify rare events [35]. Dooley et al. [35] and Zhu et al. [36] clustered patches into different clusters, and used the frequency of patches in each cluster as a new feature of WSIs for heart transplant rejection classification. Similarly, Abbet [34] proposed a self-supervised learning method that jointly learns a representation of tissue regions as well as a metric of the clustering to obtain their underlying spatial features of tissues (e.g., cluster probabilities and cluster transition probabilities).

In addition, unsupervised clustering has been commonly used in image-based deep-learning survival analysis. Yao et al. clustered the patches in each WSI individually into different phenotype clusters, and sampled one patch from each cluster to represent the WSI to predict the survival of colorectal cancer (CRC) patients [37]. Sharma et al. deployed a local cluster-based (clustering patches from a single WSI) sampling approach for classifying
children with celiac disease and healthy children [38]. Zhu et al. and Yue et al. performed global clustering of the patches from all patients and training a survival model for each cluster. The features from the most predictive clusters were then aggregated across the patches from each cluster to predict the outcome. Muhammad et al. used patch features grouped by global centroids to calculate local slide-level centroid and concatenated the nearest patches to local centroids to represent each slide and trained with survival data. Their approach performed better than other approaches in modelling intrahepatic cholangiocarcinoma (ICC) [39].

Despite of the wide range of applications of unsupervised clustering in digital pathology, very few studies have been performed to use unsupervised clustering to select and identify image patches to optimize the prediction of gene mutations. In this study, we show that unsupervised clustering can enhance the model prediction ability by dividing the image into different areas, assisting in tissue type annotation, and selecting the most predictive image patches. By applying our method to WSIs from three cancer types in The Cancer Genome Atlas (TCGA) studies (CRC, LUAD, and HNSCC), we demonstrate that unsupervised clustering improved and optimized prediction on gene mutations in different cancer types. In addition, clustering of image patches helped identify predictive patches with greater association with mutation of genes and exclude patches with no predictive values. Furthermore, unsupervised clustering assisted annotation of tissue types could also facilitate understanding the morphologic features associated with certain gene mutations.

Methods
**TCGA dataset**

The Cancer Genome Atlas (TCGA) WSI dataset was available at the TCGA portal (https://portal.gdc.cancer.gov). TCGA gene mutation data and subtype data were downloaded from the website https://xenabrowser.net/datapages/. Datasets were obtained for three tumor types, namely colorectal (COAD and READ), head and neck squamous cell carcinoma (HNSCC), and lung adenocarcinoma (LUAD).

**Image preprocessing**

To exclude the background area of each hematoxylin and eosin (H&E) stained whole slide image (WSI) where no tissue was present, we adapted Otsu’s method [40] to separate pixels in each image into two classes on greyscale space, foreground, and background. By removing the background areas, we obtained the areas of the image that contained the tissue. The tissue areas of the image were then tiled into small non-overlapping patches each with dimension of 224 × 224 pixels. Macenko’s method was used to normalize the color of patches synchronously (Supplementary Figure 1) [41]. For LUAD, HNSCC, and CRC, the number of patches extracted from the images ranged from <100 to 50,000 (average = 12,664), >100 to 30,000 (average = 12,772), and a few hundred to 30,000 (average = 7,888), respectively.

**Feature extraction**

We employed a fine-tuned Xception model to extract features of image patches [42], which was pretrained on the ImageNet dataset, and fine-tuned on a CRC dataset [43]. A feature vector with a dimension of 256 was extracted from each patch. For each patient, a n × 256 feature
matrix (n = the number of patches of that patient) was obtained.

*Unsupervised clustering*

We randomly selected 100 patients from each of the three tumor types (LUAD, HNSCC, and CRC) in TCGA dataset. After pooling all patches from the 100 patients of each cancer type, we used K-means clustering to cluster these patches into four groups. To our knowledge, we are the first to predict mutations using unsupervised clustering and only a few papers have explored the effect of the number of clusters on classification tasks using WSI. Sharma et al. presented the impact of cluster numbers on the performance of classification of celiac disease vs. histologically normal [38]. Referring to their results and the trade-off between computational efficiency and clustering performance on deep learning platform, we used cluster numbers k=4. Finally, a k-NN algorithm was used to assign cluster labels to the rest of patches of that cancer type, which were not included in the process of building the clustering model.

*Clinically relevant gene mutations*

Clinically relevant genes for each cancer type [44, 45] were selected for the analysis (Table 1). We assigned each patch with label 1 or 0, depending on the presence or absence of the mutation in that patient.

*Best-cluster optimized multiple-instance learning (BCOMIL)*

To study the effect of clustering, on each cluster we trained a patch-level multilayer perceptron (MLP) classifier that used the features of the patches only in each cluster as the input to estimate
the mutation probability of each patch. The Adam algorithm was used to optimize the cross entropy. After averaging the predicted probability, we obtained a slide-level classifier. The test was then achieved on test slides in TCGA dataset. Figure 1 shows the pipeline of our proposed method.

**Comparison to WSI-based approach without unsupervised clustering**

As a benchmark comparison, we also trained an MIL classifier using all patches from patients as input without clustering the patches. Briefly, all the patches from a WSI were used to train a patch-level network. The average predicted probability of patches was used to predict slide-level mutation [24].

**Comparison to approach based on tumor regions (CRC)**

Patches from tumor areas have often been used to train prediction models for gene mutations [18, 24, 32, 33]. For CRC, we selected tumor tiles using a fine-tuned Xception-based tissue-type classifier [42], trained mutation prediction models on tumor patches, and compared the performance of these models with BCOMIL-based models. Since patch-level labels for tissue types were only available for CRC, we compared these two methods only on CRC cohort.

**Comparison to other published baseline algorithms**

In addition, we compared our method with two recently published baseline methods that also utilized unsupervised clustering to assist their prediction models.

1. The method of Dooley et al. [35]: Use unsupervised clustering to group patches into
different clusters, and then use the frequency of patches of each cluster as a new feature for classification.

2. The method of Zhu et al. [46]: Zhu et al. developed an algorithm to use unsupervised clustering to assist development of a deep-learning survival model. The main idea of this method is using K-means cluster instead of manual annotated region, and selecting the effective cluster for the final prediction. We adapted several steps similar to those of Zhu et al. method. Briefly, we (1) performed global clustering of the patches from all patients, (2) trained a separate MLP classifier for each cluster, (3) selected the clusters with average AUC values from the 5-fold cross validation greater than 0.5 on that gene and (4) calculated the weight for cluster \( j \) in patient \( i \) as,

\[
w_{ij} = \frac{n_{ij}}{n_i}, i \in \{1, \ldots, N\}, j \in \{1, \ldots, J\},
\]

where \( n_{ij} \) represented number of patches patient \( i \) had in cluster \( j \) and \( n_i \) represented patient \( i \) total patches.

Similar to the Zhu et al. method, the features for that patient in the selected cluster were calculated as,

\[
x_{ij} = w_{ij} \sum_{k=1}^{K} x_{ijk}/K,
\]

where \( x_{ij} \) is the output features in cluster \( j \) for patient \( i \). We concatenated weighted feature from selected clusters, then trained a SVM classifier for mutation prediction.

**Experiments**

In all experiments, we used stratified five-fold cross validation according to the mutation status for each gene to divide TCGA dataset into five folds for each cancer type and to avoid data
imbalance problem. In each fold, 80% of the data were used for model training and 20% of the data were used for model validation. During training, we used the Adam optimization with initial learning rate of 0.00005, a cosine annealing schedule with maximum number of iterations of 20 and training and validation were done over 1000 iterations. We evaluated the model performances with area under the ROC curve (AUC). When calculating cross entropy loss, we assigned more weight to classes with a small number of training images so that the network was punished more if it falsely predicted the labels of these classes.

**Annotation of image patches**

1. Automated annotation with tissue-type classifier for CRC

Kather et al. developed a deep-learning classifier to classify CRC image tiles into eight tissue types: adipose tissue (ADI), background (BACK), debris (DEB), lymphocytes (LYM), mucus (MUC), smooth muscle (MUS), normal colon mucosa (NORM), cancer-associated stroma (STR), and colorectal adenocarcinoma epithelium (TUM) [47]. We used pathologist annotated NCT-CRC-HE-100K and CRC-VAL-HE-7K image sets provided by Kather et al. to train and validate, respectively, a similar tissue-type classifier to identify the tissue types for CRC images [42]. The overall accuracy of the tissue-type classification model was 99% based on the training dataset NCT-CRC-HE-100K and 94.4% based on the validation image set CRC-VAL-HE-7K (Supplementary Figure 2). The tissue type of each image tile from the TCGA dataset was predicted using the fine-tuned Xception-based tissue-type classifier.

2. Semi-automatic annotation

For LUAD and HNSCC, we developed a semi-automatic procedure to assist pathologists’
annotation of the image patches from each cluster. For each cluster of LUAD and HNSCC, an additional K-means clustering (k =4) was performed. Then, we selected four neighborhood patches around the center of each subcluster for pathologists’ annotation. In total, the pathologist annotated 64 patches (out of a total of 5,496,176 patches for LUAD, 5,504,732 patches for HNSCC, and 3,265,632 patches for CRC; 4 clusters x 4 subclusters x 4 neighborhood patches around the center of each subcluster) for each cancer type. We used CRC to confirm that the semi-automatic annotation approach could identify similar tissue types compared to the tissue-type classifier for CRC.

Results

Tissue clusters in LUAD, HNSCC, and CRC

Figure 3 shows that K-means clustered the tiles into four distinct clusters for the three TCGA datasets (LUAD, HNSCC, and CRC). Figures 4-6 show the image tiles representing the most common tissue type among the four neighborhood patches near the center of each subcluster (four subclusters for each cluster) for LUAD, HNSCC, and CRC, respectively. For LUAD, Cluster 2 mainly consisted of tumor tissues, while Cluster 4 primarily included stromal cells. Clusters 1 and 3 were a mix of red blood cells, stromal cells, pulmonary alveolus, tumor, lymphocytes, proliferating fibroblasts and other non-tumoral cells (Figure 4). For HNSCC, Clusters 3 and 4 were the non-tumor and tumor compartments, respectively. Cluster 1 of HNSCC was a mix of lymphocytes and tumor cells, while Cluster 2 comprised mostly non-tumor cells with some tumor cells (Figure 5).
For CRC, we examined the tissue types of each cluster using the CRC tissue type classifier [47]. Figures 6 and 7 shows that Cluster 1 was mainly consist of tumor and mucin cells, while almost all patches in Cluster 2 were tumor cells. Cluster 3 of CRC primarily included muscular and stromal cells with some debris and tumor cells as well. A wide range of tissue types were present in Cluster 4, which included all 8 tissue types (plus small amount of background patches) in the classifier (Figure 7). It is worth mentioning that the annotations based on the tissue classifier were generally consistent with manual annotations provided by a pathologist (Figure 6a and 6b).

**Prediction of gene mutations by tissue clusters**

Tables 2-4 illustrate the average AUC values from the five-fold cross validation using the TCGA datasets (LUAD, HNSCC, and CRC, respectively) for the four prediction models based on the image tiles from the four individual clusters. It is obvious that image tiles from different clusters had different predictive ability. For LUAD (Table 2), tumor cells (Cluster 2) clearly provided the best prediction for TP53 and STK11, suggesting that the mutant-like image features for TP53 and STK11 may be mainly contained in the tumor regions (refer to the heatmap in Figure 8). This is consistent with the findings from Coudray et al. [24]. Tumor cells also predicted EGFR mutation well (Table 2). Stromal cells in Cluster 4 provided the highest AUC for prediction of ALK gene mutation (Table 2) and the image tile with the highest likelihood of ALK mutation demonstrated stromal features (Figure 8). Models based on image tiles from Clusters 1 and 3 (a mix of red blood cells, stromal cells, interalveolar septum cells, and other non-tumoral cells) provided the best prediction for KEAP1 and KRAS respectively.
Similarly, for HNSCC, a remarkable difference in AUC for prediction of gene mutations was observed for different clusters (Table 3). The tumor cells (Cluster 4) provided high predictive ability for TP53, HRAS, CASP8, and NSD1. The heatmap in Figure 9 shows that the TP53-mutant like features are highly present in the tumor compartment of HNSCC. Interestingly, non-tumor cells in Cluster 3 best predicted the mutation of NSD1, while the mix of lymphocytes and tumor cells in Cluster 1 best predicted the CASP8 mutation. The image tiles with the highest likelihood of NSD1 and CASP8 were non-tumoral cells (Figure 9). The models based on the non-tumor cells (including blood vessel, debris, etc.) in Cluster 2 outperformed the other three clusters in terms of the prediction for DNAH5, HRAS, and PTEN mutations. Consistently, the image tiles with the highest likelihood of DNAH5, HRAS, and PTEN were red blood cells, stroma/red blood cells, and tumor/blood cell, respectively (Figure 9).

For CRC (Table 3), features from both Cluster 1 (primarily tumor and mucin tissues) and Cluster 4 (a mixture of all types of tissues) had a better predictive performance for the vast majority of the genes, although the tumor compartment (Cluster 2) also had a decent prediction performance. The image features best predicted PIK3CA, ATM, BRAF, and MET genes with an AUC > 0.7. In addition, TP53 and RET can also be consistently predicted by the digital pathology based deep learning models (AUC > 0.65). These results suggest that the mutant-like image features for these clinically relevant genes in CRC are not exclusively confined in the tumor regions (Figure 10). Other tissues, particularly mucin, has a great predictive value for mutation of CRC genes which is consistent with the work of Nguyen et al., where image
patches for tumor and mucus regions tended to better predict MSI status for CRC patients [48]. In addition, the image tiles demonstrated in Cluster 4 lymphocytes most likely related to ERBB2, ATM, and MET mutations, while an image tile with normal-tissue-like features had the highest likelihood of PIK3CA mutation (Figure 10).

**Comparison of prediction: the best cluster vs. WSI**

We used the average AUC from the five-fold cross-validation to compare our best-cluster-based approach with the model using all patches from a WSI, and we focused on genes with an average AUC greater than 0.6 (Figure 11). The cross validation showed that the selected best cluster consistently provided improvement for prediction of gene mutations for all the three cancer types (i.e., LUAD, HNSCC, and CRC) (Figures 11). Overall, the improvement of the average AUC from the cross validations was up to 0.08 (Figure 11b). In lung adenocarcinoma (LUAD), remarkable improvement of AUC (ΔAUC) was observed for STK11 (0.061), ALK (0.051), and KRAS (0.046). For HNSCC, DNAH5 (0.083), HRAS (0.067), and PTEN (0.057) had the most improved AUC, while the prediction of ERBB2 (0.066) and RET (0.054) genes was remarkably improved for CRC.

**Comparison of the proposed clustering algorithm with baseline algorithms**

Figure 12 shows that our proposed best-cluster method outperformed both baseline algorithms. The superiority over the Zhu et al. model suggests combining and sampling from clusters with AUC > 0.5 tended to produce a suboptimal predictive ability. In addition, the cluster distribution algorithm of Dooley et al. generally does not perform well for gene mutation data,
producing AUC mostly around or lower than 0.5.

These results suggest that unsupervised clustering can facilitate identification of patches with better predictive values and exclude patches lack of predictive information. Furthermore, as expected, introducing less predictive clusters could compromise model predictive ability as our proposed best-cluster approach outperformed Zhu’s method where combined top clusters (AUC > 0.5) were used to construct prediction model.

*Comparison of prediction: the best cluster vs. tumor area*

Figure 13 shows the average AUC of model trained on the best clusters and model trained only on tumor patches. BCOMIL outperformed models trained only on tumor regions. The predictions of RET and BRAF mutations were improved by 0.062 and 0.042, respectively. In addition, Table 4 shows that patches in Cluster 1 (tumor + mucin) and Cluster 4 (a mix of all tissue types) provided more predictive models than the models based on only tumor patches identified by the CRC tissue classifier.

*Discussion*

WSIs are widely used in digital pathology to predict gene mutations, molecular subtypes, and clinical outcomes. Since WSIs are too large (Giga pixels) to fit on a GPU at once, they are usually split into small image patches for training neural networks and prediction models. However, since patch-level labels are usually not available, we cannot directly perform classification on each patch. Therefore, multiple instance learning is often implemented to
develop prediction models at patient level. It is commonly assumed that tumor regions carry most predictive information. Manual annotations by pathologists and classification models for predicting tumor tissues are usually used to select tiles for development of deep-learning models. In this paper, we proposed an unsupervised clustering method to segment WSIs according to the different morphologic features and select tiles for prediction deep-learning models.

We demonstrated that different clusters possessed different predictive ability. In addition, clustering of image patches helped to identify predictive patches and therefore improved prediction on gene mutations in three different cancer types (LUAD, HNSCC, and CRC from TCGA) compared to the approach using all patches from WSIs. These results suggest that unsupervised clustering can facilitate identification of patches with better predictive values and exclude patches lack of predictive information. Furthermore, our proposed algorithm outperformed two recently published baseline algorithms leveraging unsupervised clustering to assist model prediction. Finally, the unsupervised-clustering-based deep learning mutation prediction models allows understanding the spatial locations of each cluster, to identification of the spatial regions related to mutation of a specific gene via the resolved probability scores, and highlighting the heterogeneity of a predicted genotype in the tumor microenvironment.

Image tiles from tumor regions of a WSI are usually selected for constructing deep-learning digital pathology models based on the assumption that tumor cells possess the most predictive information. It appears that this hypothesis is true for HNSCC as tumor tissues (Cluster 4)
tended to provide a better prediction of gene mutation for HNSCC (Table 3). However, for LUAD, tumor-like image tiles seem to be less predictive for genes ALK, KRAS, and KEAP1. Similarly, for CRC, either the tumor tiles (Cluster 2) identified by the unsupervised clustering or the tumor patches identified by a supervised classifier of CRC (the far right column in Table 4) did not provide a superior prediction performance for gene mutation status. Conversely, tumor and mucin tiles (Cluster 1) of CRC as well as a mix of a variety of tissue types (Cluster 4) better predicted CRC mutations compared to tumor tiles (Table 4). This suggests that selection of tumor regions of WSIs is not always the best way to identify patches for prediction of gene mutations, and other tissue types in the tumor micro-environment may provide a better prediction ability for certain phenotypes than tumor tissues. Studies show that mucin-to-tumor area ratio is highly correlated with CMS groups, MSI status, and expression of mucin-producing genes [48].

Finally, we also demonstrated that unsupervised clustering can help reduce the workload for pathologist-based manual annotation. We assumed that a limited number of tissue types are present in WSIs, and repeated clustering of the tiles could separate individual tissue types based on their morphologic appearance. By further clustering of each cluster, we selected a small number of the tiles near the center of subclusters (e.g., 4 tiles) of each cluster for pathologist’s annotation. We showed that this semi-automatic annotation approach could identify similar tissue types for CRC WSIs compared to an automatic tissue classifier for CRC, and improve the interpretability of unsupervised-clustering-based deep learning models.
Acknowledgements

The research of Zihan Chen, Xingyu Li, and Hong Zhang was partially supported by National Natural Science Foundation of China (No. 11771096, 72091212), Anhui Center for Applied Mathematics, and Special Project of Strategic Leading Science and Technology of CAS (No. XDC08010100). This material is based upon work supported by the Google Cloud Research Credits program with the award GCP19980904.

Author contributions

X.S.X., X.L., Z.C., and H.Z. contributed to design of the research; X.L., Z.C., and X.S.X. contributed to data acquisition; X.L., Z.C., and X.S.X. contributed to data analysis. Z.C., X.L., X.S.X., M.Y., and H.Z. contributed to data interpretation. Z.C., X.S.X., X.L., and H.Z. wrote the manuscript; and all authors critically reviewed the manuscript and approved the final version.

Data availability

The TCGA dataset is publicly available at the TCGA portal (https://portal.gdc.cancer.gov). The public TCGA clinical data is available at the website(https://xenabrowser.net/datapages/). Xception model weights are available at (https://github.com/fchollet/deep-learningmodels/releases/download/v0.4/xception_weights_tf_dim_ordering_tf_kernels_notop.h5).

Code availability

Source code is available at https://github.com/ChenZHUSTC/BCOMIL.
Reference

1. Abeshouse, A., et al., *The molecular taxonomy of primary prostate cancer*. Cell, 2015. **163**(4): p. 1011-1025.
2. Bailey, P., et al., *Genomic analyses identify molecular subtypes of pancreatic cancer*. Nature, 2016. **531**(7592): p. 47-52.
3. Dienstmann, R., et al., *Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer*. Nature Reviews Cancer, 2017. **17**(2): p. 79-92.
4. Lindeman, N.I., et al., *Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology*. Journal of Thoracic Oncology, 2013. **8**(7): p. 823-859.
5. Russnes, H.G., et al., *Breast cancer molecular stratification: from intrinsic subtypes to integrative clusters*. The American Journal of Pathology, 2017. **187**(10): p. 2152-2162.
6. Woodman, S.E., et al., *New strategies in melanoma: molecular testing in advanced disease*. Clinical Cancer Research, 2012. **18**(5): p. 1195-1200.
7. Bazan, V., et al., *Specific codon 13 K-ras mutations are predictive of clinical outcome in colorectal cancer patients, whereas codon 12 K-ras mutations are associated with mucinous histotype*. Annals of Oncology, 2002. **13**(9): p. 1438-1446.
8. Castagnola, P. and W. Giaretti, *Mutant KRAS, chromosomal instability and prognosis in colorectal cancer*. Biochimica et Biophysica Acta (BBA)-Reviews on Cancer, 2005. **1756**(2): p. 115-125.
9. Liu, X., M. Jakubowski, and J.L. Hunt, *KRAS gene mutation in colorectal cancer is correlated with increased proliferation and spontaneous apoptosis*. American Journal of Clinical Pathology, 2011. **135**(2): p. 245-252.
10. Nosho, K., et al., *Comprehensive biostatistical analysis of CpG island methylator phenotype in colorectal cancer using a large population-based sample*. PloS One, 2008. **3**(11): p. e3698.
11. Poehlmann, A., et al., *K-ras mutation detection in colorectal cancer using the Pyrosequencing technique*. Pathology-Research and Practice, 2007. **203**(7): p. 489-497.
12. Russo, A., et al., *Prognostic and predictive factors in colorectal cancer: Kirsten Ras in CRC (RASCAL) and TP53CRC collaborative studies*. Annals of Oncology, 2005. **16**(4): iv44-iv49.
13. Suehiro, Y., et al., *Epigenetic-genetic interactions in the APC/WNT, RAS/RAF, and P53 pathways in colorectal carcinoma*. Clinical Cancer Research, 2008. **14**(9): p. 2560-2569.
14. Blumenthal, G.M., et al., *Oncology drug approvals: evaluating endpoints and evidence in an era of breakthrough therapies*. The Oncologist, 2017. **22**(7): p. 762.
15. Pérez-Soler, R., et al., *Determinants of tumor response and survival with erlotinib in patients with non–small-cell lung cancer*. Journal of Clinical Oncology, 2004. **22**(16): p. 3238-3247.
16. Rusch, M., et al., *Clinical cancer genomic profiling by three-platform sequencing of whole genome, whole exome and transcriptome*. Nature Communications, 2018. **9**(1): p. 1-13.
17. Srinidhi, C.L., O. Ciga, and A.L. Martel, *Deep neural network models for computational histopathology: A survey*. Medical Image Analysis, 2021. **67**: p. 101813.
18. Qu, H., et al., *Genetic mutation and biological pathway prediction based on whole slide images in breast carcinoma using deep learning*. NPJ Precision Oncology, 2021. **5**(1): p. 1-11.
19. Ding, K., et al. *Feature-Enhanced Graph Networks for Genetic Mutational Prediction Using Histopathological Images in Colon Cancer*. in International Conference on Medical Image
20. Liao, H., et al., *Deep learning-based classification and mutation prediction from histopathological images of hepatocellular carcinoma*. Clinical and Translational Medicine, 2020. **10**(2).

21. Chen, M., et al., *Classification and mutation prediction based on histopathology H&E images in liver cancer using deep learning*. NPJ Precision Oncology, 2020. **4**(1): p. 1-7.

22. Wang, X., et al., *Prediction of BRCA gene mutation in breast cancer based on deep learning and histopathology images*. Frontiers in Genetics, 2021: p. 1147.

23. Kather, J.N., et al., *Pan-cancer image-based detection of clinically actionable genetic alterations*. Nature Cancer, 2020. **1**(8): p. 789-799.

24. Coudray, N., et al., *Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning*. Nature Medicine, 2018. **24**(10): p. 1559-1567.

25. Gurcan, M.N., et al., *Histopathological image analysis: A review*. IEEE Reviews in Biomedical Engineering, 2009. **2**: p. 147-171.

26. Yuan, Y., et al., *Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling*. Science Translational Medicine, 2012. **4**(157): p. 157ra143-157ra143.

27. Zhu, X., J. Yao, and J. Huang. *Deep convolutional neural network for survival analysis with pathological images*. in *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 2016. IEEE.

28. Zhu, X., et al. *Lung cancer survival prediction from pathological images and genetic data—an integration study*. in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. 2016. IEEE.

29. Cheng, J., et al., *Identification of topological features in renal tumor microenvironment associated with patient survival*. Bioinformatics, 2018. **34**(6): p. 1024-1030.

30. Sirinukunwattana, K., et al., *Image-based consensus molecular subtype (imCMS) classification of colorectal cancer using deep learning*. Gut, 2021. **70**(3): p. 544-554.

31. Levy, J., et al. *Topological feature extraction and visualization of whole slide images using graph neural networks*. in *BIOCOMPUTING 2021: Proceedings of the Pacific Symposium*. 2020. World Scientific.

32. Bilal, M., et al., *Novel deep learning algorithm predicts the status of molecular pathways and key mutations in colorectal cancer from routine histology images*. MedRxiv, 2021.

33. Jang, H.-J., et al., *Prediction of genetic alterations from gastric cancer histopathology images using a fully automated deep learning approach*. World Journal of Gastroenterology, 2021. **27**(44): p. 7687.

34. Abbet, C., et al. *Divide-and-rule: self-supervised learning for survival analysis in colorectal cancer*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2020. Springer.

35. Dooley, A.E., et al. *Prediction of heart transplant rejection using histopathological whole-slide imaging*. in *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*. 2018. IEEE.

36. Zhu, Y., et al. *Improved prediction on heart transplant rejection using convolutional autoencoder and multiple instance learning on whole-slide imaging*. in *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*. 2019. IEEE.

37. Yao, J., et al., *Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks*. Medical Image Analysis, 2020. **65**: p. 101789.
38. Sharma, Y., et al., *Cluster-to-Conquer: A Framework for End-to-End Multi-Instance Learning for Whole Slide Image Classification*. arXiv preprint arXiv:2103.10626, 2021.

39. Muhammad, H., et al., *EPIC-Survival: End-to-end Part Inferred Clustering for Survival Analysis, Featuring Prognostic Stratification Boosting*. arXiv preprint arXiv:2101.11085, 2021.

40. Otsu, N., *A threshold selection method from gray level histograms*. IEEE Transactions on Systems, Man, and Cybernetics, 1979. 9: p. 62-66.

41. Macenko, M., et al. *A method for normalizing histology slides for quantitative analysis*. in 2009 IEEE International Symposium on Biomedical Imaging: from Nano to Macro. 2009. IEEE.

42. Li, X., et al. *Improving Feature Extraction from Histopathological Images Through A Fine-tuning ImageNet Model*. 2022. arXiv:2201.00636.

43. Kather, J.N., et al., *Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study*. PLoS Medicine, 2019. 16(1): p. e1002730.

44. Mosele, F., et al., *Recommendations for the use of next-generation sequencing (NGS) for patients with metastatic cancers: a report from the ESMO Precision Medicine Working Group*. Annals of Oncology, 2020.

45. Network, C.G.A., *Comprehensive genomic characterization of head and neck squamous cell carcinomas*. Nature, 2015. 517(7536): p. 576.

46. Zhu, X., et al. *Wsisa: Making survival prediction from whole slide histopathological images*. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.

47. Li, X., et al., *A Retrospective Analysis using Deep-Learning Models for Prediction of Survival Outcome and Benefit of Adjuvant Chemotherapy in Stage II/III Colorectal Cancer*. arXiv preprint arXiv:2111.03532, 2021.

48. Nguyen, H.-G., et al., *Image-based assessment of extracellular mucin-to-tumor area predicts consensus molecular subtypes (CMS) in colorectal cancer*. Modern Pathology, 2022. 35(2): p. 240-248.
**Tables**

**Table 1. Numbers of mutation and wild type of each gene in three cancers.**

| Gene   | Total | Mutant | Wild Type | Mutation Frequency |
|--------|-------|--------|-----------|-------------------|
| **LUAD** |       |        |           |                   |
| TP53   | 434   | 224    | 210       | 0.52              |
| STK11  | 434   | 63     | 371       | 0.15              |
| KEAP1  | 434   | 77     | 357       | 0.18              |
| EGFR   | 434   | 55     | 379       | 0.13              |
| ALK    | 434   | 24     | 410       | 0.06              |
| KRAS   | 434   | 132    | 302       | 0.30              |
| **HNSCC** |      |        |           |                   |
| TP53   | 431   | 320    | 111       | 0.74              |
| CASP8  | 431   | 47     | 384       | 0.11              |
| NSD1   | 431   | 51     | 380       | 0.12              |
| HRAS   | 431   | 27     | 404       | 0.06              |
| PTEN   | 431   | 10     | 421       | 0.02              |
| DNAH5  | 431   | 58     | 373       | 0.13              |
| **CRC** |       |        |           |                   |
| TP53   | 414   | 260    | 154       | 0.63              |
| PIK3CA | 414   | 158    | 256       | 0.38              |
| ATM    | 414   | 120    | 294       | 0.29              |
| MET    | 414   | 38     | 376       | 0.09              |
| BRAF   | 414   | 91     | 323       | 0.22              |
| RET    | 414   | 27     | 387       | 0.07              |
| ERBB2  | 414   | 30     | 384       | 0.63              |
Table 2. Average AUC (standard deviation) from 5-fold cross validation for different clusters in TCGA LUAD

| Gene | Cluster 1 (N =433) | Cluster 2 (N = 428) | Cluster 3 (N = 434) | Cluster 4 (N = 434) |
|------|--------------------|---------------------|---------------------|---------------------|
| TP53 | 0.655±0.077        | **0.692±0.082**     | 0.609±0.070         | 0.584±0.075         |
| STK11| 0.608±0.095        | **0.647±0.100**     | 0.553±0.100         | 0.563±0.157         |
| EGFR | **0.649±0.126**    | **0.643±0.118**     | 0.584±0.107         | 0.595±0.123         |
| ALK  | 0.549±0.192        | 0.609±0.151         | 0.544±0.123         | **0.655±0.233**     |
| KRAS | 0.517±0.053        | 0.536±0.054         | **0.608±0.068**     | 0.564±0.070         |
| KEAP1| **0.630±0.137**    | 0.594±0.152         | 0.619±0.084         | 0.611±0.081         |
Table 3. Average AUC (standard deviation) from 5-fold cross validation for different clusters in TCGA HNSCC

| gene    | Cluster 1 (N = 431) | Cluster 2 (N = 431) | Cluster 3 (N = 430) | Cluster 4 (N = 430) |
|---------|---------------------|---------------------|---------------------|---------------------|
| TP53    | 0.690±0.073         | 0.611±0.128         | 0.596±0.068         | **0.719±0.061**     |
| DNAH5   | 0.462±0.090         | **0.604±0.064**     | 0.505±0.067         | 0.479±0.088         |
| HRAS    | 0.590±0.152         | **0.665±0.140**     | 0.454±0.103         | **0.658±0.178**     |
| CASP8   | **0.666±0.124**     | 0.564±0.105         | 0.638±0.061         | **0.665±0.072**     |
| PTEN    | 0.540±0.286         | **0.625±0.230**     | 0.577±0.204         | 0.552±0.225         |
| NSD1    | 0.630±0.100         | 0.632±0.121         | **0.657±0.089**     | **0.639±0.070**     |
Table 4. Average AUC (standard deviation) from 5-fold cross validation for different clusters in TCGA CRC

| gene   | Cluster 1 (N = 413) | Cluster 2 (N = 411) | Cluster 3 (N = 414) | Cluster 4 (N = 414) | Tumor patches* |
|--------|---------------------|---------------------|---------------------|---------------------|----------------|
| TP53   | 0.657±0.034         | 0.642±0.059         | 0.575±0.061         | **0.653±0.028**     | 0.665±0.043    |
| PIK3CA | **0.759±0.071**     | 0.706±0.083         | 0.721±0.081         | **0.766±0.048**     | 0.737±0.085    |
| BRAF   | **0.729±0.043**     | 0.666±0.066         | 0.625±0.072         | **0.703±0.078**     | 0.687±0.054    |
| ERBB2  | 0.554±0.145         | 0.551±0.135         | 0.546±0.049         | **0.633±0.167**     | 0.598±0.137    |
| ATM    | **0.738±0.036**     | 0.733±0.056         | 0.720±0.030         | **0.743±0.026**     | 0.734±0.054    |
| MET    | **0.703±0.097**     | 0.695±0.062         | 0.696±0.096         | **0.737±0.112**     | 0.697±0.052    |
| RET    | **0.685±0.094**     | 0.529±0.111         | 0.510±0.123         | **0.677±0.076**     | 0.623±0.048    |

Note: * model trained on tumor patches identified by a tissue classifier for CRC
Figure 1. Framework of unsupervised clustering-based deep-learning modeling for prediction of gene mutations. Images of three cancer tissues were first downloaded from TCGA and CPTAC databases. Each whole-slide H&E image was preprocessed to (1) remove the background areas using a U-net, (2) split into non-overlapping tiles with a size of 224 x 224 pixels, and (3) color normalized. A fine-tuned Xception model-based feature extractor was used to generate patch representations. For each cancer type, we used K-means clustering to group patches into four clusters. Neural networks with the same structure were trained on each cluster data and all-patch data to obtain patch-level classifiers. The results were finally aggregated per-slide to produce the heatmaps and the AUC statistics.

Figure 2. Framework of semi-automatic annotation of clusters. An additional K-means clustering (k =4) was performed, and four neighborhood patches around the center of each subcluster were selected for pathologists’ annotation.

Figure 3. t-SNE visualization of clustering results. For each cancer, 5,000 patches were randomly selected from each of the four clusters and displayed using t-distributed stochastic neighbor embedding (t-SNE) dimensionality reduction representation.

Figures 4. Representative Image tiles for each cluster for LUAD. For each subcluster, a representative patch is displayed. The tissue type for each tile is annotated using a semi-automatic annotation approach.

Figures 5. Representative Image tiles for each cluster for HNSCC. For each subcluster, a representative patch is displayed. The tissue type for each tile is annotated using a semi-automatic annotation approach.

Figures 6. Representative Image tiles for each cluster for LUAD. For each subcluster, a representative patch is displayed. The tissue type for each tile is annotated using a semi-automatic annotation approach (a) and a CRC tissue-type classifier (b).

Figure 7. Sankey diagram of CRC clustering results. Sankey diagram is used to identify tissue types in different clusters by comparing with the tissue types predicted from a CRC tissue classifier.

Figures 8. Visualization of the proposed algorithm for different genes in LUAD. The Deep learning-based unsupervised clustering and mutation predictions are visualized to understand the spatial locations of each cluster, to identify the spatial regions related to mutation of a specific gene via the resolved probability scores, and to highlight the heterogeneity of a predicted genotype in the tumor microenvironment. The heatmap showcases the probability scores of the gene mutations in the identified best cluster. The tile with the highest probability of mutations for each gene is displayed and the corresponding tissue type is provided.

Figures 9. Visualization of the proposed algorithm for different genes in HNSCC. The
Deep learning-based unsupervised clustering and mutation predictions are visualized to understand the spatial locations of each cluster, to identify the spatial regions related to mutation of a specific gene via the resolved probability scores, and to highlight the heterogeneity of a predicted genotype in the tumor microenvironment. The heatmap showcases the probability scores of the gene mutation in the identified best cluster. The tile with the highest probability of mutation for each gene is displayed and the corresponding tissue type is provided.

**Figures 10. Visualization of the proposed algorithm for different genes in CRC.** The Deep learning-based unsupervised clustering and mutation predictions are visualized to understand the spatial locations of each cluster, to identify the spatial regions related to mutation of a specific gene via the resolved probability scores, and to highlight the heterogeneity of a predicted genotype in the tumor microenvironment. The heatmap showcases the probability scores of the gene mutation in the identified best cluster. The tile with the highest probability of mutation for each gene is displayed and the corresponding tissue type is provided.

**Figure 11. Comparison of model performance (average AUC score) of the proposed clustering-based algorithm with models using whole-slide images without patch selection.** Red points = the best cluster results; green points = models using whole-slide images. The bar charts show the difference in average AUC between clustering model and all-patch model.

**Figure 12. Comparison of model performance (average AUC score) of the proposed clustering-based algorithm with two baseline methods (Dooley et al.’s method and Zhu et al.’s method)**

**Figure 13. Comparison of the proposed BCOMIL method with model trained on tumor patches in CRC.** Red points = the best cluster results; green points = model trained on tumor patches. The bar charts showcase the difference in average AUC between clustering model and all-patch model.
Figure 2.
Figure 3.

LUAD

HNSCC

CRC
Figure 4.

| Subcluster | Cluster 1                  | Cluster 2                     | Cluster 3                      | Cluster 4             |
|------------|----------------------------|-------------------------------|--------------------------------|-----------------------|
| Subcluster 1 | Stroma, Atypical cells     | Tumor                         | Stroma, Atypical cells       | Stroma                |
| Subcluster 2 | Pulmonary alveolus         | Tumor                         | Pulmonary alveolus            | Stroma                |
| Subcluster 3 | Suspected tumor cells      | Tumor                         | Lymphocytes/proliferating fibroblasts | Stroma                |
| Subcluster 4 | Tumor                      | Tumor                         | Non-tumor                     | Stroma/Lymphocytes    |
| Subcluster 1 | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|-------------|-----------|-----------|-----------|-----------|
| Tumor       | Non-tumor Blood vessels | Non-tumor | Tumor     |
| Subcluster 2 | Lymphocytes | Non-tumor | Non-tumor | Tumor     |
| Tumor       | Tumor     | Non-tumor | Tumor     |
| Subcluster 4 | Tumor     | Necrotic debris | Muscle, adipose | Tumor |

**Figure 5.**
| Subcluster 1 | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|-------------|-----------|-----------|-----------|-----------|
| Tumor       | Tumor     | Debris    | Lymphocytes |

| Subcluster 2 | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|--------------|-----------|-----------|-----------|-----------|
| Tumor/Mucin  | Tumor     | Non-tumor | Adipose   |

| Subcluster 3 | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|--------------|-----------|-----------|-----------|-----------|
| Mucin        | Tumor     | Stroma    | Tumor     |

| Subcluster 4 | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|--------------|-----------|-----------|-----------|-----------|
| Tumor        | Tumor     | Muscule   | Non-tumor |
| Subcluster       | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|------------------|-----------|-----------|-----------|-----------|
| Subcluster 1     | Tumor     | Tumor     | Debris    | Lymphocytes |
|                  | ![Image](image1) | ![Image](image2) | ![Image](image3) | ![Image](image4) |
| Subcluster 2     | Mucin     | Tumor     | Stroma    | Adipose   |
|                  | ![Image](image5) | ![Image](image6) | ![Image](image7) | ![Image](image8) |
| Subcluster 3     | Mucin     | Tumor     | Stroma    | Mucin     |
|                  | ![Image](image9) | ![Image](image10) | ![Image](image11) | ![Image](image12) |
| Subcluster 4     | Tumor     | Tumor     | Muscle    | Normal    |
|                  | ![Image](image13) | ![Image](image14) | ![Image](image15) | ![Image](image16) |
| Gene   | WSI | Clustering | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Heatmap | Image tile |
|--------|-----|------------|-----------|-----------|-----------|-----------|---------|------------|
| TP53   |     |            |           |           |           |           |         |            |
| STK11  |     |            |           |           |           |           |         |            |
| EGFR   |     |            |           |           |           |           |         |            |
| ALK    |     |            |           |           |           |           |         |            |
| KRAS   |     |            |           |           |           |           |         |            |
| KEAP1  |     |            |           |           |           |           |         |            |
| Gene   | WSI  | Clustering | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Heatmap | Image tile |
|--------|------|------------|-----------|-----------|-----------|-----------|---------|------------|
| TP53   | ![Image](image1.png) | ![Image](image2.png) | ![Image](image3.png) | ![Image](image4.png) | ![Image](image5.png) | ![Image](image6.png) | ![Image](heatmap1.png) | ![Image](tile1.png) |
| DNAH5  | ![Image](image7.png) | ![Image](image8.png) | ![Image](image9.png) | ![Image](image10.png) | ![Image](image11.png) | ![Image](image12.png) | ![Image](heatmap2.png) | ![Image](tile2.png) |
| HRAS   | ![Image](image13.png) | ![Image](image14.png) | ![Image](image15.png) | ![Image](image16.png) | ![Image](image17.png) | ![Image](image18.png) | ![Image](heatmap3.png) | ![Image](tile3.png) |
| CASP8  | ![Image](image19.png) | ![Image](image20.png) | ![Image](image21.png) | ![Image](image22.png) | ![Image](image23.png) | ![Image](image24.png) | ![Image](heatmap4.png) | ![Image](tile4.png) |
| PTEN   | ![Image](image25.png) | ![Image](image26.png) | ![Image](image27.png) | ![Image](image28.png) | ![Image](image29.png) | ![Image](image30.png) | ![Image](heatmap5.png) | ![Image](tile5.png) |
| NSD1   | ![Image](image31.png) | ![Image](image32.png) | ![Image](image33.png) | ![Image](image34.png) | ![Image](image35.png) | ![Image](image36.png) | ![Image](heatmap6.png) | ![Image](tile6.png) |

Figure 9.

Gene WSI Clustering Cluster 1 Cluster 2 Cluster 3 Cluster 4 Heatmap Image tile
TP53 ![Image](image1.png) ![Image](image2.png) ![Image](image3.png) ![Image](image4.png) ![Image](image5.png) ![Image](heatmap1.png) ![Image](tile1.png)
DNAH5 ![Image](image7.png) ![Image](image8.png) ![Image](image9.png) ![Image](image10.png) ![Image](image11.png) ![Image](heatmap2.png) ![Image](tile2.png)
HRAS ![Image](image13.png) ![Image](image14.png) ![Image](image15.png) ![Image](image16.png) ![Image](image17.png) ![Image](heatmap3.png) ![Image](tile3.png)
CASP8 ![Image](image19.png) ![Image](image20.png) ![Image](image21.png) ![Image](image22.png) ![Image](image23.png) ![Image](heatmap4.png) ![Image](tile4.png)
PTEN ![Image](image25.png) ![Image](image26.png) ![Image](image27.png) ![Image](image28.png) ![Image](image29.png) ![Image](heatmap5.png) ![Image](tile5.png)
NSD1 ![Image](image31.png) ![Image](image32.png) ![Image](image33.png) ![Image](image34.png) ![Image](image35.png) ![Image](heatmap6.png) ![Image](tile6.png)
| Gene       | WSI           | Clustering | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Heatmap | Image tile |
|------------|---------------|------------|-----------|-----------|-----------|-----------|---------|------------|
| TP53       | ![Image](image1.png) | ![Heatmap](heatmap1.png) | ![Image](image2.png) | ![Image](image3.png) | ![Image](image4.png) | ![Image](image5.png) | ![Image](heatmap2.png) | ![Image](image6.png) |
| PIK3CA     | ![Image](image7.png) | ![Heatmap](heatmap3.png) | ![Image](image8.png) | ![Image](image9.png) | ![Image](image10.png) | ![Image](image11.png) | ![Image](heatmap4.png) | ![Image](image12.png) |
| BRAF       | ![Image](image13.png) | ![Heatmap](heatmap5.png) | ![Image](image14.png) | ![Image](image15.png) | ![Image](image16.png) | ![Image](image17.png) | ![Image](heatmap6.png) | ![Image](image18.png) |
| ERBB2      | ![Image](image19.png) | ![Heatmap](heatmap7.png) | ![Image](image20.png) | ![Image](image21.png) | ![Image](image22.png) | ![Image](image23.png) | ![Image](heatmap8.png) | ![Image](image24.png) |
| ATM        | ![Image](image25.png) | ![Heatmap](heatmap9.png) | ![Image](image26.png) | ![Image](image27.png) | ![Image](image28.png) | ![Image](image29.png) | ![Image](heatmap10.png) | ![Image](image30.png) |
| MET        | ![Image](image31.png) | ![Heatmap](heatmap11.png) | ![Image](image32.png) | ![Image](image33.png) | ![Image](image34.png) | ![Image](image35.png) | ![Image](heatmap12.png) | ![Image](image36.png) |

Figure 10.
Figure 11.
Figure 12.

|      | LUAD AUC | HNSCC AUC | CRC AUC |
|------|----------|-----------|---------|
| TP53 | ![Bar Chart](#) | ![Bar Chart](#) | ![Bar Chart](#) |
| ALK  | ![Bar Chart](#) | ![Bar Chart](#) | ![Bar Chart](#) |
| EGFR | ![Bar Chart](#) | ![Bar Chart](#) | ![Bar Chart](#) |
| STX11| ![Bar Chart](#) | ![Bar Chart](#) | ![Bar Chart](#) |
| KEAP1| ![Bar Chart](#) | ![Bar Chart](#) | ![Bar Chart](#) |
| KRAS | ![Bar Chart](#) | ![Bar Chart](#) | ![Bar Chart](#) |

Method:
- Proposed best cluster
- Zhu
- Doeley

AUC values range from 0.3 to 0.7 for LUAD and HNSCC, and from 0.3 to 0.8 for CRC.
**Supplementary:**

Supplementary Table 1. A summary of the number of patients of each cancer in different tasks.

| cancer   | LUAD mutation | HNSCC mutation | CRC mutation |
|----------|--------------|----------------|-------------|
| tcga_all | 434          | 431            | 414         |
| tcga_c1  | 433          | 431            | 413         |
| tcga_c2  | 428          | 431            | 411         |
| tcga_c3  | 434          | 430            | 414         |
| tcga_c4  | 434          | 430            | 414         |

Supplementary Figure 1. Target patch of Macenko’s normalization
Supplementary Figure 2. Confusion matrix of CRC tissue-type classification