Spatial-context-aware RNA-sequence prediction from head and neck cancer histopathology images

Shreya Sharma*, Srikanth Ragothaman, Abhishek Vahadane, Devraj Mandal and Shantanu Majumdar

Abstract-Molecular profiling of the tumor in addition to the histological tumor analysis can provide robust information for targeted cancer therapies. Often such data are not available for analysis due to processing delays, cost or inaccessibility. In this paper, we proposed a deep learningbased method to predict RNA-sequence expression (RNA-seq) from Hematoxylin and Eosin whole-slide images (H&E WSI) in head and neck cancer patients. Conventional methods utilize a patch-by-patch prediction and aggregation strategy to predict RNA-seq at a whole-slide level. However, these methods lose spatial-contextual relationships between patches that comprise morphology interactions crucial for predicting RNA-seq. We proposed a novel framework that employs a neural image compressor to preserve the spatial relationships between patches and generate a compressed representation of the whole-slide image, and a customized deep-learning regressor to predict RNA-seq from the compressed representation by learning both global and local features. We tested our proposed method on publicly available TCGA-HNSC dataset comprising 43 test patients for 10 oncogenes. Our experiments showed that the proposed method achieves a 4.12% higher mean correlation and predicts 6 out of 10 genes with better correlation than a state-of-the-art baseline method. Furthermore, we provided interpretability using pathway analysis of the best-predicted genes, and activation maps to highlight the regions in an H&E image that are the most salient of the RNA-seq prediction.

Clinical relevance—The proposed method has the potential to discover genetic biomarkers directly from the histopathology images which could be used to pre-screen the patients before actual genetic testing thereby saving cost and time.

I. INTRODUCTION

In cancer diagnosis and treatment, molecular profiling of patients is in increasing demand to take advantage of targeted or biomarker-based therapies. For example, patients with lung cancer who have EGFR gene mutations or patients with melanoma who have BRAF gene mutations have got approval for targeted therapies by the US Food and Drug Administration. Although molecular profiling can accurately characterize the genotype of a given tumor, these data are not routinely used in analysis due to high cost, relatively long turnaround time (days to weeks), and inaccessibility even in premier care centers. On the other hand, histopathology images such as H&E WSI can be acquired inexpensively and are accessible even under low resource settings [1].

Studies have shown that there exists a relationship between cell morphology in histology data and genetic mutations in molecular data. This relationship can be well exploited by deep learning techniques to predict one modality from another [1]-[5]. The first work in this direction was done by Coudray et al. where they proposed a multi-task Inception network to predict 10 most commonly mutated genes in lung adenocarcinoma and found that mutations of a few genes can be predicted directly from pathology images with accuracy ranging from 74% to 83% [2]. Another work done by Schaumberg et al. proposed a meta-ensemble of ResNet model to predict SPOP gene mutations from prostate cancer images and achieved more than 70% accuracy [3]. Furthermore, researchers have proposed techniques to predict clinically relevant biomarkers such as tumor mutation burden and microsatellite instability directly from H&E images [4][5].

RNA sequencing is one of the types of molecular profiling that provides an efficient high-throughput technique to robustly characterize the tumor-immune microenvironment (TME) of cancer patients [6]. The increasing use of RNAseq expression has enabled the discovery of novel biomarkers that are responsive to cancer immunotherapy. Based on the genotype-phenotype correlations, a recent study proposed a deep learning method to predict RNA-seq directly from H&E WSI [7]. The method uses a patch-based framework where an H&E WSI is first divided into patches, then a deep learning model is trained to predict RNA-seq for each patch and finally the results of all patches are aggregated to get wholeslide level RNA-seq. Although this study initiated the field of RNA-seq prediction from H&E images, the results still show a low correlation between actual and predicted gene expression values in RNA-seq for most cancer types.

We hypothesize certain drawbacks of the current approach. Firstly, there is an inherent assumption that each patch has a sufficient signal to express whole-slide level RNA-seq. The method randomly samples a subset of total patches from H&E WSI that may lead to exclusion of informative regions or inclusion of noisy regions in the image that leads to low correlation. Secondly, by dividing H&E WSI into patches and treating each patch individually, we lose the spatial-contextual relationship between patches that contain cell morphology patterns important for predicting RNA-seq. For example, spatial structures in TME at the boundary of the tumor and non-tumor cells are very sensitive for RNA-seq prediction but these structures are lost in the conventional patch-by-patch-based technique [7].

In this paper, we proposed a spatial-context-aware RNAseq prediction method that preserves spatial relationships between patches and utilizes the entire H&E WSI instead of randomly sampled patches to predict RNA-seq. We employed neural image compression to compress a whole-slide image

^{*}Corresponding author is available at shreya.sharma@rakuten.com. All authors are associated with Rakuten Institute of Technology India, Rakuten Group, Inc.



Fig. 1. Block diagram of our proposed method, (a) input H&E whole-slide image, (b) extracted tissue-regions marked in yellow color, (c) extracted patches from the tissue-regions marked by blue crosses, (d) compressed spatial-contextual features where each small image in the grid is a WSI-level feature representation, (e) deep learning regression model and (f) output RNA-seq prediction comprising of three gene expression values as an example.

into a compressed feature representation that maintained the spatial consistency between the patches and designed a deep learning regressor based on the VGG-Net [8] architecture to infer RNA-seq from the compressed feature representation. With this framework, we avoided the assumption that each patch has RNA-seq information and used both the local cellular features and the global spatial features for inferring RNA-seq at the whole-slide level. We tested our method on a publicly available dataset for head and neck cancer [9] and showed improvements over the conventional method that ignored spatial-contextual information.

II. METHODOLOGY

Fig. 1 shows a block diagram of our proposed method which comprises of three main steps: 1) dataset preprocessing, 2) spatial-contextual feature extraction, and 3) deep-learning regression.

A. Dataset preparation

In this study, we used the publicly available dataset from the Cancer Genome Atlas (TCGA) [9]. We selected primary tumor samples for head and neck squamous cell carcinoma (HNSCC), for which both H&E WSI and RNA-seq data were available. Since tumors are heterogeneous and diverse, a pan-cancer model might not have generalized well to all cancer types and provides fewer insights about cancer. Thus, we focused our study on a single type of cancer, namely, HNSCC, which is not well-examined in literature yet [2]-[5]. HNSCC includes cancers of multiple sites in head and neck region which are morphologically different and challenging. We selected 462 patients from TCGA-HNSC database with matched H&E WSI and RNA-seq data in FPKM-UQ format.

B. Dataset pre-processing

We pre-processed both H&E WSI and RNA-seq data for all 462 HNSCC patients. Due to memory constraints, we first downsampled an input H&E WSI by $4\times$ and then applied Otsu thresholding to separate tissue regions from background regions. Next, we extracted patches of size 128×128 , a common patch size used in histopathology image analysis literature [10], with their corresponding locations from the tissue regions only. For RNA-seq data pre-processing, we excluded genes with a median expression value equal to zero, i.e., those genes that are not expressed in more than 50% of patients. We further shortlisted relevant genes based on GEPIA software [11] and the OncoKB database [12]. The selection was based on differential expression genes between normal and tumor samples with a log-fold change greater than 1 and q-value significance less than 0.01. Only those genes which were annotated as cancer-related and curated by OncoKB were considered for further analysis. This led to the selection of 10 genes for our analysis: PTK6, DKK4, EGR1, HIST1H2BD, HIST1H2BK, HIST1H3H, INHBA, SOCS1, TAP1, OXC11. Since gene expression FPKM-UQ values have varied magnitude scales, we applied $a \rightarrow \log_{10}(1 + a)$ transformation to normalize the magnitude scales.

C. Spatial-contextual feature extraction

After converting H&E WSI into patches, we extracted features from each patch to represent a low-dimensional pixel space into a high-dimensional feature space. To extract features while retaining the spatial-contextual information, we employed a Neural Image Compression (NIC) technique for compressing H&E WSI [13]. NIC uses a neural network to map patches into feature vectors and then places each feature vector into an array that keeps the original spatial arrangement intact such that neighboring feature vectors in the array represent neighboring patches in the original H&E WSI. Several neural network architectures have been proposed in NIC, out of which we selected variational autoencoder (VAE) which has shown good performance in the medical image classification tasks. We used a pre-trained VAE from [13] trained on H&E datasets that are quite similar to our dataset for obtaining compressed features.

D. Deep learning regression

We designed a deep learning regression model based on the VGGNet which is a well-established deep learning architecture. Our model inputs the compressed feature representation of H&E WSI as feature vectors and outputs RNAseq prediction comprising of 10 gene expression values. All the feature vectors are resized to size 224×224 as per



Fig. 2. Comparison between proposed and baseline methods: (a) overall correlation and (b) gene-wise correlation. Bars indicate mean performance on test patients with 95% confidence interval error bars. ns: not significant, $+: p < 10^{-1}$, $*: p < 5 \times 10^{-2}$, $**: p < 10^{-2}$, $***: p < 10^{-3}$, $****: p < 10^{-4}$.

the VGGNet architecture setting with the number of input channels set to 128 which is the length of each feature vector. The proposed architecture contains 8 convolutional layers of kernel size (3,3) with stride = 1 and padding = 1 each followed by a batch-normalization layer and a max-pooling layer with stride = 2 and padding = 2. Also, we added 2 hidden fully-connected layers with 512 units each and an output fully-connected layer with 10 units. The convolutional layers learn local patch-level and global imagelevel features through a hierarchical learning process and the fully-connected layers regress the gene values based on the learned features.

We optimized the loss function of the proposed deep learning model with the mean-squared error (MSE) loss function that computes the mean of the summation of the squared difference between the ground-truth and predicted RNA-seq gene expression values over the H&E WSI used during training. Mathematically, MSE is defined as follows,

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (y_{true}^{i} - y_{pred}^{i})^{2}$$
(1)

where, i denotes the i^{th} image out of total N training images, y^i_{true} denotes the ground-truth and y^i_{pred} denotes the predicted RNA-seq for the i^{th} training image.

III. EXPERIMENTS

We divided the TCGA-HNSC 462 patient data into 375 training patients, 44 validation patients, and 43 test patients. The training set is used to train the deep learning regression model, the validation set is used for hyperparameter tuning and the test set is used for evaluating performance against a baseline. We used PyHist software [14] for H&E WSI pre-processing, PyTorch [15] framework for creating deep learning models, and NVIDIA Tesla V100 GPU for training.

We have selected the HE2RNA model as our baseline method which is the current state-of-the-art in the RNAseq prediction task [7]. The HE2RNA model is a pancancer model trained and tested on 28 different cancer types. To conduct a fair comparison between the baseline and the proposed methods, we re-implemented the HE2RNA



Fig. 3. Top-8 immunological pathways activated by the 6-best predicted genes with statistical significance measured in $-\log(p\text{-}value)$.

model on our TCGA-HNSC dataset using the same training and testing protocol as the proposed method. We used the Pearson correlation coefficient between the ground-truth and predicted RNA-seq gene expression values as an evaluation metric. Furthermore, we used Wilcoxon signed-rank test to compare the distribution of gene values in the ground-truth and predicted RNA-seq with a significance level of $\alpha = 0.05$.

We set model hyperparameters as follows. The learning rate is equal to $3e^{-5}$ with 0.1 factor decay after 5 consecutive epochs if the validation metric saturates, the number of epochs is equal to 200 for training the model and batch-size is equal to 32 samples. To prevent overfitting, we used early stopping during training with patience set to 10 epochs. We used ReLU activation in all the layers except output layer.

IV. RESULTS AND DISCUSSION

A. Correlation with the RNA-seq ground-truth

Fig. 2 shows the comparison between proposed and baseline methods for overall mean correlation and gene-wise mean correlation. The proposed method outperforms the baseline method by 4.12% in overall correlation and predicts 6 out of 10 genes with higher correlation than the baseline method. Since the TCGA-HNSC cohort is diverse with biopsies from multiple sites and different subtypes, our method is promising in predicting the genes relevant for HNSCC. The proposed method demonstrates the ability of a deep learning



Fig. 4. Activation maps generated using GRAD-CAM++ on few samples from our dataset. Yellow color indicates high attention region in H&E WSI.

network to predict RNA-seq in smaller cohorts (order of 100's) which is relevant in a real case scenario. Further training on relevant image datasets will help a researcher to study genes and their associated morphological changes which as we show can be well-exploited by a non-linear deep learning network. For example, PTK6 gene influences morphology in normal tissue by promoting cellular differentiation and apoptosis, and in cancer tissue by sensitizing cells to mitogenic signals and enhancing proliferation.

B. Activation of immunological pathways

We also examined the immunological pathways activated by the best-predicted genes using Reactome software [16]. Fig. 3 shows the top-8 pathways based on their statistical significance measured in $-\log(p\text{-value})$ activated by the 6best predicted genes. The genes for which the expression was most accurately predicted are associated with pathways important for HNSCC. For example, the altered DNA methylation pathway is an important factor associated with HNSCC development [17] and SIRT1 expression is a good indicator of HNSCC prognosis [18]. This shows that our method has the potential to predict genes with high clinical importance.

C. Model interpretability with activation maps

We provided interpretability of the proposed model using activation maps generated by GRAD-CAM++ [19]. The maps highlight the regions in H&E WSI which are the most salient of the RNA-seq prediction. Unlike the baseline method, we retain the spatial locations of all the patches which makes it possible to generate such saliency maps and makes our model interpretable for biologists. Fig. 4 shows activation maps for few samples from our dataset. In all samples, our model gives high attention to tissue regions and ignores the background. We observed that a certain region of the H&E WSI gets high attention. On zooming-in the image (e.g. Fig. 4(d)-(e)), we found visual differences in cell morphology in high (A) and low (B) attention regions that can help in discovering RNA-seq associated visual patterns.

V. CONCLUSION

We proposed a spatial-context-aware RNA-seq prediction method from H&E WSI. Our framework has a neural image compressor that compresses a WSI into features while retaining spatial-contextual relationships between patches. Secondly, the framework has a deep learning regressor to infer RNA-seq from the compressed features. Experiments on the publicly available TCGA-HNSC dataset for the prediction of 10 oncogenes showed that the proposed method outperforms the conventional method. We provided interpretability through activated pathways and activation maps that will serve as a tool for biologists to understand the black-box deep learning model. In future work, we will test the proposed method for more cancer types and enhance biological interpretability.

References

- N. Coudray et al., "Deep learning links histology, molecular signatures and prognosis in cancer," Nature Cancer, vol. 1, pp. 755-757, 2020.
- [2] N. Coudray et al., "Classification and mutation prediction from nonsmall cell lung cancer histopathology images using deep learning," Nature medicine, vol. 24, no. 10, pp. 1559-1567, 2018.
- [3] A. J. Schaumberg et al., "H&E-stained WSI deep learning predicts SPOP mutation state in prostate cancer," BioRxiv, p. 064279, 2018.
- [4] J. N. Kather et al., "Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer," Nature medicine, vol. 25, no. 7, pp. 1054-1056, 2019.
- [5] H. Zhang et al., "Predicting tumor mutational burden from liver cancer pathological images using convolutional neural network," IEEE Int. Conf. on Bioinformatics and Biomedicine (BIBM), pp. 920-925, 2019.
- [6] D. Lau et al., "RNA sequencing of the tumor microenvironment in precision cancer immunotherapy," Trends in cancer, vol. 5, 2019.
- [7] B. Schmauch et al., "A deep learning model to predict RNA-Seq expression of tumours from whole slide images," Nature communications, vol. 11, no. 1, pp. 1-15, 2020.
- [8] K. Simonyan et al., "Very deep convolutional networks for large-scale image recognition," Int. Conf. on Lear. Repr. (ICLR), 2015.
- [9] TCGA Database: https://portal.gdc.cancer.gov/repository
- [10] P. Bándi et al., "Comparison of different methods for tissue segmentation in histopathological whole-slide images," IEEE 14th Int. Symposium on Biomedical Imaging (ISBI), pp. 591–595, 2017.
- [11] GEPIA Software: http://gepia2.cancer-pku.cn/
- [12] OncoKB Database: https://oncokb.org/
- [13] D. Tellez et al., "Neural image compression for gigapixel histopathology image analysis," IEEE trans. on pat. ana. and mach. intell., 2019.
- [14] M. Muñoz-Aguirre et al., "PyHIST: A Histological Image Segmentation Tool," PLoS Comp. Bio., vol. 16, no. 10, p. e1008349, 2020.
- [15] A. Paszke et al., "Pytorch: An imperative style, high-performance deep learning library," Adv. in Neural Info. Pro. Systems (NIPS), 2019.
- [16] Reactome Software: https://reactome.org/
- [17] N. M. Gašperov et al., "DNA Methylome Distinguishes Head and Neck Cancer from Potentially Malignant Oral Lesions and Healthy Oral Mucosa," Int. Jour. of Mole. Sci, vol. 21, pp. 6853, 2020.
- [18] A. Noguchi et al., "SIRT1 expression is associated with good prognosis for head and neck squamous cell carcinoma patients," Oral surgery, oral medicine, oral pathology and oral radiology, vol. 115, 2013.
- [19] A. Chattopadhay et al., "Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks," IEEE Winter Conf. on Appl. of Computer Vision (WACV), 2018.