Deep Multi-Instance Learning to Predict Mismatch Repair Deficiency in Colon Biopsies

Sasa Andjelkovic[1], Sinisa Todorovic[1], Djordje Pavlovic[1], Stacy Littlechild[1], Igor Mihajlovic[1], Ivana Mikic[1], Claire Weston[1], Casey Laris[1], Timothy Moran[1],
Michael Quick[2], Sid Mayer[2], Raymond Jenoski[2], Kate Mansfield[2], Kristina Weatherhead[3], Kathy Murphy[3].
[1]Reveal Biosciences Inc., San Diego, CA, USA. [2]Hologic, Marlborough, MA, USA. [3]Propath, Dallas, Texas, USA. Contact email: imikic@revealbio.com.

REVEAL BIOSCIENCES Data Powered Pathology
HOLOGIC The Science of Sure
ProPath

ABSTRACT

DNA mismatch repair deficiency (MMR) is a form of genetic instability that impacts approximately 15% of all colorectal cancer patients and a hereditary disorder that can increase a patient's risk of cancer. Identification of MMR in patients impacts their treatment strategy. We present a deep learning system for predicting the MMR abnormal or normal status from Hematoxylin and Eosin (H&E) stained slides of colon cancer biopsies, collected at ProPath and scanned on the Hologic Genius platform.

Our deep learning system is an attention-based multi-instance learning (MIL) model aimed at addressing the computational issue that H&E images are too large for processing as a whole. Our MIL, first, randomly samples foreground tiles from the H&E image, and then extracts and fuses their deep features for predicting the MMR status. In the fusion step, higher attention is assigned to tiles estimated as important for the prediction. Our evaluation of the MIL model on 246 MMR abnormal slides and 136 MMR normal slides, with 80% of these H&E images used for training, gives average accuracy 92.1%.

BACKGROUND

DNA mismatch repair deficiency (MMR) impacts approximately 15% of all colorectal cancer patients. Identification of positive or negative MMR status in patients impacts their treatment.

An artificial intelligence system for predicting MMR status from images of colon cancer biopsies stained with H&E could serve as a decision support tool for pathologists.

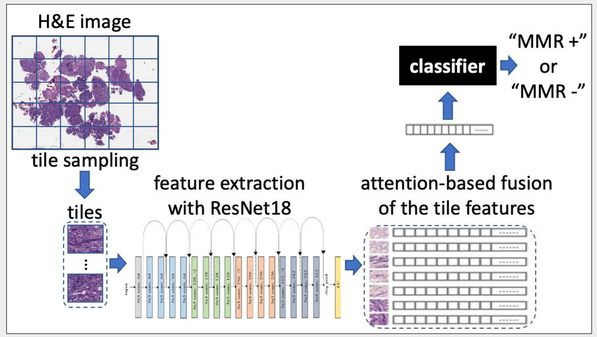
MOTIVATION FOR OUR AI DESIGN

Recent advances in deep learning have led to breakthroughs in medical image analysis. However, deep neural networks:

- Cannot directly process H&E images as a whole,
- Require large training image sets, which are typically not available in medical applications.

KEY ASPECTS OF OUR APPROACH

- Uses ResNet18 for feature extraction from images, since it has a balanced trade-off between performance and the required training dataset size [1].
- Samples a number of smaller tiles from the large H&E image for **Multi-instance learning** (MIL) [2].
- Ensures that the tile sampling selects informative image parts for MMR status prediction by leveraging a public dataset of colon cancer biopsies [3].
- Estimates a relative importance of each sampled tile for prediction, called **attention**.



METHODS

H&E-stained slides of colorectal cancer samples from 382 patients were collected at ProPath and scanned on the Hologic Genius platform.

Attention-based multi-instance learning (MIL) [2] was used to predict positive or negative MMR status as follows:

1. **Tile sampling**: ResNet18 [1] was trained to select tiles with size 128x128 pixels from cancer image regions in the image on the public NCT-CRC-HE-100K dataset of colon cancer biopsies [3] with 9 classes: Adenocarcinoma epithelium, Cancer-associated stroma, Normal colon mucosa, Mucus, Debris, Adipose tissue, Lymphocytes, Smooth muscle, and Background.
2. **Tile attention**: A fully-connected module was used to estimate the importance of each tile for MMR status.
3. **Tile fusion**: A weighted sum of the tile features was passed to a multilayer perceptron to classify the H&E image as MMR positive or negative, where the weights in the sum were set to the tile attentions. In this way, the tile fusion pays more attention to tiles estimated as more important for the MMR prediction.

IMPLEMENTATION DETAILS

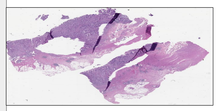
Dataset: 246 MMR abnormal and 136 MMR normal H&E images at resolution 0.255 μm/pixel.

Training: 80% of the dataset; 50 epochs of ADAM training with the cross-entropy loss [1] and L1-norm regularization [1] on data mini-batches consisting of 512 128x128 tiles.

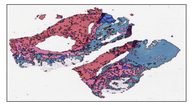
Tile classifier: 9-class average accuracy on 20K test 128x128 tiles from NCT-CRC-HE-100K dataset [3] is 98.5%.

RESULTS

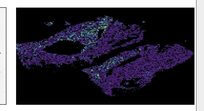
Model	Specificity	Sensitivity	Accuracy
Random tile sampling + fusion by averaging	80.3%	84.1%	82.5%
Tile classifier + fusion by averaging	87.5%	89.7%	89.6%
Tile classifier + fusion by attention Training mini-batch: 128 tiles with size 256x256	86.9%	91.3%	91.2%
Tile classifier + fusion by attention	**88.8%**	**93.8%**	**92.1%**



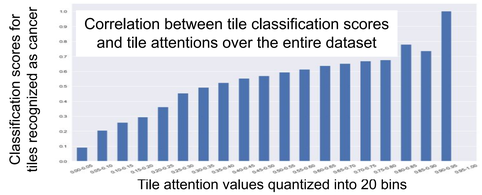
H&E image of a colon cancer biopsy.



Tile classification: Warmer colors indicate our higher scores for cancer regions.



Attentions: Brighter colors indicate higher attentions



The above figure shows strong correlation between the tile classification scores and the tile attentions. This suggests that our MIL model assigns higher importance for MMR recognition to tiles that are estimated as showing cancer.

CONCLUSIONS

Our MIL model estimates relative importance of H&E image tiles for MMR status recognition, and achieves classification accuracy of 92.1% on 382 H&E Hologic slides of colon cancer biopsies. The tiles estimated as important show cancer image regions. Our results strongly suggest that **a reliable prediction of MMR status from H&E images** is possible, when only slide-level annotations are provided in training.

REFERENCES

[1] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition," in IEEE Conf. on Computer Vision and Pattern Recognition, 2016.

[2] M. Ilse, J. Tomczak, M. Welling, "Attention-based Deep Multiple Instance Learning", in Proc. International Conf. on Machine Learning, 80:2127-2136, 2018.

[3] J. Kather, N. Halama, A. Marx, "NCT-CRC-HE-100K" v0.1, 2018. Zenodo.