

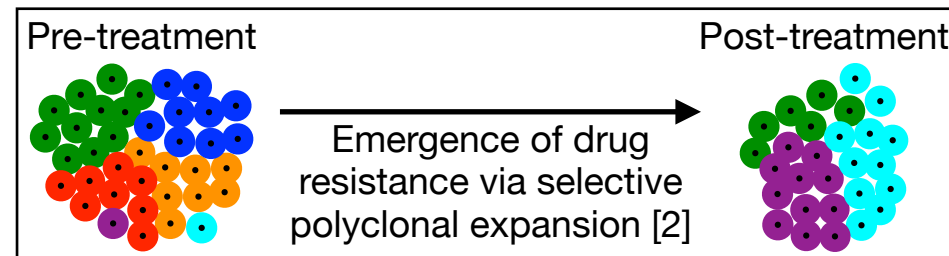
Predicting Genetic Intra-tumor Heterogeneity From Digital Histopathology Slides

Mustafa Umit Oner^{1,2}, Jianbin Chen³, Weiwei Zhai^{7,8}, Wing-Kin Sung^{1,3}, Hwee Kuan Lee^{1,2,4,5,6}

¹School of Computing, National University of Singapore, Singapore. ²A*STAR Bioinformatics Institute, Singapore. ³A*STAR Genome Institute of Singapore, Singapore. ⁴Singapore Eye Research Institute (SERI), Singapore. ⁵Image and Pervasive Access Lab (IPAL), Singapore. ⁶Rehabilitation Research Institute of Singapore, Singapore. ⁷Key Laboratory of Zoological Systematics and Evolution, Chinese Academy of Sciences, China. ⁸Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, China.

Introduction

- ❖ Genetic intra-tumor heterogeneity (ITH) results in therapeutic failure and drug resistance [1]

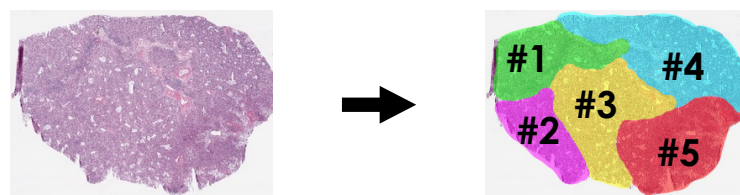


- ❖ Genetic ITH is quantified by genomic sequencing
 - ❖ **Not applicable on small tissue samples**
 - ❖ **Poor scalability: requires fresh/frozen tissue**
 - ❖ **Destructive: spatial information is lost**
 - ❖ **Expensive**

Objective

Developing a machine learning model predicting genetic ITH from hematoxylin and eosin (H&E) stained whole-slide images (WSIs) to provide clinicians with new tools to plan treatments and monitor therapeutic response.

- ❖ **H&E stained WSI is a routine diagnostic tool**
 - ❖ **Widely applicable**
 - ❖ **Highly scalable**
- ❖ **It is cheaper than genomic sequencing**
- ❖ **It can enable us to infer spatial organization of subpopulations (SPs) in the tissue**



Input: H&E stained WSI

Output: # of SPs and their spatial organization

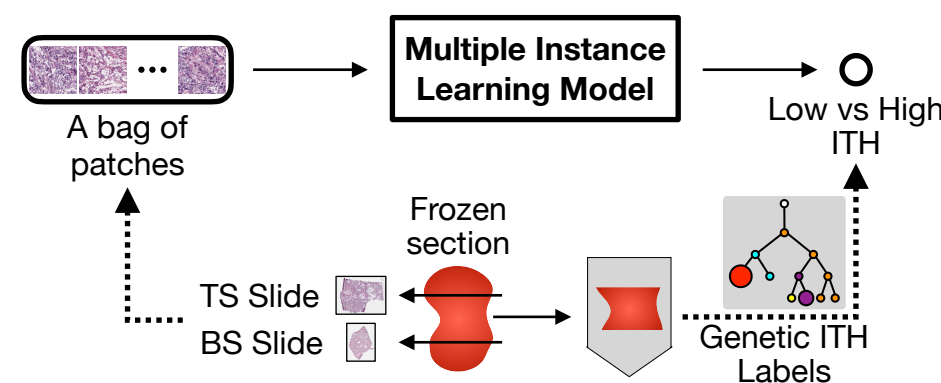
Patient Cohort

- ❖ The Cancer Genome Atlas (TCGA) Lung Adenocarcinoma (LUAD) cohort
- ❖ **Genetic ITH metric:** # of SPs obtained from genomic sequencing data using EXPANDS [3]

| | Low ITH (# SPs < 5) | High ITH (# SPs > 7) |
|----------|------------------------|-------------------------|
| Training | 34 patients | 29 patients |
| Test | 11 patients | 12 patients |

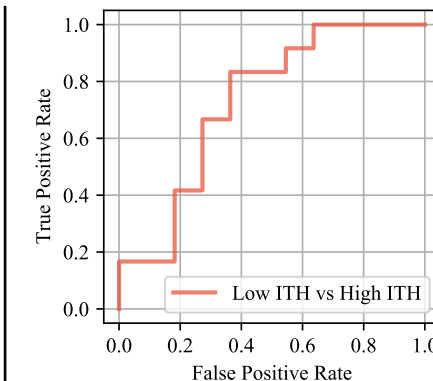
Multiple Instance Learning Model

- ❖ A sample is represented as a bag of patches cropped from the sample's slides and the sample's genetic ITH label, i.e., number of subpopulations, is used as the bag label.
- ❖ The model is trained on 63 patients in the training set and evaluated on 23 patients in the hold-out test set.



Results

- ❖ For a sample, 100 predictions are obtained and mean value is used as the sample's prediction
- ❖ Evaluation metric: area under the receiver operating characteristic curve (AUC) (Fig. 1)
- ❖ 95% confidence interval (CI) is calculated using percentile bootstrap method [4]



AUC = 0.727
CI: 0.485 - 0.938

Fig. 1: The receiver operating characteristic curve

- ❖ Prediction box plots for test set patients (Fig. 2)
 - ❖ The stars show mean values
 - ❖ Different colors represent different ITH groups

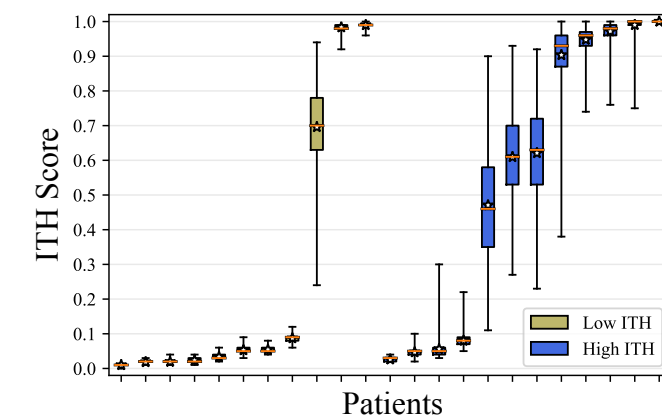


Fig. 2: The prediction box plots for test set patients

Conclusion

- ❖ Our model produces promising results to explore further genetic intra-tumor heterogeneity prediction from H&E stained WSIs as a new tool.

Future Work

- ❖ Semantic segmentation of WSIs using the trained model to infer spatial organization of SPs

References

- [1] McGranahan, N. and Swanton, C., 2015. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer cell*, 27(1), pp.15-26.
- [2] Burrell, R.A. and Swanton, C., 2014. Tumour heterogeneity and the evolution of polyclonal drug resistance. *Molecular oncology*, 8(6), pp.1095-1111.
- [3] Andor, N., Harness, J.V., Mueller, S., Mewes, H.W. and Petritsch, C., 2014. EXPANDS: expanding ploidy and allele frequency on nested subpopulations. *Bioinformatics*, 30(1), pp.50-60.
- [4] Efron, B., 1992. Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics* (pp. 569-593). Springer, New York, NY.