A Deep Learning solution for triaging patient with cancer from their predicted mutational status using histopathological images

Louis-Oscar Morel¹, Valentin Derangère², Laurent Arnould², Sylvain Ladoire², and Nathan Vinçon¹

¹Ummon HealthTech, Dijon, France
²Centre de Lutte contre le Cancer, Dijon, France

Introduction

Targeted cancer therapies are specialized and efficient therapies that have revolutionized the treatment of cancer in the last few years (1, 2). The higher specialization of targeted cancer therapies requires to know more and more information about the patient and getting personalized information requires making more specialized diagnostic tests (3). As an example, the presence or the absence of genomic mutations can be associated with a response to a targeted cancer therapy. Therefore it has become an important information while it helps giving a more specialized and more efficient treatment for each patient. Detection of mutation is routinely made by DNA-sequencing diagnostic tests. Recently, a new strategy of deep learning based method of prediction of mutation from histopathological images has shown promising results. However it is still unknown whether these methods can be useful as screening tests aside from sequencing methods for efficient population diagnosis. Here we developed our own deep learning based prediction pipeline for the detection of mutation in breast, lung, colon and ovarian cancer and find clinically relevant genes that have easy-to-detect mutational status, and especially TP53 which can be detected in these four tumors. We then propose 3 potential screening strategies and show the effectiveness of our deep learning pipeline for optimizing diagnosis in the patient population.

deep learning | cancer | molecular diagnostics | histopathology

Correspondence: NathanVincon@UmmonHealthTech.com

Material and Methods

Study Design. All experiments were conducted in accordance with the Declaration of Helsinki and the International Ethical Guidelines for Biomedical Research Involving Human Subjects. Anonymized scanned WSIs were retrieved from the TCGA project through the Genomic Data Commons Portal (https://portal.gdc.cancer.gov/). We applied our method to the following tumor types: breast (BRCA), lung (LUAD), colon (COAD) and ovary (OV).

Ethics oversight of the TCGA study is described at https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/
Datasets. All data, including histological images and information about the participants from the TCGA database are available at https://portal.gdc.cancer.gov/. Genetic data for patients in the TCGA cohorts are available at https://portal.gdc.cancer.gov/. The corresponding authors of this study are not involved in data sharing decisions of the TCGA database.

Molecular labels. Molecular labels were determined from the masked somatic mutations maf file of somatic mutation using the MuTect2 (14) algorithm corresponding to the dataset. An IMPACT value of “HIGH” or “MODERATE” categorized by VEP software (15) was considered positive while other values were considered negative.

Pipeline. All analyses were performed using Python.

Image preprocessing. Aperio SVS files of diagnostic slides (labeled by DX in their name) from the 4 datasets were first selected. We then extracted the foreground, then tiled the images in non-overlapping patches of 600x600 pixels. These patches are associated with the label 1 if the gene is mutated, 0 if the gene is non-mutated. These patches and labels were used as inputs of the EfficientNetB7 (16) neural network. Output prediction was a float between 0 and 1 for each patch. For the slide-level scoring, we extracted the 99th percentile of the patch prediction score distribution.

Neural network training, model selection. We used an EfficientNetB7 neural network architecture with a single sigmoid output layer after the global average pooling layer. The model was trained for 5 epochs where each epoch included all patches without further refinement.

Implementation and hardware. Experiments were run with a NVIDIA RTX A4000 graphic card and TensorFlow v2.8.0-rc0, keras v2.8.0, CUDA 11.5.

Results

A significant proportion of mutations are detectable with deep learning methods on whole-slide imaging. We set up a deep learning pipeline predicting solid tumor mutations from WSI based on an EfficientNetB7 (16) pre-trained on ImageNet dataset. In this study, 4 datasets from GDC Portal were analyzed, a lung cancer dataset, a colorectal cancer dataset, a breast cancer dataset and an ovarian cancer dataset, respectively TCGA-LUAD, TCGA-COAD, TCGA-BRCA and TCGA-OV. We systematically trained the deep learning pipeline on every mutation having a prevalence greater than 10% in the dataset to focus on mutations that can show statistical significance. These correspond to 922 tested genes. On TCGA-LUAD, TCGA-COAD and TCGA-BRCA, 43% of the gene had a statistically significant predictable mutational status (Figure 1). The TCGA-OV dataset had too few cases to reach statistical significance for most of the genes, therefore we removed it for later analyses. However, we noticed that the TP53 gene had a predictable mutational status with an area under the roc curve (AUC) of 0.65.

Table 1. AUC scores for clinically relevant (top) and non clinically relevant (bottom) genes

| Gene     | Dataset  | AUC |
|----------|----------|-----|
| TP53     | TCGA-COAD| 0.60|
| TP53     | TCGA-LUAD| 0.64|
| TP53     | TCGA-BRCA| 0.74|
| TP53     | TCGA-OV  | 0.65|
| KRS5     | TCGA-COAD| 0.63|
| KRS5     | TCGA-LUAD| 0.61|
| EGFR     | TCGA-COAD| 0.05|
| APC      | TCGA-LUAD| 0.90|
| SCN1A    | TCGA-LUAD| 0.83|
| CDH1     | TCGA-BRCA| 0.82|
| GRIK2    | TCGA-LUAD| 0.82|

Screening strategies can optimize the diagnosis in the patient population. In the next step, we analyze the performance of the most predictable of the previously identified genes in realistic screening scenarios. We define 3 potential strategies and show the associated performance of the algorithm as a discriminative test. First strategy is “Save-all”, where we ask whether the algorithm can filter non-mutated patients while having almost no false-negative in order to avoid unnecessary diagnostic tests. More specifically, we are interested in the number of diagnostic tests
that can be avoided while preserving a sensitivity of 95% or 99% in our screening test (Figure 2). Second strategy is “Fixed-Capacity”, where we ask whether we can optimize the number of mutated patients found in the case of a limited number of sequencing tests. In this configuration, the goal is to optimize the number of patients that will finally benefit from the associated targeted therapy. More specifically, we are interested in the increase of the NGS sensitivity if their capacity to test is 30%, 50% or 70% of the patient population (Figure 2). Third strategy is “Prioritize” where we ask whether we can find a small proportion of patients that are highly likely to be positive in order to prioritize them for the sequencing diagnostic test. The rationale behind the “Prioritize” strategy is that in a context of cancer, time is against the patient and days can significantly change the prognosis (17). More specifically, we are interested in the positive predictive value (PPV) for a cutoff of the 5% and 10% most likely to be mutated for the given gene (Figure 2). We have chosen 5% and 10% cutoffs such that it is high enough to have an influence at the population level and low enough to be realistic for a short-track.

For each screening strategy our deep learning pipeline performance is compared to a random screening test. Intuitively, the random screening test would provide 5% and 1% avoided tests for 95% and 99% sensitivity respectively.

In the “Save-all” strategy (Figure 3), for the non-clinical genes, proportion of avoided tests reached 73% for SCN1A in TCGA-LUAD dataset (Figure 3). For the clinical genes, proportion of avoided tests using screening range from 7.8 to 26.3 for a 95% sensitivity and range from 2.6 to 14.1 for a 99% sensitivity. This means that a quarter of the diagnostic tests can be avoided in the best case, which is with the EGFR gene in TCGA-LUAD. For 95% sensitivity, in all genes but APC in TCGA-COAD, more than 10% tests could be avoided and other genes with highest score were KRAS in TCGA-LUAD with 14.4%, TP53 in TCGA-LUAD with 14.7% and TP53 in TCGA-BRCA with 18.5%.

In the “Fixed-Capacity” strategy (Figure 3), for the non-clinical genes, sensitivity improvement reached +230% for SCN1A in TCGA-LUAD for a capacity of 30% and +101% for GRIK2 in TCGA-LUAD for a capacity of 50%. For the clinical genes, the sensitivity improvement ranges from +19% to +80% for a capacity of 30% of the patient population, from +15% to +48% for a capacity of 50% and from +13% to +26% for a capacity of 70%. Intuitively, the lower is the capacity, the higher is the beneficial effect of the screening strategy. For a 30% capacity, the genes with highest score were TP53 in TCGA-BRCA with +80% sensitivity improvement, then KRAS in TCGA-COAD with +41% and KRAS in TCGA-LUAD with +38% improvement.

In the “Prioritize” strategy (Figure 3), for the non-clinical genes, the increase in PPV reached +1166% for SCN1A in TCGA-LUAD with a 5% cutoff and +633% for the same gene with a 10% cutoff. For the clinical genes, the increase in PPV ranges from +24% to +181% with a 5% cutoff and from +12% to +117% with a 10% cutoff. Best performances were once more obtained by TP53 in TCGA-BRCA dataset, and EGFR in TCGA-LUAD have an increase in PPV of +125% (2.25 times higher risk of being mutated when our test is positive) and +87% for a respective cutoff of 5% and 10%.

Interestingly, having a high AUC score is not necessarily associated with having better performance in the different screening scenarios, the difference can mostly be explained by the prevalence of the mutation in the patient population. Intuitively, the less there are mutated patients, the more the screening can help find more of them with a higher
In conclusion, among the 3 proposed strategies, the “Fixed-Capacity strategy” shows that screening can provide significant improvement in the case where only a limited number of diagnostic tests can be done. The “Prioritize strategy” appears to be efficient for a 5% cutoff but quickly becomes less efficient at a 10% cutoff. Finally, while the “Save-all strategy” can drastically reduce the number of patients to test for non-clinical genes, its effect for clinical genes is less important as it only removes 18.5% of the population for a sensitivity of 95% in the best case. These results would require a medico-economic analysis to clearly see potential advantages of a deep-learning based screening in application.

**Discussion**

Cancer is a very diversified disease where no treatment is ubiquitously sufficiently efficient. Targeted therapies have emerged in response to this diversity, being specialized on
Fig. 3. Performance of the screening pipeline for the 3 screening strategies on multiple clinical and non-clinical genes. (A) and (B) Percentage of avoided tests at a sensitivity threshold of 5% and 1% respectively (save-all strategy). The colors correspond to the dataset of origin using the same code as in figure 1. (C-E) Percentage increase in sensitivity for the fixed-capacity strategy with a capacity of respectively 30%, 50% and 70%. The colors correspond to the dataset of origin using the same code as in figure 1. (F-G) Percentage increase in positive predictive value for the priority strategy with a threshold of 5% and 10%. The y-axis is shown in log scale. The colors correspond to the dataset of origin using the same code as in figure 1.
subtypes of cancer and belonging to a more personalized medicine. However, these new treatments require acquiring more information about the cancer by characterizing its molecular and mutational statuses. Acquiring such information can be very expensive, especially for innovative DNA-sequencing methods which, added to their lack of scalability, makes their democratization difficult. In this paper, we propose credible screening strategies based on new deep learning methods, cheaper, quicker and more scalable. We show that predicting mutational status from WSI is efficient and can optimize the allocation of diagnostic test resources in 3 credible scenarios for clinically relevant genes TP53, KRAS and EGFR.

Clinically relevant genes were defined as genes having an approved or in-development drug associated with their mutational status. TP53 mutation is an indication for the WEE1 inhibitor in a clinical trial of phase 3 in pancreas cancer, for ibrutinib (18) and idealisib (19) in chronic lymphocytic leukemia. The absence of KRAS mutations is an indication for cetuximab (20) in colorectal cancer. Finally, EGFR mutation is an indication for erlotinib (21) or gefitinib (22) in lung cancer.

As for each cancer type only one dataset was used, we cannot provide an estimation of the robustness of mutation detectability to change in the data distribution that would be caused by images coming from a new center. In other words, robustness to inter-center variation is completely unknown for the specific task of mutational status prediction. This is a known barrier for clinical application of deep learning based methods. In this paper, we approximated the population prevalence by the dataset prevalence for the mutation. While prevalence does not affect sensitivity, it has influence over the PPV and the threshold of a given sensitivity. Therefore, results shown for “Fixed-Capacity” strategy would be similar for small prevalence adjustment but results for “Prioritize” and “Save-all” strategy are highly dependent on the true mutation prevalence.

To describe benefits of the shown screening strategy more accurately, a medico-economic analysis could be done integrating the cost and availability of a diagnostic test in a first part, then the clinical benefits of the considered targeted therapy in a second part. Finally, with the ever-changing possibilities in medical cancerology, non-clinically relevant genes but with highly detectable mutational status might become relevant in the future.

Bibliography
1. Kelly E. McCann, Sara A. Hurvitz, and Nicholas McAndrew. Advances in Targeted Therapies for Triple-Negative Breast Cancer. Drugs, 79(11):1217–1230, July 2019. ISSN 1179-1950. doi: 10.1007/s40265-019-01155-4.
2. Sabine Kayser and Mark J. Lewis. Advances in targeted therapy for acute myeloid leukaemia. British Journal of Haematology, 180(4):484–500, 2018. ISSN 1385-2411. doi: 10.1111/bjh.15032. _eprint:_https://onlinelibrary.wiley.com/doi/pdf/10.1111/bjh.15032.
3. Meagan B. Myers. Targeted therapies with companion diagnostics in the management of breast cancer: current perspectives. Pharmacogenomics and Personalized Medicine, 9:7–16, January 2016. ISSN 1778-7066. doi: 10.2147/PGPM.S50555.
4. Karolina N. Dziadkowiec, Emilia Gąsiorowska, Ewa Nowak-Markwitz, and Anna Jankowska.
Michele Milella, Noemi Reguart, Giuseppe Altavilla, Ulpiano Jimenez, Mariano Provenzio, Miguel Angel Moreno, Josefa Terrasa, Jose Muñoz-Langa, Javier Valdivia, Dolores Isla, Manuel Domine, Olivier Molinier, Julien Mazieres, Nathalie Baize, Rosario Garcia-Campelo, Gilles Robinet, Delvys Rodriguez-Abreu, Guillermo Lopez-Vivanco, Vittorio Gabbia, Lieba Ferrera-Delgado, Pierre Bombaron, Reyes Bernabe, Alessandra Beato, Angel Arsal, Enrico Cortesi, Christian Rolfo, Maria Sanchez-Ronco, Ana Drozdowskyj, Cristina Queralt, Itziar de Aguirre, Jose Luis Ramirez, Jose Javier Sanchez, Miguel Angel Molina, Miquel Taron, and Luis Paz-Ares. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. The Lancet Oncology, 13(3):239–246, March 2012. ISSN 1470-2045. doi: 10.1016/S1470-2045(11)70393-X.

Jean-Charles Soria, Yi-Long Wu, Kazuhiko Nakagawa, Sang-We Kim, Jin-Ji Yang, Myung-Ju Ahn, Jie Wang, James Chih-Hsin Yang, You Lu, Shinji Atagi, Santiago Ponce, Dae Ho Lee, Yinpeng Liu, Kiyotaka Yoh, Jian-Ying Zhou, Xiaojin Shi, Alan Webster, Haiyi Jiang, and Tony S K Mok. Gefitinib plus chemotherapy versus placebo plus chemotherapy in EGFR-mutation-positive non-small-cell lung cancer after progression on first-line gefitinib (IMPRESSION): a phase 3 randomised trial. The Lancet Oncology, 16(8):990–998, August 2015. ISSN 1470-2045. doi: 10.1016/S1470-2045(15)00121-7.