Abstract of Dissertation submitted by KRIS LAMI

Overcoming the interobserver variability in lung adenocarcinoma subtyping: A clustering approach to establish a ground truth for downstream applications

肺腺癌亜型における診断者間不一致の克服:クラスター解析による AI 用正解データの構築

Kris Lami, Andrey Bychkov, Keitaro Matsumoto, Richard Attanoos, Sabina Berezowska, Luka Brcic, Alberto Cavazza, John C. English, Alexandre Todorovic Fabro, Kaori Ishida, Yukio Kashima, Brandon T. Larsen, Alberto M. Marchevsky, Takuro Miyazaki, Shimpei Morimoto, Anja C. Roden, Frank Schneider, Mano Soshi, Maxwell L. Smith, Kazuhiro Tabata, Angela M. Takano, Kei Tanaka, Tomonori Tanaka, Tomoshi Tsuchiya, Takeshi Nagayasu, Junya Fukuoka

Archives of Pathology & Laboratory Medicine, in press

Department of Medical and Dental Sciences,
Nagasaki University Graduate School of Biomedical Sciences

（Supervisor：Professor Junya Fukuoka, MD, PhD）

Introduction

Lung cancer is the leading cause of cancer-related deaths globally, accounting for 18% of all cancer deaths. One of its histological and most common subtypes is adenocarcinoma (ADC), representing more than 40% of total lung cancer cases. In 2015, the World Health Organization adopted five main histological patterns of lung ADC, which correspond to its subtypes, namely lepidic, acinar, papillary, micropapillary, and solid, and several other variants, such as invasive mucinous ADC. The accurate identification of different lung adenocarcinoma histologic subtypes is important for determining prognosis of patients, but can be challenging due to overlaps in the diagnostic features, leading to considerable interobserver variability. The objectives of this study are to provide an overview of the diagnostic agreement for lung ADC subtypes among expert pulmonary pathologists and to create a set of ground truth images using a clustering approach for downstream applications.

Materials and Methods

A series of 191 surgically resected lung ADC cases encompassing all histologic subtypes have been retrieved from the Nagasaki University Hospital. Representatives slides for each case have been scanned, to obtain a total of 330 whole slide images (WSI). The resulting WSIs have been divided into three sets with different evaluation levels (small patches, areas with relatively uniform histology, and WSI). The first set included 12 WSIs, which were segmented into small patches of 1 mm^2 to obtain 4,702 patches. The second set included 79 WSIs and areas representing the dominant subtype were annotated. The third set included the remaining 239 WSIs. Seventeen expert pulmonary pathologists and one pathologist in training evaluated the histologic subtypes of the 4,702 small patches of the first set and the 79 annotated areas of the second set. For the third set, 15 pathologists from the same group were asked to provide a case-level diagnosis of 142 patients (representing the 239 WSIs), by determining the dominant and minor subtypes of lung ADC for each patient and estimated their percentages in 5% increments.

Results

Among the 4,702 patches of the first set, 1,742 (37%) had an overall consensus among all pathologists, including 1,520 patches labeled as “no carcinoma cells”. The overall Fleiss’ κ score for the agreement of all subtypes was 0.58. “Solid” pattern was the subtype with the most consensus patches, with all 18 pathologists having complete agreement on 180 patches. Pairwise agreements for invasive versus non-invasive patches ranged from 0.05 to 0.76, while agreements for cancer versus non-cancer patches had a narrower range and higher values. Using the cluster analysis from the answers of the first set, pathologists were hierarchically grouped into 2 clusters, with overall κ scores of 0.588 and 0.563 in Clusters 1 and 2, respectively. Similar results were obtained for the second and third sets, with fair-to-moderate agreements. The highest agreements among the 18 pathologists and for both clusters were seen for categories separating cancer and non-cancer patches as well as solid and other subtypes patches, both of them achieving an almost perfect agreement, followed by invasive mucinous ADC and other subtypes, with a substantial to an almost perfect agreement.

Case-level diagnoses of the third set given by 12 pathologists belonging to a cluster were used to evaluate the lung ADC tumor grades. Survival analysis was conducted to evaluate the ability of pathologists and clusters to separate non-invasive predominant tumors from invasive-predominant tumors. The plotted Kaplan–Meier curves showed statistical significance for both clusters, with *P*-values of .03 and .02 for Clusters 1 and 2, respectively. Cluster 1 showed a better significance than 4 pathologists of its cluster, while Cluster 2 showed a better stratification than 3 pathologists of its cluster, only outperformed by one pathologist.

A subsequent number of patches from the first two sets that obtained lung ADC histologic subtype consensus from pathologists of Clusters 1 and 2 were retrieved. For the first set, a total of 3,733 and 4,214 consensus patches were obtained from Clusters 1 and 2, respectively. For the second set, 6,409 patches and 7,188 patches met consensus for Cluster 1 and 2, after segmentation of the annotated areas. These resulting patches were regarded as ground truth of lung ADC subtypes

Discussion

To date, this is the first study to evaluate interobserver agreement of lung ADC subtypes using different evaluation levels, and attempting to obtain a ground truth using a clustering approach. The moderate to almost perfect agreement seen in the recognition of histologic patterns from patches of the first set were similar to previous studies focused on the interobserver variability of lung ADC subtypes. The second and the third set showed a decrease in the κ score, suggesting that inspection of small areas with minimal morphological features on a low power results in better agreements when evaluating lung ADC subtypes. The clustering approach has been evaluated by its ability to predict the overall survival based on the predominance of the invasive or the non-invasive features of a given tumor. The resulting Kaplan–Meier curves from clusters showed a better stratification of invasive-predominant and non-invasive predominant tumors, Cluster 2 surpassing the majority of pathologists for the particular cohort used in this study. We, therefore, believe that consensus patches derived from these two clusters can help in accurate recognition of lung ADC subtypes and by extension, assessment of invasion, which can be challenging and problematic in some cases. These patches, considered as ground truth, can be therefore used for downstream computational applications, such as training of deep learning algorithms capable of automatically distinguish and separate lung ADC subtypes.

(907 words)