UiT Norges arktiske universitet

Heterogeneity Assessment of PD-L1 Expression in Biopsy Versus Resected Specimens of Lung Adenocarcinoma - Machine Learning Approach

Kiniena Tekie, MK-16, Master's Thesis, UiT

Primary supervisor: Mehrdad Rakaee, PhD, Department of Clinical Medicine, UiT
Secondary supervisor: Elin Richardsen, MD-PhD, Department of Clinical Pathology, UNN
Secondary supervisor: Thomas Berg, PhD, Department of Clinical Pathology, UNN

Acknowledgement

The work presented in this thesis was carried out at the Translational Cancer Research Group, Department of Clinical Medicine, Faculty of Health Science, UiT the Arctic University of Norway under the supervision of Dr. Mehrdad Rakaee, Prof. Elin Richardsen, and Dr. Thomas Berg. I would like to thank Prof. Lill-Tove Rasmussen Busund the leader of Translational Cancer Research Group for financial support of this work. I also would like to thank her for her zeal and success in creating research favorable environment, which I believe will have long lasting influence on research productivity of this group.

I would like to thank Elin Richardsen for introducing me to the subject and to the research group, for her dedicated support and guidance, and for caring so much about my work and my well being. Thomas Berg for the conversations we have had on genomics, although they were few and in between, they are my most favorite memories from the faculty. Ultimately, I thank Mehrdad for teaching me everything I know about machine learning and statistics, I could not have written such a wonderful thesis without his help.

I also would like to thank the Qupath Group for creating the bioimage analysis software and for making it user-friendly and open-source software. The digital scoring was simplified by the easily accessible user manuals and extensive forum discussions.

My thanks go to numerous individuals who over many years have shaped my views about scientific inquiry through their writings, their inquisitive conversations, and their friendship. I would not have seen this far if I did not stand on the shoulders of giants.

Tromsø, May 2022
Kiniena Tekie

Table of Contents

Summary ..... **iv**

Abbreviations v

Keywords v

Introduction ***1***

Lung Cancer **1**

Epidemiology 1

Risk factors 1

Diagnostics 2

Treatment 4

Immunotherapy 5

PD-L1 as a predictive biomarker for immunotherapy **6**

PD-L1 positive cancer cells evade the immune system 6

The PD-L1 score is the primary biomarker for immunotherapy 6

High PD-L1 score is not the only criteria for immunotherapy 7

Heterogenicity between biopsy and resection 7

Technical challenges with PD-L1 IHC biomarker 8

Machine learning models **8**

The aim of this thesis **10**

Methods ***11***

Material **11**

Patient cohort 11

PD-L1 immunohistochemistry staining and scoring 11

Image analysis **12**

Digital histological slides 12

Automated digital scoring 14

Definition of PD-L1+ cell in digital model 14

Statistical analysis **15**

Results ***16***

Clinical attributes/Data curation 16

Manual vs Digital Scoring 17

PD-L1 Expression: Resected vs Biopsy specimens 20

Discussion ***23***

The predictive role of the PD-L1 biomarker 23

Factors that contribute to heterogenous PD-L1 detection 23

Digital vs manual PD-L1 TPS 24

Morphological heterogeneity 25

Biopsy vs Resected specimens 25

Heterogeneity of PD-L1 expression in a tumor core 26

Limitations of the study 27

Conclusion 27

Bibliography 28

Summary

Introduction

PD-L1 is a biomarker that is used to predict the response to immunotherapy of patients with lung cancer. A pathologist manually assesses the PD-L1 expression mainly on biopsy specimens. Even though PD-L1 is an established biomarkers in the routine practice, still some challenges remain. This thesis aims to address two issues: 1) whether an automated scoring system can objectively and reliably reproduce the manual scores, and 2) how the PD-L1 score on biopsy is comparable with surgically resected samples.

Methods

Paired biopsy and resection tumor tissues from 26 patients with adenocarcinoma of the non-small cell lung cancer (NSCLC) were included in this thesis. Immunohistochemistry was used to detect the expression of PD-L1. The slides were digitalized, and the clinical data was retrieved from the patient's journals. A supervised machine learning model (ML) was developed to assess the tumor proportion score (TPS) of PD-L1 expression in the whole slide images. Sensitivity and specificity of ML-derived PD-L1 scores and the intra-class correlation coefficient (ICC) of biopsy vs resected PD-L1 scores were computed.

Results

There was a moderate correlation ($r=0.59$, $P<0.001$) between digital and manual scores using PD-L1 TPS as continuous variable. The ML model showed high performance in PD-L1 scoring with a sensitivity/specificity of 0.88/0.92 and 0.85/0.96 at both <1 vs ≥ 1 % and <50 vs ≥ 50 % TPS cutoffs, respectively. No correlation was observed between biopsy vs resected PD-L1 scores at 1% cutoff using either digital or manual scores. However, at 50% cutoff, both digital and manual scores show high level of consistency (manual TPS ICC, 0.82, $P<0.001$; digital TPS ICC: 0.7, $P=0.01$) across paired biopsy and resection tissues.

Conclusion

The biopsies were found to be equivalent to corresponding resected specimens for determining PD-L1 expression at 50% cutoff in adenocarcinoma of NSCLC. Our machine learning algorithm is found to be robust in detecting PD-L1 positive tumor cells with an accuracy like pathologists. Validation of the findings and algorithms is warranted in large-scale cohort.

Abbreviations

AC	Adenocarcinoma of the lung
DAB	Diaminobenzidine, used in IHC as a brown chromogen
DAB OD	DAB optical density after color deconvolution
ECOG PS	Easter Cooperative Oncology Group Performance Scale
FDA	US Food and Drug Administration
FN	False negative
FP	False positive
FPR	False positive rate
IARC	International Agency for Research on Cancer
ICI	Immune checkpoint inhibitors
IHC	Immunohistochemistry
IHC	Immunohistochemistry
KRAS	Kirsten rat sarcoma viral oncogene homologue
LC	Lung cancer
NGS	Next Generation Sequencing
NSCLC	Non-small cell lung cancer
PD-L1	Programmed death ligand 1
PD-L1–	PD-L1 negative. A tissue where below <1% of the cells express PD-L1.
PD-L1+	PD-L1 positive. A tissue where above 1% of the cells express PD-L1
PD1	Programmed death receptor 1
ROC	Receiver operating characteristic curve
SCLC	Small cell lung cancer
TPR	True positive rate
TPS	Tumor percentage score

Keywords

Programmed cell death ligand-1 (PD-L1), adenocarcinoma of the lung (AC-NSCLC), biopsy specimens, resected specimens, manual scoring, digital scoring, and machine learning

Introduction

Lung Cancer

Epidemiology

The mortality rate of lung cancer closely fluctuated with smoking habits of the population, it drastically increased since 1930s and have been in gradual decline the last 20 years[1-4]. Approximately 2 million deaths/year globally are attributed to lung cancer[4,5]. Lung cancer has maintained its over 40-year reign as the leading cause of cancer-related death[6,7]. 1/3 of all cancer related deaths are due to lung cancer, which accounts for more deaths each year than breast, prostate, and colon cancer combined[8,9]. Comparing lung cancer related deaths with other causes of deaths for all countries in the world might provide a contextual picture. Approximately 56 million people died in 2017, approximately 18 million from cardiovascular diseases, 10 million from cancers, 27 000 died due to terrorism, and 2 million people died due to lung cancer. In other words 20% of cancer related deaths and 3,5% of all deaths are attributable to lung cancer[10].

Risk factors

The lungs are uniquely exposed to inhaled toxic insult, especially cigarette smoking. Tobacco, be it smokeless, secondhand smoking, or direct smoking is a group 1 carcinogen (IARC list of carcinogens)[11]. A toxin is group 1 carcinogen when there is enough evidence to conclude it can cause cancer in humans. In the mid-20th century, while tobacco companies were deliberately claiming health benefits of smoking[12], it is documented by doctors at that time, the individuals apparently in their own minds associated their lung cancers with smoking, and they hoped by stopping the habit to improve their prognosis[13]. A significant number of ex-smokers are entering retirement and the age of maximum vulnerability to lung cancer.

Lung cancer carcinogenesis proceeds over 10-20 years[14], the time it might take for a series of 10-20 genetic mutations to accumulate and the additional necessary events that induce phenotypical change in the tissue to occur[15]. The carcinogenesis window, if correctly identified, can be used for prevention by interrupting the process of carcinogenesis or preventing or delaying tumor occurrence[15]. On the other hand, experience from smoking cessation campaigns simply

demonstrates the difficulty of motivating asymptomatic individuals to commit to preventative medicine. Better molecular selection of patients and improved precision of targeted therapy may make apparently healthy individuals with preneoplastic disease eligible for preventative therapy[14]. Lung cancer, though proved difficult to cure, may be simpler to prevent.

In addition to smoking, silica dust (crystalline, in the form of quartz or cristobalite) and asbestos (all forms) are classified group 1 carcinogenic, and are associated with cancers of mesothelium or lung in humans[11]. Inhaled silica or asbestos, accumulated over 15-20 years, especially ascribed to work place conditions, usually leads to chronic, gradually progressive, and incurable pulmonary interstitial fibrosis[16] and at times lung cancer. Asbestosis and silicosis are not eradicated despite knowledge of the causes and effective means for prevention.

Diagnostics

Lung cancer is a heterogenous disease and the sequence of treatment options requires better targeting diagnostics. Standard classification of histological types of lung cancer has been under constant revision, ever since the heterogeneity of lung cancer became evident around 1960s[13,17]. The precision of the diagnostic tool determines made grouping arbitrary and comparing different studies difficult[13]. Small cell lung cancer (SCLC) distinguished itself early from the rest, the non-small cell lung cancers (NSCLC), with its frequency among heavy smoker, its presenting symptoms attributable to metastatic disease than to primary lesion, and its aggressive early metastases while the primary tumor is undetectably small, and its "oat-cell" histology[13].

NSCLC are pathologically and clinically distinct from the neuroendocrine cells' small cell lung cancer[18,2,19]. Rapid growing lung cancer, characterized by early metastases, typically central in location, with extensive mediastinal adenopathy, bronchoscopy is typically positive, with submucosal spread in many cases, very strong association with smoking[2]. SCLC have cancer cells that are usually smaller and have scanty cytoplasm and often show solid proliferation[20] than NSCLC. SCLC responds better to cytotoxic chemotherapy (cisplatin/etoposide) and radiation therapy[7] than NSCLC. About 80-85% of lung cancer cases are NSCLC[6,21], which is far more often

candidate for surgery than SCLC[19] (*Figure 1*). Only 30% of NSCLC are resectable at presentation[19], and 65% of these patients have advanced stage (IIIB/IV) disease at diagnosis[6,9].

Adenocarcinomas (AC) constitute of 40% of NSCLC and is commonly appeared in peripheral in terms of location, with malignancy detectable in pleural effusion, and is often already metastasized at time of diagnosis[2,18]. Reports of the peripheral location of adenocarcinoma tumors is well documented, for example a study documents out 308 patients with lung adenocarcinoma 267 patients (87%) had peripherally located tumors[22]. Patients with adenocarcinoma are often non-smokers or have often a smoking history of cigarettes with moderate to low tar content.

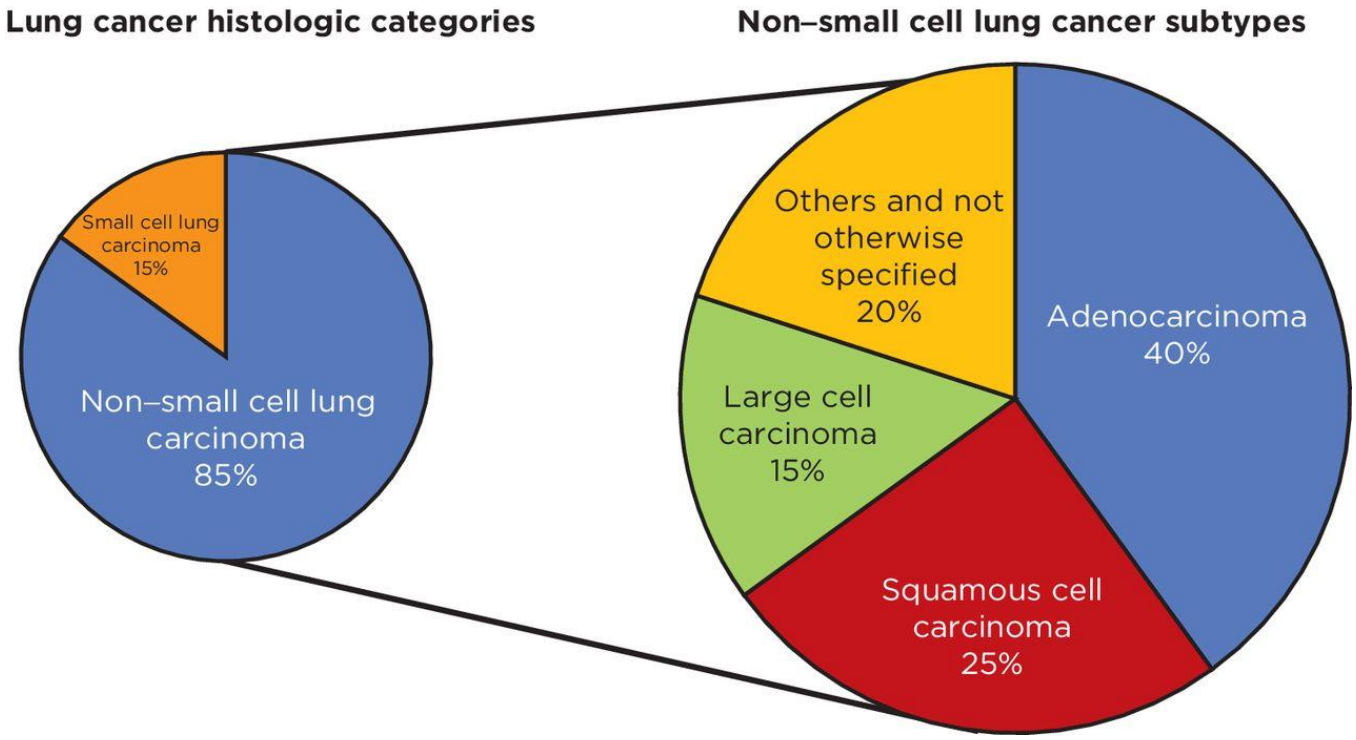


Figure 1 ***Histological classification of lung cancer****. Lung cancer is divided into two major histologic categories (SCLC and NSCLC). Adenocarcinoma is the major histological type of NSCLCs. (Image is reused with permision[23]).*

The emerging molecular targeted therapy of lung cancer require testing for increasing number of cancer cell mutations. The method used for molecular diagnostics in this study is Next Generation Sequencing (NGS), as it is ideal for testing multiple target genes in parallel. It is cost effective and requires only small amounts of tumor tissue[24]. NGS has comprehension, i.e., massive parallel sequencing can be done[25] and it can reveal many mutations of a given gene. For instance, all

relevant EGFR mutation subtypes in the tumor tissue are detectable in one run. NGS also has high sensitivity, i.e. the potential for detecting copy number alteration in tumor cells within a huge excess of normal cells[3]. Alternative methods include Sanger sequencing and mutation specific PCR kits. Sanger sequencing, while being capable of detecting all mutations within a given genomic region, has a greater risk of obtaining false-negative results when the analyzed sample is not sufficiently enriched by tumor cells[3], making it less sensitive than targeted NGS. Mutation specific PCR kits are usually capable of detecting the mutation even if the proportion of tumor cells in the specimen is as low as 1-5%[3], but are limited to detecting a single, or a few mutations per analysis.

Treatment

Cancer services have developed in an ad-hoc fashion. Surgical cancer care was initially undertaken by general surgeons before subspecializing took over in thoracic surgery[26]. The evidence for benefits from specialization, i.e. specialized training, higher caseload, and formation of multidisciplinary teams, came from the success of specialized breast/ovarian cancer care[26]. Outcomes of specialized cancer care varied widely from cancer to cancer when assessing for survival, quality of life, and patient satisfaction. Lung cancer had only benefit in immediate surgical outcomes from specialized practice[26]. The median survival for patients with advanced stage NSCLC treated with platinum-based chemotherapy is a disappointing 8-10 months (2006)[6,9]. Most patients will experience disease progression within 3-6 months diagnosis independent of therapy[9]. The 5-year survival is below 10%[21]. Cure rates are dismal with currently available treatment[27].

Surgical resection in early-stage localized diseases offers substantial cure rates[27], yet have high risk for recurrence of lung cancer[28]. Small tumors (≤2 cm) resected with ≥ 2 cm resection margin have the best chance for non-recurrence[29]. However, the 5-year survival rate after surgery, even in the best stages (I and II) does not exceed 50%[19].

Stage III lung cancer requires a multimodal approach, which may include surgery after neoadjuvant chemo-radio-therapy[19]. This is because in surgically respectable stage III NSCLC, surgery alone was worse than surgery preceded by neoadjuvant chemotherapy[30].

25 years ago, the benefits of chemotherapy was so overshadowed by its toxic effects, that best supportive care was considered an acceptable option for most lung cancer patients with advanced or metastatic NSCLC[31]. With this standard of care, advanced NSCLC had median survival of 4-5 months, the 1-year survival rate was 10%[8]. After two decades of clinical trials with the refinement of chemotherapy regiments and management of adverse events, best supportive care is no longer an acceptable alternative to chemotherapy of advanced NSCLC[21,31,32]. Chemotherapy, especially platinum containing regiments, have been established as cornerstone for treatment of advanced NSCLC[33]. Two decades of refinement of chemotherapy, have increased response rates and improved tolerability[33], but only attained modest survival benefit[19,27]. Even though chemotherapy alone has reached a therapeutic plateau, the majority of lung cancer still inevitably remain candidates for it[33].

Activating EGFR mutations of NSCLC are druggable mutations that are treated with targeted therapy, TKIs[7]. Even when they overexpress PD-L1, they do not respond well to immune-checkpoint inhibitors (ICIs)[34]. Erlotinib, an EGFR tyrosine kinase inhibitor (TKI), have cytotoxic effect that results from G1 cell cycle arrest in cell lines with wild type EGFR mutation[35] and leads to reduced cell growth or to cell death. Erlotinib and gefitinib are both small molecule reversible TKIs[27]. In patients with previously treated NSCLC, those treated with erlotinib have a median survival 2 months longer (6.7 vs 4.7 months) than those given placebo, constituting the rationale the approval by the United States food and drug administration (FDA) of erlotinib for relapsed NSCLC[36].

Immunotherapy

The fundamental concepts of immunotherapy are based on the premise that the development of most tumors is governed by the tumors ability to make immune escape alterations to their microenvironment[37]. An immunogenic tumor, the tumor that is detectable by the immune surveillance, expresses cancer antigens that can induce immune reaction[38]. The accumulated genetic changes, the tumor mutation burden (TMB), of cancer cells, also lead to expression of cancer antigens, i.e. increase the neoantigen burden[39]. In fact, TMB estimation has been investigated as an alternative biomarker to PD-L1[40] expression. The survival of these by default

immunogenic tumors depends on their ability to develop immune resistance mechanisms, such as PD-L1 expression, that uses the immune-checkpoint pathways to induce and maintain tumor-tolerance[39]. Early reports of PD-L1 expressing tumors promoted T-cell apoptosis and proliferation of highly immunogenic tumors[41], supported by even earlier findings of PD1 deficiency leads to development of systemic autoimmune diseases[42], led to the research and initial success of T-cell based cancer immunotherapy[40]. Only one in five unselected patients responded to blockade of the PD-1-axis immunotherapy and PD-L1 expression was suggested as the selection biomarker[43].

PD-L1 as a predictive biomarker for immunotherapy

PD-L1 positive cancer cells evade the immune system

Cancer immunoediting, the balance between immune surveillance and cancer progression, have three possible outcomes: cancer elimination, equilibrium/dormancy, and immune escape[44]. The tumor that develops mechanisms to escape the immune system, like PD-L1+ tumor, may develop into a clinically apparent cancer, and if left untreated, would have worse prognosis than a PD-L1– tumor[44]. Immune escape, the ability of cancer cells to evade the immune system, in addition to expression of PD-L1, may be due to lack of B7 T-cell co-receptor, increased downregulating cytokines, production of immunosuppressive molecules[45]. There are multiple pathways that regulate the immune system: The classical immunosuppressives, like TGF-beta, prostaglandin E2, IL-10, and COX-2, affect antigen presenting cell processing and the acquisition and expression of CTL function[45]; and the PD-L1[46] inhibits the immune cell.

The PD-L1 score is the primary biomarker for immunotherapy

Immune checkpoint inhibitors (ICI), initially in 2nd-line, recently 1st-line treatment of PD-L1+ lung cancer, with strict selection criteria, still include the minority of patients[33]. Benefit from anti-PD-(L)1 therapy is related to its primary biomarker PD-L1 detected by IHC-staining of biopsied specimens. The rule of thumb for initiating immunotherapy: the stronger the PD-L1 positivity (≥50% PD-L1 staining tumor cells), the greater the benefit is. Example of an immunotherapy regiment could be anti-PD-L1 infusion every 3. week for 2 years[47,48].

High PD-L1 score is not the only criteria for immunotherapy

NSCLC patients with EGFR mutations do not respond well to immunotherapy, while patients with KRAS mutations respond well[49]. Some researchers even argue that anti-PD-L1 "poisons the pond" for EGFR-TKI, thus initiating anti-PD-L1 treatment for a patient who later proves to be EGFR-mutation positive, would somewhat reduce potential benefit from TKIs[24]. The patient selection criteria for immunotherapy thus rapidly becomes multifactorial: Patient with NSCLC with high PD-L1 score, high tumor infiltrating lymphocytes and EGFR-mutation negative higher probability of benefit from immunotherapy. The prognostic value of PD-L1 expression in NSCLC patients is disputed. Only 3 out of 11 FDA ICI approvals of NSCLC require PD-L1 testing before treatment with immunotherapy, which are one approval in 2015 and two approvals in 2016[50], the remaining FDA approvals, most of which came after 2016 do not require PD-L1 testing, which is a reflection on that PD-L1 has not been a broadly applied biomarker. It remains nonetheless the only immune-based biomarker for selecting patients with NSCLC for immunotherapy in current clinical practice[51].

Heterogenicity between biopsy and resection

Although surgical specimens are generally considered of suitable size for biomarker analysis, most patients with NSCLC present with unresectable or advanced disease and are often investigated with noninvasive procedures, including bronchial biopsy, endobronchial ultrasound-guided transbronchial needle biopsy, and computed tomography (CT)-guided needle biopsy. These procedures generally provide only small biopsy and cytology size tissue specimens. Researcher in the field continue to investigate the reliability of small biopsy specimens to determine PDL1 expression on tumor core of NSCLC.

Since Ilie et al. first proposed that there is difference in PD-L1 expression detection rate in tumor cells from biopsy and surgical resection specimens[52,53], the potential problem for using biopsy in determining PDL1 was confirmed later by another team of researchers[54]. However, it was later refuted by two other team of researchers[55,56], who showed that biopsy specimens are equivalent to surgically resected specimens for the detection of PD-L1 expression in tumor cells of NSCLC. Therefor the potential role of inaccurate representation of the biopsy specimens to represent PD-

L1 expression in tumor core continues to be investigated, as it may help explain the reason for why studies correlating PD-1/ PD-L1 expression and prognosis in NSCLC have yielded variable results[57].

Technical challenges with PD-L1 IHC biomarker

Uruga et al.[40] systematically addressed the pre-analytical, analytical, and post-analytical issues associated with the PD-L1 biomarker. The pre-analytical issues like specimen-age and sampling to fixation time, the analytical issues involve the selection of a reliable and reproducible immuno-assay, and the post-analytical issues including assessment method and cut-off choice. It is suggested that these issues might be the reason of high response rate (10%) of lung cancer patients with negative PD-L1 expression to anti-PD-1/PD-L1 inhibitors[58].

Machine learning models

Computer-aided decision systems are being developed for application in many areas of diagnosis and treatment of cancer[59-63]. Combining well established image analysis algorithms, such as quantitative histomorphometry[64] that measure shape, size and texture of tumor vs normal tissue images, with artificial intelligence that learns and recognizes patterns, have had very promising initial results[65]. Although a trained human has impressive visual capabilities of pattern recognition, it has less than optimal reproducibility for the standards of precision medicine[66,67]. For detecting and quantifying the increasing amount of medical data, the machine that can learn patterns is uniquely suited. The machine learning algorithms do this with ease in efficiently reproducing its results with precision [68].

Generalization of a successful machine learning algorithms requires verification for use in different scenarios. Investigations are required to make sure that it works well for various staining conditions and tissue types. One must make sure the assumptions employed in training the algorithm, especially the *apriori* knowledge input required by some algorithms is representative for the population in general. Many of the machine learning algorithms are still trained by manually labelling the positive and negative detection examples, this supervision phase, required due to the lack of ground truth training objects, also contributes to the limitations of free generalization of the machine learned algorithm. A continuous human computer interaction for fine-tuning is

therefore necessary before a fully automated digital classifier can function in clinical practice, i.e., the algorithm must first function as an aid for the pathologist, while continuously learning and fine-tuning its results[69].

Full digitalization of the microscopic evaluation of stained tissue slides has made possible for computer aided diagnostics in pathology. Many machine learning algorithms are under continuous development for object recognition problems such as detection of metastasis[67] and quantification of lymphocytes[70]. The research on computer-assisted solutions was initially driven by the need to automate the time-consuming and error-prone manual scoring by pathologist[71], so that the pathologist may spend more time on higher-level decision-making tasks. The digital scorer has also shown promising results at higher-level tasks such as predicting diagnosis and prognosis, for example using the degree of malignancy, architecture and arrangement of tumor infiltrating lymphocytes to predict the risk of recurrence in early-stage NSCLC[64,72]. The computer aided diagnostics can effectively extract appropriate image features in tumor environment (detection and segmentation), classify objects based on the learned patterns (machine learning) and predict outcome.

In addition to helping the pathologist, the machine learning approaches in digital pathology have been helping to address issues faced by oncologists, for example, through the development of prognostic assays to evaluate disease severity and outcome[64,72] as well as features of cancer nuclei[73] and spatial arrangement and clustering of tumor infiltrating lymphocytes[74] to predict response to immunotherapy in NSCLC. In our context, the subjective scoring of PD-L1 IHC is reported as a contributing factor to the inconsistent research results on PD-L1's predictive role of clinical outcome. Improving the accuracy of digital scoring is a hot topic in digital pathology, as it is anticipated to increase the consistency of the results by applying the well documented objective and reproducible method of image analysis[75].

The aim of this thesis

One of the important goals in oncology is to define with greater precision adenocarcinoma patients that will benefit from immunotherapy[76]. This thesis focuses on whether PD-L1 expression on adenocarcinoma of the lung is spatially heterogenous in paired biopsy and surgically resection samples considering following objectives: 1) developing a machine learning algorithm to quantify the PD-L1 TPS. 2) determining whether digital scoring of PD-L1 expression is equivalent to manual scoring. The findings in this thesis may add to literature on which of the pre- and post-analytical issues are major contributors to the controversial results of PD-L1 expression's ability to predict immunotherapy response.

Methods

Material

Patient cohort

Initially 31 patients diagnosed with NSCLC of the type of adenocarcinoma, with pathological stage I-IIIA, at the University Hospital of North Norway (UNN), Tromsø, during 2017-2018 were considered for this study, as part of following clinical trial (ClinicalTrials.gov ID: NCT03299478). This was a consecutive collection of patients from Tromsø center by including those that have paired biopsy and resection samples. Available tumor blocks were collected, and patient demographic and clinico-pathological data were compiled into a database and de-identified (REDCap). Five patients were excluded from the study due to inadequate tumor tissue in the blocks, leaving 26 patients for the analysis. The tumors were staged according to the 8th edition of International Union Against Cancer's TNM classificatio[77].Histological classification was done according to the to the most recent updated version of World Health Organization guidelines[78]. One experienced pathologist (Elin Richardsen) blinded to any pathological or clinical information reviewed all the cases. The whole study (clinical trial: NCT03299478) was ethically approved by the Regional Committee (REK) and Norwegian Data Protection Organization.

PD-L1 immunohistochemistry staining and scoring

The biopsy and resection blocks were sectioned at a thickness of 4 µm and baked overnight at 37°C. The slides were processed using the Ventana Discovery-Ultra automated platform (Roche, Tucson, AZ) using validated anti-PD-L1 antibody (clone: SP263) as part of routine clinical care. Regarding scoring, multiple cell types, including tumor cells and tumor infiltration lymphocytes, in the tumor microenvironment may express PD-L1 detectable by IHC. Multiple PD-L1 scoring formulas are employed in literature, Doroshow et al. listed the most common formulas[79]. TPS is the scoring formula used in this thesis which is the percentage of PD-L1 expressing tumor cells, compared to the total amount of nucleated tumor cells in tumor core. The percentage of cells, and not the number makes the data more reproducible independent of slide thickness[80]. The percentage of tumor cells that were PD-L1+ was assessed manually and digitally in this study.

$$TPS = \frac{\text{nr. of PD} - \text{L1 stained tumor cells}}{\text{nr. of viable tumor cells}} * 100\%$$

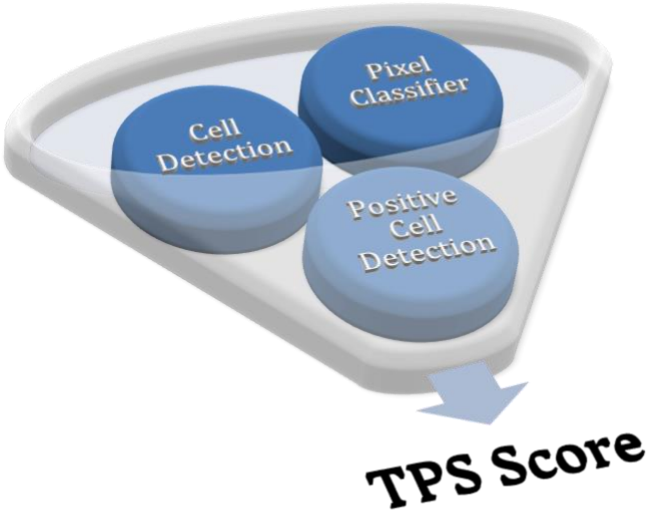
The scoring used a three-point scoring system (TPS ≥ 50%, 1-49%, <1%). A tumor core is scored strongly positive when TPS ≥ 50 %, i.e., above 50% of the tumor cells are PD-L1+. Intermediate score when TPS is between 1% and 49%, and negative score when TPS<1%, i.e., less than 1% of the tumor cells are PD-L1+. The TPS score can be transformed into a categorical variable: 0 = negative (TPS<1%); 1 = intermediate (1-49%); 2 = high (>50%) PD-L1 expression (***Figure 2***).

Image analysis

Digital histological slides

Both the biopsy and resected slides were digitalized using Panoramic 250 Flash III (0.24 microns/pixel, 3DHistech, Hungary) scanner. Supervised machine learning algorithms (QuPath v.0.3.0, Queen's University, Northern Ireland; open access program)[81-83] were sequentially employed to build an automated PD-L1 scoring model.

A



B

Training Pixel Classifier	Applying Pixel Classifier	Cell Detection	Classify by centroid	Positive Cell Detection	TPS score
•Training images from 50% of the sample were used to train the pixel classifier.	•The pixel classifier was employed on all the digital slides, labeling all the pixels as Tumor and NotTomor	•**Hematoxylin OD mean:** Threshod 0.05 was used to detect a nucleus object •The boundaries of the cells are superimposed on the classified pixels	•Classifies detected cells as Tumor or NotTumor depending on the class of the pixels of that belong to each nucleus.	•**Cytoplasm DAB OD mean threshold of 0.05** was used to define a cell as PD-L1-positive.	

Figure 2 Digital Scoring Algorithm. (A) An overview of the three central elements of the digital detection algorithm: Cell detection, Pixel classifier, Positive Cell Detection. (B) Detailed overview of the sequence of the algorithm employed in this study.

Automated digital scoring

The digital scoring algorithm involves the pixel classifier, followed by cell detection, and finally the positive cell detection algorithm to give the TPS score (***Figure 3***). The pixel tissue classifier was used in this study to automatically identify pixels that belong to tumor cells from non-tumor cells. The cell detection algorithm in Qupath[84] identifies the cells and superimposes these on the classified pixels. The "Classify by Centroid" program classifies the detected cells according to majority of the classified pixels in the nucleus. The positive cell detection algorithm then used the diaminobenzidine (DAB) intensity threshold to classify the tumor cells as positive or negative PD-L1 expressing tumor cells. TPS is automatically calculated, and the measurements are exported into an excel sheet.

Definition of PD-L1+ cell in digital model

PD-L1 positive cell is thus defined as a brown stained cell with mean cytoplasmic stain ≥ 0.05. "Cytoplasmic DAB optical density (DAB OD) mean", is the name of the vector on QuPath, and the mean cytoplasmic brown stain (DAB stain) is automatically generated for each cell. If a cell has a mean cytoplasmic brown stain ≥ 0.05, it is classified as PD-L1 positive cell. The Cytoplasm DAB OD mean threshold was manually compared with other possible thresholds (0.03, 0.08, 0.1). The threshold (0.05) had 96% sensitivity and 87% specificity (true negative rate), as calculated from the given manual count of positive and negative cells before and after threshold and was found to be better than the other juxtaposing thresholds. This was an ad-hoc solution, the threshold has a potential to be more precise, with more automated method of collecting and comparing measurements from QuPath, one could find a better cut-off.

A single threshold is used across all slides, both biopsy and resected specimens. The clinical trials that require PD-L1 testing are by convention categorical, partially because human eye is better suited for categorical scoring than continuous scoring, therefore no attempt was made to give a continuous manual score. Although multiple studies do include a continuous manual score[85,86]. All manual scores were given as categorical scores (<1%, 1-49%, and >50%) and all digital scores, automatically calculated as continuous scores, were later converted into corresponding categories for correlative statistics.

Statistical analysis

R programming with relevant packages was used for statistics and data visualization. A p-value <0.05 was chosen to be statistically significant. Weighted Cohen's kappa (κ) and intraclass correlation coefficient (two-way random-effects model with an absolute agreement definition) was employed to evaluate inter-observer agreement[87]. The receiver operating characteristic (ROC) is used to compare the performance of the automated digital scores with the pathologist manual scores. For graphing purposes, the TPS scores in their original percentage values (continuous variable from 0% to 100%) before being categorized into their respective cutoffs (0,1, and 2) were used.

Results

Clinical attributes/Data curation

The tissue that was included in this study was from 26 patients with adenocarcinoma of the NSCLC. The patients had median age of 65 years (range, 51-84 years), a male representing 53% of the cohort, with majority of heavy smokers (n=14, 54%) and only 2 never smokers, almost half had mixed pattern adenocarcinoma (n=12, 46%), and most had early stage adenocarcinoma at resection (n=18, 69%) ***Table 1***.

Table 1 Clinical characteristics of patients

Clinical characteristics	N (%)	**Tumor characteristics**	N (%)
Age (range)		**AC differentiation**	
≤65 (51-65) years	10 (38)	Single pattern	14 (54)
>65 (66-84) years	16 (62)	Mixed pattern	12 (46)
Gender		**AC subtype**	
Female	12 (46)	Solid	6 (23)
Male	14 (53)	Papillary	4 (15)
Smoking status		Acinar	2 (8)
Current smoker	8 (30)	Lepidic	2 (8)
Former smoker	15 (58)	Mixed pattern	12 (46)
Never smoker	3 (12)	**AC stage**	
Smoking history		Stage I	18 (69)
Smoked <20 pky	9 (35)	Stage II	3 (12)
Smoked >20 pky	14 (54)	Stage IIIA	5 (19)
Weight loss *		**ECOG PS**	
<10%	23 (88)	0	22 (85)
≥10%	3 (12)	1	4 (15)

Abbreviations: ECOG PS: Eastern Cooperative Oncology Group Performance Status; AC: adenocarcinoma of the lung; pky – pack-years, * weight loss in % of body weight.

Manual vs Digital Scoring

Examples of TPS calculation by ML model at PD-L1 subgroups are shown in ***Figure 3***. The digital and manual score agreement regarding PD-L1 expression with respect to clinical cutoffs were evaluated using ROC curves. The cut-offs, TPS≥1 and TPS≥50%, are used for comparison as they are the most common cut-offs used to evaluate PD-L1 IHC expression. The manual PD-L1 scores, which are the established method of evaluation PD-L1 expression, are used as ground truth values, and the accuracy of the digital scorer is determined in comparison to the manual score.

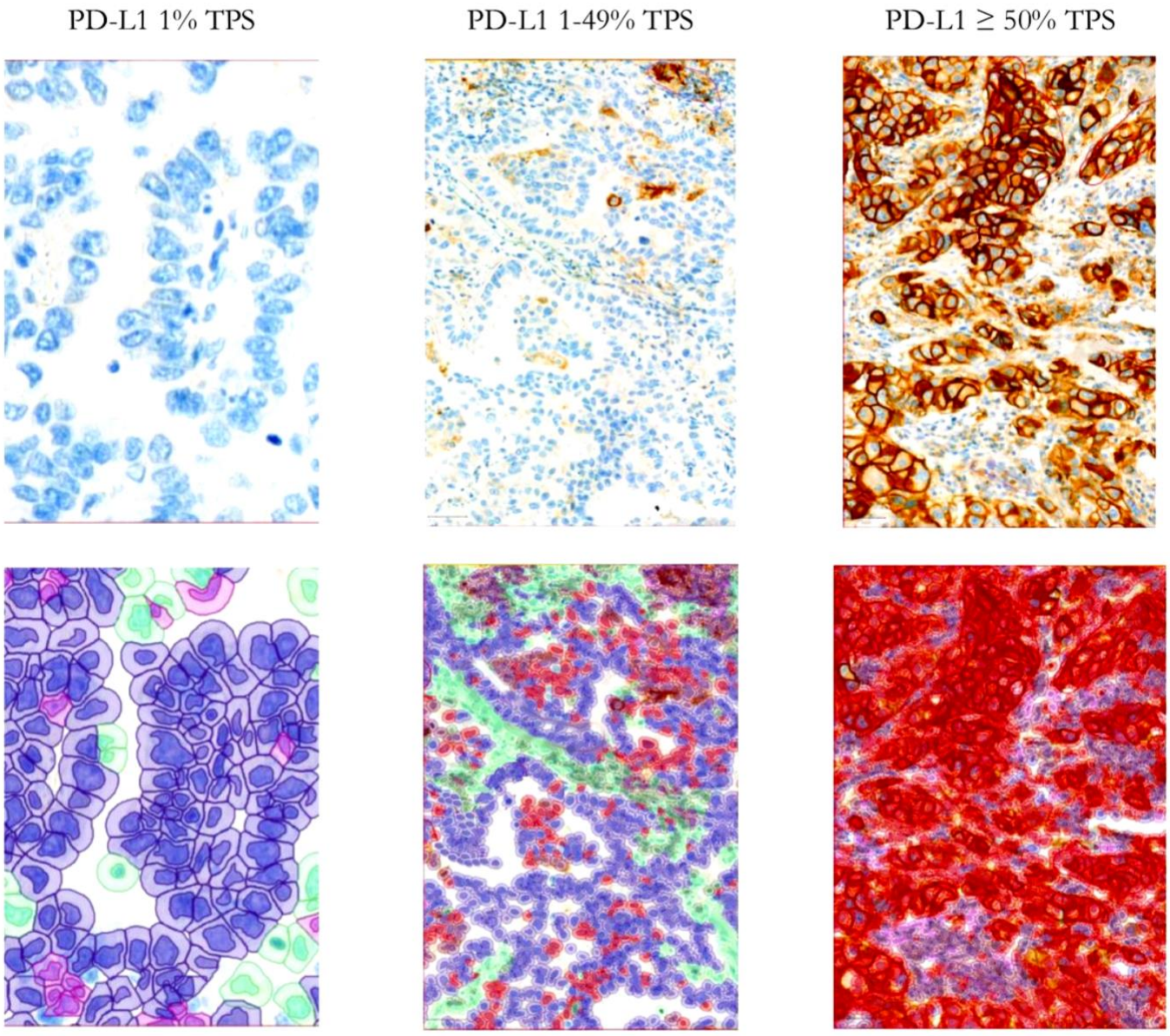
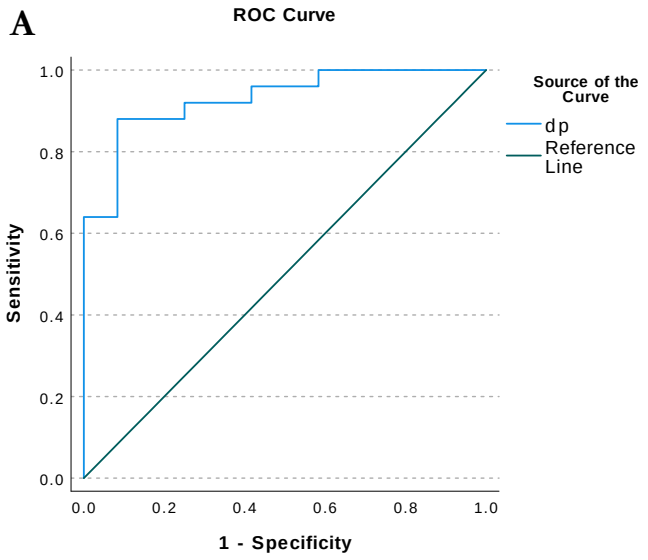
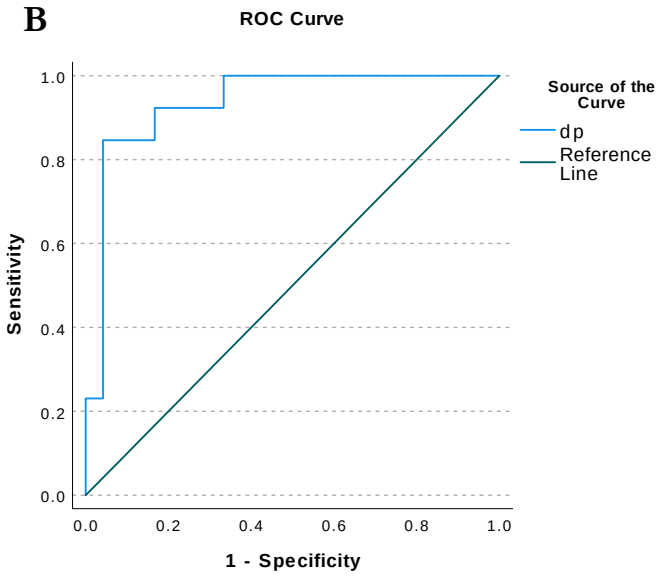


Figure 3 Representative Images of PD-L1 expression subgroups (1%, 1-49%, ≥50%) with machine learning overlay.

A



B



C

	Cutoff 1% (95% CI)	Cutoff 50% (95% CI)
AUC	0.93 *[0.85,1.10]*	0.94 *[0.87,1.20]*
Sensitivity	0.88 *[0.69,0.98]*	0.85 *[0.55,0.98]*
Specificity	0.92 *[0.62,1.00]*	0.96 *[0.79,1.00]*
Positive Predictive Value	0.96 *[0.77,0.99]*	0.92 *[0.64,1.00]*
Negative Predictive Value	0.79 *[0.52,0.99]*	0.92 *[0.72,1.00]*
Positive Likelihood Ratio	10.5 *[1.60, 69.3]*	20.3 *[2.9, 140.3]*
Negative Likelihood Ratio	0.13 *[0.05,0.38]*	0.16 *[0.05, 0.58]*

Figure 4 ROC Curves demonstrating the sensitivity and specificity of the digital scoring algorithm. (A) ROC Curve for cutoff of 1% and (B) ROC Curve for cutoff of 50%. The straight diagonal line represents a strictly random classifier. The closer the line is to the upper left corner represent the better classifier. (C) Table lists the numeric results of the ROC analysis. Accuracy (AUC) is the ratio of correctly predicted observation to the total observations. ROC curve – receiver operating characteristic curve, dp – digital percentage scoring compared to the manual percentage scoring.

Our highly automated digital PD-L1 scoring algorithm achieves an AUC-ROCs of 0.93 for 1% cutoff, corresponding to 88% sensitivity and 92% specificity for this cutoff (***Figure 4A and C***), while it achieves a slightly better test results for 50% cutoff with AUC of 0.94 corresponding to 85% sensitivity and 96% specificity of the digital scorer(***Figure 4B and C***).

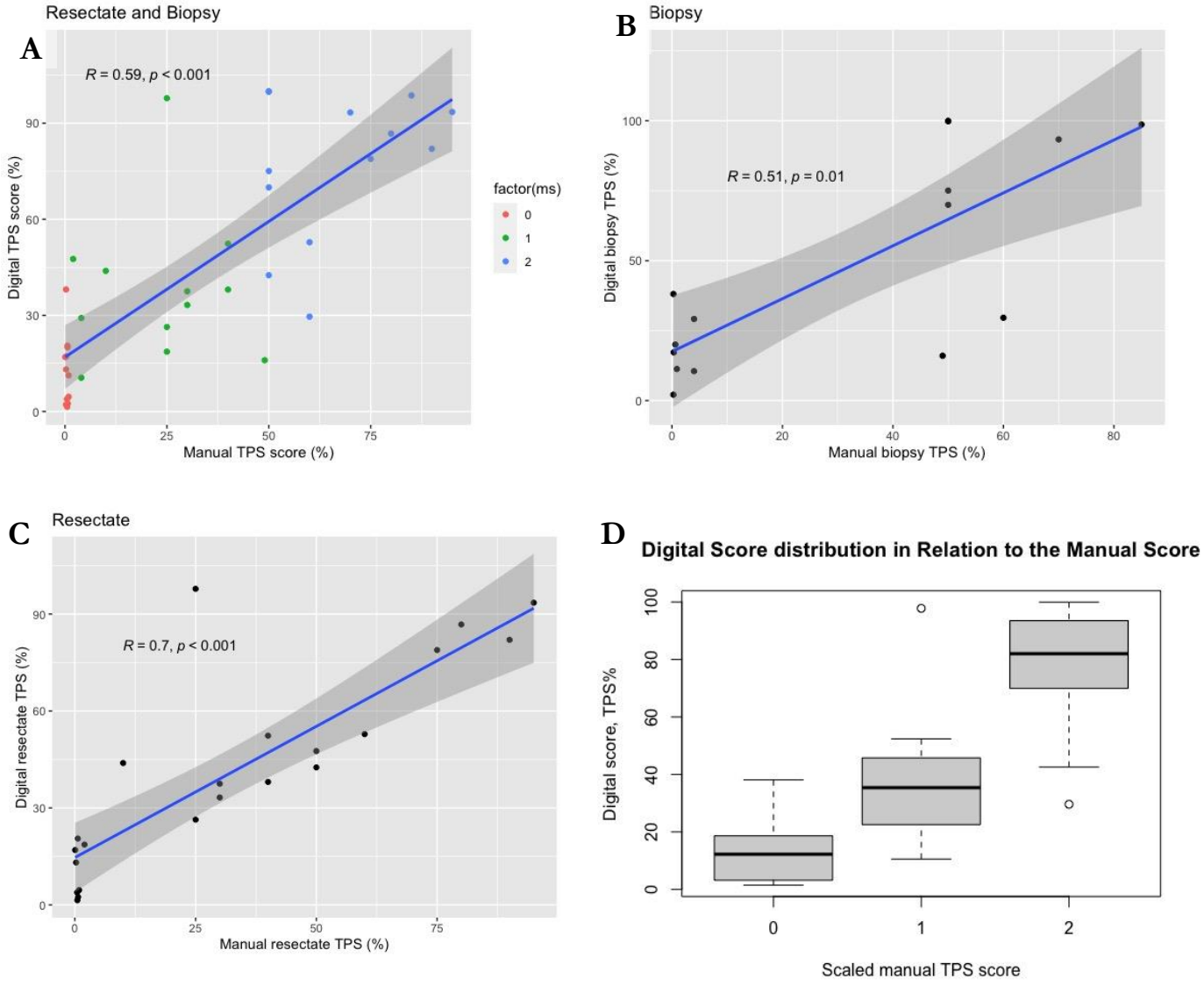


Figure 5 Correlation of tumor cell programmed death-ligand 1 (PD-L1) expression between, (A) paired manual score and digital score of all biopsy and resected tissue, (B) paired manual score and digital score of all biopsy tissue, (C) paired manual score and digital score of all resected tissue, (D) boxplot illustrating the distribution of the digital score in relation to the stratified manual score. The x-axis and the y-axis in A, B & C represent the percentage of total tumor cells that demonstrate membranous and/or cytoplasmic PD-L1 expression. R indicates the Kendall correlation coefficient.

Even though this is not a fair comparison as manual pathologist scores are more a semi-quantitative assessment, but using the scoring system as continuous variable, the digital scores showed a moderate correlation (***Figure 5A***,*R = 0.59, p<0.001*) with manual pathologist scores in the overall set including both biopsies and surgically resected samples. Similar results were also found when comparing scores for biopsy and resection samples when analyzed distinct (***biopsy: Figure 5B,*** *R=0.51, p = 0.01;* ***resected specimens: Figure 5C,*** *R=0.70, p <0.01*). The boxplot in ***Figure 5D*** demonstrates that the digital classifier quantitively scores clustering the scores around their respective true cutoffs in the overall cohort (Kruskal–Walli's test, $p < 0.001$).

PD-L1 Expression: Resected vs Biopsy specimens

Biopsy is the most common procedure to determine whether the tumor expresses PD-L1 for clinical decision making. The reliability of results from small biopsy samples to represent the true expression of the whole tumor tissue assumes that the tumor cells express PD-L1 homogenously across the tumor tissue. This assumption is investigated by comparing the manual scores of the resected specimens with the manual scores of the biopsy of the same tumor sample.

Manual scoring: At TPS cutoff ≥50% using manual score 35% biopsy slides were scored positive and 35% paired resection slides were scored positive, with a concordance rate of 82 % (*ICC=0.82, p<0.001,* A significant moderate-to-strong agreement observed between biopsy and resected specimens at TPS cutoff ≥ 50% (*κ=0.69, p<0.001*). While at TPS cutoff ≥ 1 % using manual score 55% of biopsy samples were scored positive and 75% of paired resections were scored positive, giving a non-significant concordance rate of 40 (ICC=0.444, *κ=0.28, p*=0.204). The results show no agreement between biopsy and resected specimens at TPS cutoff ≥1% (*κ=0.28, p*=0.2).

Digital scoring: Like the manual scoring results, moderate agreement was found at TPS cutoff ≥50% using the digital scoring algorithm to compare biopsies vs resected specimens (n=14, *ICC=0.70, p=0.01, κ=0.52*). No agreement was found at TPS cutoff ≥1% and the kappa value was not significant.

When treating the PD-L1 TPS as a continuous variable, in difference from the above that uses a categorical variable, no statistically significant consistency was found between the resected specimens scores and the biopsy scores (***Figure 6A***, *R = 0.18, p=0.274, κ=0.36*) with manual wise scores. Likewise, the same comparison showed no correlation with digital scores across biopsy and resection samples (***Figure 6B***, *R=0.25, p = 0.233, κ=0.38*).

Using 3-scale TPS cutoffs, even though there was an overall a significant relationship between the manual detection rates of resected and biopsy tissue samples (χ^2 = 10.09, p<0.05), which is supported by the V_{cramer} value of 0.41 indicating a relatively strong correlation (V_{cramer} = 0.41, $_{95\%}CI^{HDI}$= 0.23 – 0.57), a closer look at each cutoff confirmed weak correlation especially for the <1% TPS cutoff. The detection rate of biopsy vs resected specimens at different cutoffs (0: <1%, 1: 1-49%, and 2: ≥50%) is shown in ***Figure 6C manual scoring***. The digital scorer in similar manner to the manual scorer demonstrates the poor detection rate of PD-L1 at the 1% cutoff in biopsy versus resected specimens ***Figure 6D***. As discussed earlier, a moderate to strong agreement observed with PD-L1 IHC staining at cutoff of ≥50%, with an improved correlation in detection rate (p=0.07).

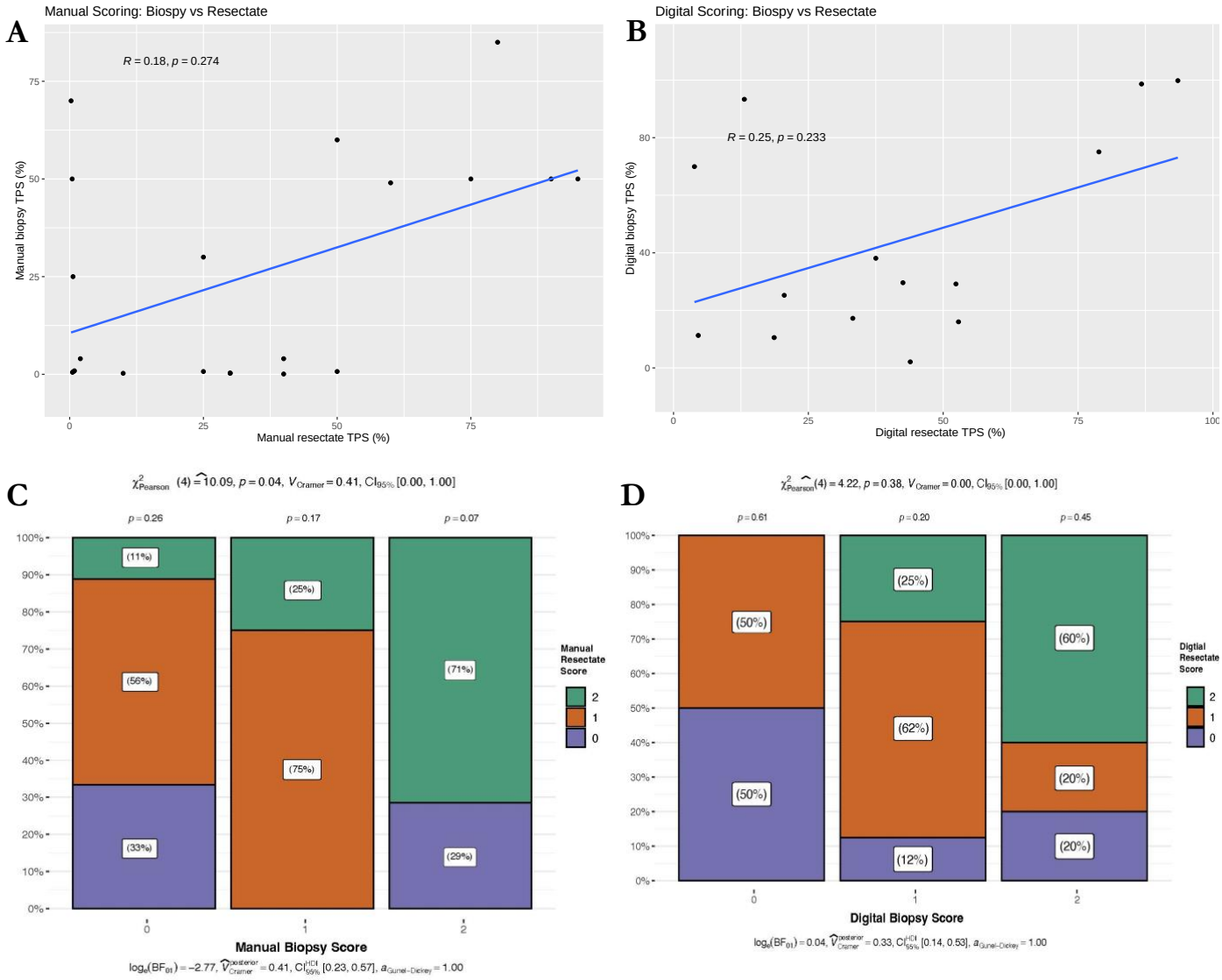


Figure 6 Correlation of tumor cell programmed death-ligand 1 (PD-L1) expression between biopsy and resected specimens, (A) paired manual score (TPS%) of biopsy vs resected slide, (B) paired digital score (TPS%) of biopsy vs resected slide, (C) paired manual score and digital score of all resected tissue, (D) boxplot illustrating the distribution of the digital score in relation to the stratified manual score. *The x-axis and the y-axis* in A, B & C *represent the percentage of total tumor cells that demonstrate membranous and/or cytoplasmic PD-L1 expression.* R *indicates the Kendall correlation coefficient.*

Discussion

In this relatively small sample sized study, an automated/ML-based algorithm was built to quantitively assess PD-L1 TPS. The PD-L1 score outputs of ML model showed a high harmony with manual scores of the pathologist. However, the main objective of the study was to analyze the concordance of PD-L1 expression level detection in paired biopsy and surgically resected specimens. With this regard, we found no significant difference in the detection rate of PD-L1 at the 50% cutoff between biopsy and resection regardless of scoring strategy, manual or digital. While there was a high heterogeneity and disconcordance in patients with low expression (1% TPS) of PD-L1 comparing biopsy with resection.

The predictive role of the PD-L1 biomarker

PD-L1, a crucial protein for immune escape, is a targetable biomarker, as NSCLC patients with especially high PD-L1 score are expected to greatly benefit from anti-PD-(L)1 treatment. The increased focus on targetable biomarkers has consequently increased the demand for efficient extraction of information from smaller tissue material obtained by minimally invasive techniques. This thesis adds to the limited available literature comparing PD-L1 IHC testing in biopsy and resected specimens[53,88]. In addition this thesis addresses the efficiency of automated digital scoring to further refine PD-L1 as a useful and objectively reliable predictive biomarker[79].

Factors that contribute to heterogenous PD-L1 detection

The predictive role of PD-L1 has shown varied results, it was for example found not to be predictive for more than 50% of patients in several studies[50], limiting its clinical value. Several parameters may account for PD-L1s limited utility, which majorly lies in the generalizability of the PD-L1 scoring results between studies. When comparing the results of this study to other studies, care must be taken to check the cancer type, tissue type, PD-L1 staining assay, and type of cell used for calculating the PD-L1 score are the same. Our study compared PD-L1 expression in matched biopsy & resection tissues of adenocarcinoma of the lung. The staining assay used for all the slides is VENTANA's SP263 antibody[89]. The types of tissues tested in this thesis are archival biopsy and/or resected tissue. Tumor cells are used to calculate the PD-L1 expression

using the tumor percentage score (TPS) formula. The TPS scores are usually manual scores which introduces uncertainty due to the well known intra- and interscorer variability[86].

Another issue that limits the generalizability of PD-L1 scores is the conventional cutoffs. This is especially due to PD-L1 being a relatively new biomarker, it is prone to increased error rates in its introduction phase, especially at the threshold cut-offs[89]. These PD-L1 scoring cut-offs were originally designed to achieve statistically predefined outcomes and do not reflect a uniform biological state[79]. The inaccuracy of the PD-L1 biomarker, such as found in a retrospective analysis of data (2011-2019) which identified that PD-L1 positivity was predictive for less than 30% of the patients[50], was in part be due to these conventional cut-offs.

Digital vs manual PD-L1 TPS

The need to eliminate issues related to manual PD-L1 IHC scoring method of surgical specimens – biopsy or resected tissue – has been the driving force for the research to improve the digital scoring method to initially aid the traditional manual method until it reaches its fully automated potential. The pixel classifier was found to be an adequate substitution of the manual classifier for segmentation and classification of histological material in various fields of pathology[90].

There was a moderate correlation ($r=0.59$, $p<0.001$) between the digital and manual scores using PD-L1 TPS as continuous variable. In line with previous reports[75,85], as the ROC curves in ***Figure 4*** demonstrate, the weakly supervised machine learned classifiers has great[63] discrimination capacity to distinguish PD-L1 positive from PD-L1 negative tumor with AUC-ROC of 0.93 cutoff 1%, this corresponds to 88% sensitivity and 92% specificity of the automated digital scorer for TPS ≥1%, i.e., an excellent positive predictive value of 96%, and a suboptimal negative predictive value of 79%. The digital scorer has higher sensitivity (96%) 50% respectively, while the majority of misclassification was reported to commonly occur around the 1% cut-off[89].

The discordance between the manual and digital scores in resected tissue is clustered around the 1-49% TPS. Most of the disagreements between digital and manual scoring lie at the extreme borders of the weak PD-L1 score: 1-49%., especially around the 1% cutoff. Significantly higher

percentage of cases in this study had 1-49% score on semi-automated compared to manual scoring, this is due to overestimation of low scores and underestimation of high scores (see ***Figure 5***). Previous studies also report that digital scoring overscores the borderline <1% and underscores the borderline >50% into TPS 1-49% category[86]. The clustering into 1-49% category was found to be independent of tumor type[86]. The different PD-L1 positivity definition applied between the digital scoring method (cytoplasmic staining) and the manual scoring methods (distinct membrane staining) might partially explain the remaining disagreement between the digital and the manual scoring methods.

Morphological heterogeneity

The machine learning (ML) algorithm employed in this thesis uses the pixel classifier to detect tumor cells, while object classifiers are still in use in other research communities. The object classifier was first tested for this thesis, but it was found to require heavy manual supervision[62,84]. Heavy manual annotations of numerous whole slide images were required for training the object classifier, this albeit being manually tasking, was not able compensate for the morphological heterogeneity of adenocarcinoma tumor cells (adenocarcinoma subtypes: solid, papillary, acinar and lepidic, and even undifferentiated/mixed AC) and to distinguish tumor from immune cell accurately. Identifying tumor from immune cell was specially challenging for the object classifier. A pixel classifier in the other hand is not dependent on the morphological parameters of the cell-object. It uses texture and color in each pixel to learn to distinguish tumor from not tumor, with astounding accuracy and without being confounded by the morphological heterogeneity of adenocarcinoma cells in this small sample, achieving a more generalizable detection algorithm[85].

Biopsy vs Resected specimens

Treatment decisions are often based on the PD-L1 score of biopsies. Although PD-L1 staining of resected specimens is not a requirement for clinical decision making, it was done in this study as paired tissues were available for the patients who were operable at the time of diagnosis. This made it possible to address the question of whether PD-L1 staining of resected specimens showed any significant difference from that found by biopsy.

This thesis had 26 paired cases for comparing biopsy vs resected specimens scores. To evaluate how the results of PD-L1 estimation are influenced by the tissue sample size, a concordance analysis was done comparing the TPS scores of the biopsy to the TPS scores of the resected specimens. The results show an overall significant relationship between the detection rates for resected and biopsy tissue samples (***Figure 6C,*** $p<0.05$). At TPS cutoff ≥ 1 % using either manual score or digital scores the results show no correlation. At TPS cutoff ≥50% using manual score the results show a significantly better concordance rate of 80 % ($p<0.05$). Comparing PD-L1 results to the literature have variability issues. A study[92] using 80 cases showed a concordance rate of 92.4% with a tumor and immune cell score of ≥ 1 as positive for PD-L1 staining. Another study[88], in line to our findings, did not find satisfactory concordance when using the 1% cutoff, but found a concordance of small biopsy and surgical resection of 92.2% when using a cut off ≥50% tumor staining with PD-L1.

Heterogeneity of PD-L1 expression in a tumor core

Only rarely does a tumor core heterogeneously express the PD-L1 protein. PD-L1 expression in a single tumor core was not found to be significantly heterogenous when comparing ROIs within each tumor core with overall tumor core score. Four ROIs manually drawn on each tumor core of a resected specimen were manually scored before an overall tumor core score was given. The four ROIs most often had the same PD-L1 score. Only 2 out 26 WSI had more than one discordant PD-L1 score. In total, only 8 out of 88 ROIs were discordant to their respective tumor core, this gives a concordance rate of 91%. Thus, 9% of the discordance might be due to PD-L1 heterogeneity in tumor core.

The weak concordance rate found in the results might also be due to the time gap from biopsy of tumor to surgical resection. The treatment the patient may have received in this time space, and the evolution of the tumor in response of treatment might affect the PD-L1 expression of resected tissue compared to biopsy. Although the patients in this cohort did not receive neoadjuvant therapy in the time between biopsy and resection. Thus, both the resected and biopsy methods may still be considered equivalent in terms of tumor immune microenvironment.

Limitations of the study

The sample size of this study was relatively small, consequently the results of this analysis need to be verified on larger cohorts. The cohort had mainly early-stage adenocarcinoma, there is potential room to further explore the ML model in advanced stage adenocarcinoma, most of whom have inoperable lung cancer at diagnosis, the biopsy is often the only tissue specimen available for immunotherapy treatment decision. This makes the findings regarding PD-L1 digital quantification from this thesis even more valuable in advanced stage adenocarcinoma. This study is limited to adenocarcinoma, while nearly 25% of the NSCLC are squamous cell carcinoma, therefore results need to be verified on those patients as well as adenocarcinoma subtypes.

Conclusion

Biopsy was found to be equivalent to resected specimens when determining PD-L1 expression of lung adenocarcinoma at a particular cutoff (50% TPS), this could not be attributed to the heterogeneity of tumor cells in tissue core but might be due to pre-analytical issues affecting PD-L1 staining, such as time between biopsy and resected specimens' acquisition from the patient. This thesis demonstrates the robustness of fully automated pixel classifier in detecting tumor cells and the positive cell detection algorithm in detecting PD-L1 positive cells. Automated digital scoring of IHC slides could be of help for the pathologist. Digital pathology can reliably be used alone in clear-cut cases, and as a companion diagnostic aid in difficult cases[91].The precision of the classifier may further be improved by increasing the exposure to varying histological presentation of tumor cells. The current ML model, albeit being trained on small dataset, had the advantage of exhaustive pixel-wise labels, with the potential of generalizability to clinical-grade and real-world data.

Bibliography

1. Mason G. CANCER OF THE LUNG REVIEW OF A THOUSAND CASES. *The Lancet.* 1949;254(6579):587-591.
2. Hoffman PC, Mauer AM, Vokes EE. Lung cancer. *Lancet.* 2000;355(9202):479-485.
3. Imyanitov EN, Iyevleva AG, Levchenko EV. Molecular testing and targeted therapy for non-small cell lung cancer: Current status and perspectives. *Critical Reviews in Oncology/Hematology.* 2021;157:103194.
4. Safiri S, Sohrabi MR, Carson-Chahhoud K, et al. Burden of Tracheal, Bronchus, and Lung Cancer and Its Attributable Risk Factors in 204 Countries and Territories, 1990 to 2019. *J Thorac Oncol.* 2021;16(6):945-959.
5. Canceratlas.cancer.org. The burden of lung cancer. *The Cancer Atlas* 2018; https://canceratlas.cancer.org/the-burden/lung-cancer/, 2021.
6. Stinchcombe TE, Lee CB, Socinski MA. Current approaches to advanced-stage non-small-cell lung cancer: first-line therapy in patients with a good functional status. *CLIN LUNG CANCER.* 2006;7 Suppl 4:S111-117.
7. Lara PN, Jr., Gandara DR, Natale RB. Randomized phase III trial of cisplatin/irinotecan versus cisplatin/etoposide in patients with extensive-stage small-cell lung cancer. *CLIN LUNG CANCER.* 2006;7(5):353-356.
8. Schiller JH, Harrington D, Belani CP, et al. Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer. *N Engl J Med.* 2002;346(2):92-98.
9. Stinchcombe TE, Socinski MA. Treatment Paradigms for Advanced Stage Non-small Cell Lung Cancer in the Era of Multiple Lines of Therapy. *Journal of Thoracic Oncology.* 2009;4(2):243-250.
10. Ritchie H, Roser M. Causes of Death. *Our World in Data* 2019; https://ourworldindata.org/causes-of-death. Accessed 14. may 2021.
11. Society AC. Known and Probable Human Carcinogens. *American Cancer Society medical and editorial content team* 2019; August 14, 2019:https://www.cancer.org/cancer/cancer-causes/general-info/known-and-probable-human-carcinogens.html.
12. Stine JK. 20th century tobacco advertisments. 2014; https://americanhistory.si.edu/blog/2014/03/smoke-gets-in-your-eyes-20th-century-tobacco-advertisements.html.
13. Watson WL, Berg JW. Oat cell lung cancer. *Cancer.* 1962;15:759-768.
14. Mann JR, DuBois RN. Cancer chemoprevention: myth or reality? *Drug Discovery Today: Therapeutic Strategies.* 2004;1(4):403-409.
15. Soria JC, Kim ES, Fayette J, Lantuejoul S, Deutsch E, Hong WK. Chemoprevention of lung cancer. *Lancet Oncol.* 2003;4(11):659-669.
16. Wagner GR. Asbestosis and silicosis. *The Lancet.* 1997;349(9061):1311-1315.
17. Shinton NK. THE HISTOLOGICAL CLASSIFICATION OF LOWER RESPIRATORY TRACT TUMOURS. *British journal of cancer.* 1963;17(2):213-221.
18. Hanna JM, Onaitis MW. Cell of origin of lung cancer. *J Carcinog.* 2013;12:6-6.
19. Ferreira CG, Huisman C, Giaccone G. Novel approaches to the treatment of non-small cell lung cancer. *Critical Reviews in Oncology/Hematology.* 2002;41(1):57-77.
20. Yazawa T, Kamma H, Fujiwara M, et al. Lack of class II transactivator causes severe deficiency of HLA-DR expression in small cell lung cancer. *J Pathol.* 1999;187(2):191-199.
21. Belani CP, Dakhil S, Waterhouse DM, et al. Randomized phase II trial of gemcitabine plus weekly versus three-weekly paclitaxel in previously untreated advanced non-small-cell lung cancer. *Ann Oncol.* 2007;18(1):110-115.
22. Moon Y, Lee KY, Sung SW, Park JK. Differing histopathology and prognosis in pulmonary adenocarcinoma at central and peripheral locations. *J Thorac Dis.* 2016;8(1):169-177.
23. Schabath MB, Cote ML. Cancer Progress and Priorities: Lung Cancer. *Cancer Epidemiology Biomarkers & Prevention.* 2019;28(10):1563.
24. Kerr KM, Bibeau F, Thunnissen E, et al. The evolving landscape of biomarker testing for non-small cell lung cancer in Europe. *Lung Cancer.* 2021;154:161-175.

25. Lindeman NI, Cagle PT, Aisner DL, et al. Updated Molecular Testing Guideline for the Selection of Lung Cancer Patients for Treatment With Targeted Tyrosine Kinase Inhibitors: Guideline From the College of American Pathologists, the International Association for the Study of Lung Cancer, and the Association for Molecular Pathology. *The Journal of Molecular Diagnostics.* 2018;20(2):129-159.
26. Selby P, Gillis C, Haward R. Benefits from specialised cancer care. *The Lancet.* 1996;348(9023):313-318.
27. Wozniak A. Challenges in the current antiangiogenic treatment paradigm for patients with non-small cell lung cancer. *Crit Rev Oncol Hematol.* 2012;82(2):200-212.
28. Klempner SJ, Ou S-HI, Costa DB, et al. The Clinical Use of Genomic Profiling to Distinguish Intrapulmonary Metastases From Synchronous Primaries in Non–Small-Cell Lung Cancer: A Mini-Review. *CLIN LUNG CANCER.* 2015;16(5):334-339.e331.
29. Blasberg JD, Pass HI, Donington JS. Sublobar Resection: A Movement from the Lung Cancer Study Group. *Journal of Thoracic Oncology.* 2010;5(10):1583-1593.
30. Rosell R, Gomez-Codina J, Camps C, et al. A randomized trial comparing preoperative chemotherapy plus surgery with surgery alone in patients with non-small-cell lung cancer. *N Engl J Med.* 1994;330(3):153-158.
31. Herbst RS, Lynch TJ, Sandler AB. Beyond doublet chemotherapy for advanced non-small-cell lung cancer: combination of targeted agents with first-line chemotherapy. *CLIN LUNG CANCER.* 2009;10(1):20-27.
32. Wakelee H, Dubey S, Gandara D. Optimal adjuvant therapy for non-small cell lung cancer--how to handle stage I disease. *Oncologist.* 2007;12(3):331-337.
33. Baxevanos P, Mountzios G. Novel chemotherapy regimens for advanced lung cancer: have we reached a plateau? *Ann Transl Med.* 2018;6(8):139.
34. Isomoto K, Haratani K, Hayashi H, et al. Impact of EGFR-TKI Treatment on the Tumor Immune Microenvironment in EGFR Mutation-Positive Non-Small Cell Lung Cancer. *Clin Cancer Res.* 2020;26(8):2037-2046.
35. Davies AM, Ho C, Lara PN, Mack P, Gumerlock PH, Gandara DR. Pharmacodynamic Separation of Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitors and Chemotherapy in Non–Small-Cell Lung Cancer. *CLIN LUNG CANCER.* 2006;7(6):385-388.
36. Johnson BE, Janne PA. Rationale for a phase II trial of pertuzumab, a HER-2 dimerization inhibitor, in patients with non-small cell lung cancer. *Clin Cancer Res.* 2006;12(14 Pt 2):4436s-4440s.
37. Cesano A, Marincola FM, Thurin M. Status of Immune Oncology: Challenges and Opportunities. *Methods Mol Biol.* 2020;2055:3-21.
38. Rousseaux S, Debernardi A, Jacquiau B, et al. Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers. *Sci Transl Med.* 2013;5(186):186ra166.
39. Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer.* 2012;12(4):252-264.
40. Uruga H, Mino-Kenudson M. Predictive biomarkers for response to immune checkpoint inhibitors in lung cancer: PD-L1 and beyond. *Virchows Archiv.* 2021;478(1):31-44.
41. Dong H, Strome SE, Salomao DR, et al. Tumor-associated B7-H1 promotes T-cell apoptosis: a potential mechanism of immune evasion. *Nat Med.* 2002;8(8):793-800.
42. Nishimura H, Nose M, Hiai H, Minato N, Honjo T. Development of lupus-like autoimmune diseases by disruption of the PD-1 gene encoding an ITIM motif-carrying immunoreceptor. *Immunity.* 1999;11(2):141-151.
43. Topalian SL, Hodi FS, Brahmer JR, et al. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N Engl J Med.* 2012;366(26):2443-2454.
44. Yu J, Wang X, Teng F, Kong L. PD-L1 expression in human cancers and its association with clinical outcomes. *OncoTargets and Therapy.* 2016;Volume 9:5023-5039.
45. Kelly RJ, Gulley JL, Giaccone G. Targeting the Immune System in Non–Small-Cell Lung Cancer: Bridging the Gap Between Promising Concept and Therapeutic Reality. *CLIN LUNG CANCER.* 2010;11(4):228-237.
46. Okazaki T, Honjo T. PD-1 and PD-1 ligands: from discovery to clinical application. *International Immunology.* 2007;19(7):813-824.
47. Herbst RS, Giaccone G, de Marinis F, et al. Atezolizumab for First-Line Treatment of PD-L1–Selected Patients with NSCLC. *New England Journal of Medicine.* 2020;383(14):1328-1339.

48. www.cancertherapyadvisor.com. Cancer Therapy. 2022; https://www.cancertherapyadvisor.com/home/cancer-topics/lung-cancer/lung-cancer-treatment-regimens-landing-page/non-small-cell-lung-cancer-treatment-regimens/.
49. Kauffmann-Guerrero D, Tufman A, Kahnert K, et al. Response to Checkpoint Inhibition in Non-Small Cell Lung Cancer with Molecular Driver Alterations. *Oncol Res Treat.* 2020;43(6):289-298.
50. Davis AA, Patel VG. The role of PD-L1 expression as a predictive biomarker: an analysis of all US Food and Drug Administration (FDA) approvals of immune checkpoint inhibitors. *J Immunother Cancer.* 2019;7(1):278.
51. Proto C, Ferrara R, Signorelli D, et al. Choosing wisely first line immunotherapy in non-small cell lung cancer (NSCLC): what to add and what to leave out. *Cancer Treat Rev.* 2019;75:39-51.
52. Ilie M, Long-Mira E, Bence C, et al. Comparative study of the PD-L1 status between surgically resected specimens and matched biopsies of NSCLC patients reveal major discordances: a potential issue for anti-PD-L1 therapeutic strategies. *Ann Oncol.* 2016;27(1):147-153.
53. Wang Y, Wu J, Deng J, She Y, Chen C. The detection value of PD-L1 expression in biopsy specimens and surgical resection specimens in non-small cell lung cancer: a meta-analysis. *J Thorac Dis.* 2021;13(7):4301-4310.
54. Li C, Huang C, Mok TS, et al. Comparison of 22C3 PD-L1 Expression between Surgically Resected Specimens and Paired Tissue Microarrays in Non-Small Cell Lung Cancer. *J Thorac Oncol.* 2017;12(10):1536-1543.
55. Munari E, Zamboni G, Sighele G, et al. Expression of programmed cell death ligand 1 in non-small cell lung cancer: Comparison between cytologic smears, core biopsies, and whole sections using the SP263 assay. *Cancer Cytopathol.* 2019;127(1):52-61.
56. Hernandez A, Brandler TC, Zhou F, Moreira AL, Schatz-Siemers N, Simsir A. Assessment of Programmed Death-Ligand 1 (PD-L1) Immunohistochemical Expression on Cytology Specimens in Non-Small Cell Lung Carcinoma. *Am J Clin Pathol.* 2019;151(4):403-415.
57. Russell-Goldman E, Kravets S, Dahlberg SE, Sholl LM, Vivero M. Cytologic-histologic correlation of programmed death-ligand 1 immunohistochemistry in lung carcinomas. *Cancer Cytopathol.* 2018;126(4):253-263.
58. Mino-Kenudson M. Programmed cell death ligand-1 (PD-L1) expression by immunohistochemistry: could it be predictive and/or prognostic in non-small cell lung cancer? *Cancer Biol Med.* 2016;13(2):157-170.
59. Wu J, Liu C, Liu X, et al. Artificial intelligence-assisted system for precision diagnosis of PD-L1 expression in non-small cell lung cancer. *Mod Pathol.* 2021.
60. Gould MK, Huang BZ, Tammemagi MC, Kinar Y, Shiff R. Machine Learning for Early Lung Cancer Identification Using Routine Clinical and Laboratory Data. *Am J Respir Crit Care Med.* 2021;204(4):445-453.
61. Avanzo M, Stancanello J, Pirrone G, Sartor G. Radiomics and deep learning in lung cancer. *Strahlenther Onkol.* 2020;196(10):879-887.
62. Silva F, Pereira T, Neves I, et al. Towards Machine Learning-Aided Lung Cancer Clinical Routines: Approaches and Open Challenges. *J Pers Med.* 2022;12(3).
63. Campanella G, Hanna MG, Geneslaw L, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med.* 2019;25(8):1301-1309.
64. Lee G, Sparks R, Ali S, et al. Co-occurring gland angularity in localized subgraphs: predicting biochemical recurrence in intermediate-risk prostate cancer patients. *PLoS ONE.* 2014;9(5):e97954.
65. van der Laak J, Litjens G, Ciompi F. Deep learning in histopathology: the path to the clinic. *Nature Medicine.* 2021;27(5):775-784.
66. Bulten W, Pinckaers H, van Boven H, et al. Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. *Lancet Oncol.* 2020;21(2):233-241.
67. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *Jama.* 2017;318(22):2199-2210.
68. Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology — new tools for diagnosis and precision oncology. *Nature Reviews Clinical Oncology.* 2019;16(11):703-715.

69. Chen J, Srinivas C. Automatic Lymphocyte Detection in H&E Images with Deep Neural Networks. *CoRR*. 2016;abs/1612.03217.
70. Chen J, Jiang CC, Jin L, Zhang XD. Regulation of PD-L1: a novel role of pro-survival signalling in cancer. *Ann Oncol*. 2016;27(3):409-416.
71. Liu Y, Gadepalli K, Norouzi M, et al. Detecting Cancer Metastases on Gigapixel Pathology Images. *arXiv*. 2017.
72. Corredor G, Wang X, Zhou Y, et al. Spatial Architecture and Arrangement of Tumor-Infiltrating Lymphocytes for Predicting Likelihood of Recurrence in Early-Stage Non-Small Cell Lung Cancer. *Clin Cancer Res*. 2019;25(5):1526-1534.
73. Wang X, Barrera C, Velu P, et al. Computer extracted features of cancer nuclei from H&E stained tissues of tumor predicts response to nivolumab in non-small cell lung cancer. *Journal of Clinical Oncology*. 2018;36(15_suppl):12061-12061.
74. Barrera C, Velu P, Bera K, et al. Computer-extracted features relating to spatial arrangement of tumor infiltrating lymphocytes to predict response to nivolumab in non-small cell lung cancer (NSCLC). *Journal of Clinical Oncology*. 2018;36(15_suppl):12115-12115.
75. Humphries MP, Hynes S, Bingham V, et al. Automated Tumour Recognition and Digital Pathology Scoring Unravels New Role for PD-L1 in Predicting Good Outcome in ER-/HER2+ Breast Cancer. *Journal of Oncology*. 2018;2018:2937012.
76. Pao W, Ooi CH, Birzele F, et al. Tissue-Specific Immunoregulation: A Call for Better Understanding of the "Immunostat" in the Context of Cancer. *Cancer Discov*. 2018;8(4):395-402.
77. Goldstraw P, Chansky K, Crowley J, et al. The IASLC lung cancer staging project: proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM classification for lung cancer. *Journal of Thoracic Oncology*. 2016;11(1):39-51.
78. Travis WD, Brambilla E, Nicholson AG, et al. The 2015 World Health Organization classification of lung tumors: impact of genetic, clinical and radiologic advances since the 2004 classification. *Journal of thoracic oncology*. 2015;10(9):1243-1260.
79. Doroshow DB, Bhalla S, Beasley MB, et al. PD-L1 as a biomarker of response to immune-checkpoint inhibitors. *Nature Reviews Clinical Oncology*. 2021;18(6):345-362.
80. Al-Shibli K. *The prognostic significance of the innate and adaptive immune systems in non-small cell lung carcinoma*. Tromsø, Norway: Institute for Mecdical Biology, Pathology unit Nordland Central Hopsital Bodø, UiT - University of Tromsø; 2010.
81. Bankhead P, Loughrey MB, Fernandez JA, et al. QuPath: Open source software for digital pathology image analysis. *Sci rep*. 2017;7(1):16878.
82. Qupath da. Cell Classification - QuPath 0.2.3 Documentation. 2019-2021; https://qupath.readthedocs.io/en/latest/docs/tutorials/cell_classification.html, 2021.
83. Bankhead P. QuPath tutorial #2 - IHC analysis. 2018; https://www.youtube.com/watch?v=aTVfJk6yNKs&list=PL4ta8RxZklWk_O_Z7K0bZlhmHtaH73vlh .
84. Bankhead P, Research_Associate. Cell Detection. 2019; https://forum.image.sc/t/qupath-intro-choose-your-own-analysis-adventure/27906/11.
85. Puladi B, Ooms M, Kintsler S, et al. Automated PD-L1 Scoring Using Artificial Intelligence in Head and Neck Squamous Cell Carcinoma. *Cancers (Basel)*. 2021;13(17).
86. Naso JR, Povshedna T, Wang G, et al. Automated PD-L1 Scoring for Non-Small Cell Lung Carcinoma Using Open-Source Software. *Pathol Oncol Res*. 2021;27:609717.
87. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)*. 2012;22(3):276-282.
88. Gradecki SE, Grange JS, Stelow EB. Concordance of PD-L1 Expression Between Core Biopsy and Resection Specimens of Non-Small Cell Lung Cancer. *Am J Surg Pathol*. 2018;42(8):1090-1094.
89. Keppens C, Dequeker EMC, Pauwels P, Ryska A, 't Hart N, von der Thüsen JH. PD-L1 immunohistochemistry in non-small-cell lung cancer: unraveling differences in staining concordance and interpretation. *Virchows Archiv*. 2021;478(5):827-839.
90. Vizcaíno A, Sánchez-Cruz H, Sossa H, Quintanar JL. Pixel-Wise Classification in Hippocampus Histological Images. *Comput Math Methods Med*. 2021;2021:6663977-6663977.

91. Humphries MP, McQuaid S, Craig SG, et al. Critical Appraisal of Programmed Death Ligand 1 Reflex Diagnostic Testing: Current Standards and Future Opportunities. *Journal of Thoracic Oncology*. 2019;14(1):45-53.
92. Kitazono S, Fujiwara Y, Tsuta K, et al. Reliability of Small Biopsy Samples Compared With Resected Specimens for the Determination of Programmed Death-Ligand 1 Expression in Non–Small-Cell Lung Cancer. *CLIN LUNG CANCER*. 2015;16(5):385-390.