PD-L1 expression and prediction of response to immune modulators in non-small cell lung cancer; reasons for its fragility and strategies to reduce it

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

by

Alexander Kieron Haragan

May 2021

Dedication

For my wife, Rachel, and our best friend, Jorvik.

Abstract

Introduction

Anti-PD-1/PD-L1 immunomodulatory (IM) therapy has revolutionised the treatment of non-small cell lung cancer (NSCLC). The only 'biomarker' currently-validated for predicting response of these tumours to IM therapy is the extent of PD-L1 expression as detected by immunohistochemistry (IHC). Despite the overall success of this therapy in patients with NSCLC, PD-L1 expression is an imperfect predictor, some patients with tumours displaying low expression responding strongly, and some with high expression not at all. The thesis considers why PD-L1 expression is an imperfect predictor and how it might be improved.

Methods

The research described in the first part of this thesis considered the impact of pre-analytical conditions on PD-L1 expression. This examined not only the effect of how tumours are sampled, but the influence of specimen processing and fixation and conditions of storage, the latter employing a novel tissue ageing acceleration chamber and mass-spectrometry. The second part describes examination of heterogeneity of expression in a series of resected NSCLCs in which the primary tumour was accompanied by nodal metastases. Biological and artefactual heterogeneity within and between tumour deposits was assessed at different scales using a novel 'squares method' and 'digital sampling'. The third part describes assessment of the tumour immune environment (TME), specifically interrogation of immune cell populations, employing a combination of techniques including traditional IHC, multiplex IHC, multiplex immunofluorescence and image analysis. The fourth and final part of the work involved an assessment of digital pathology and image analysis with integrated machine-learning algorithms as a tool to improve accuracy and consistency in assessing PD-L1 expression.

Results

PD-L1 expression is consistent across different types of specimen; loss of its immunogenicity can be reduced by storage in cold and dry conditions, particularly when combined with a desiccant. Approximately 20-25% of resected NSCLCs demonstrated tumoural heterogeneity such that sampling from different sites might produce clinically-relevant differences in PD-L1 expression. This can be minimised, but not reduced entirely, by generous sampling. The TME of NSCLCs can be differentiated by assessing different immune cell populations, but only in specimens containing sufficient tissue and routine, small, diagnostic specimens will prove difficult to analyse in this way. Image analysis and algorithms are potentially powerful tools that can reduce intra- and inter-observer consistency when assessing PD-L1 expression, but require learning and experience for their effective use.

Discussion

The research described in this thesis confirms that assessment of PD-L1 expression by IHC is a powerful, but imperfect biomarker, and indicates also that its utility can be improved. Accuracy and consistency in its interpretation can be increased by optimising pre-analytical conditions. Tumour heterogeneity is a more complex problem; whilst availability of multiple, generous, good quality samples improves accuracy, the confounding effect of this fundamental fact of the biology of PD-L1 expression cannot be removed entirely. Techniques to interrogate the TME yield powerful data but, at present, most are too expensive, too complicated and require too much tissue to be useful in the routine clinical setting. Image analysis, machine learning and algorithms are becoming established techniques and are clearly of value, but possibly largely in improving confidence in and consistency of interpretation.

Tables of Contents

Title Page I

Dedication II

Abstract III

Table of Contents ..... IV-XII

Acknowledgments ..... XIII-XIV

List of abbreviations ..... XV-XIX

List of tables ..... XX-XXI

List of figures ..... XXII-XXV

Published work ..... XXVI-XXVII

Chapter 1 - Introduction 1

1.0 Lung cancer and treatment options 1

1.0.0 Development of stratified treatments in advanced NSCLC 1

1.0.1 History of cancer immunotherapy 3

1.1 PD-1 and PD-L1; a brief introduction 5

1.1.0 PD-1 5

1.1.1 PD-L1 and PD-L2 7

1.2 Cancer and the immune system 8

1.2.0 Cancer cells and immune cells 8

1.2.1 Immunoediting 8

1.3 The Tumour microenvironment 9

1.3.0 Components of the TME 9

1.3.1 T-cells in the TME 10

1.3.2 Other immune cells within the TME 13

1.3.3 Cytokine activity within the TME 15

1.4 The role of PD-1/PD-L1 in mediating tumour immune escape 16

1.4.0 Alternative mechanisms of action 16

1.4.1 Regulation of PD-1 expression 17

1.4.2 Regulation of PD-L1 expression – genetic and epigenetic factors 17

1.4.3 Regulation of PD-L1 expression – cytokines and environmental factors 18
1.5 PD-1/PD-L1 inhibitor resistance 21
1.5.0 Primary resistance – PD-L1 heterogeneity and constitutive expression 21
1.5.1 Primary resistance – TILs and T-cell dysfunction 21
1.5.2 Primary resistance – TMB/Neoantigenicity 23
1.5.3 Primary resistance – other inhibitory mechanisms 24
1.5.4 Secondary resistance 25
1.6 Predicting response to PD-1/PD-L1 IM therapy 26
1.6.0 Immunohistochemistry: a brief overview of a typical pathway. 26
1.6.1 PD-L1 IHC to guide IM therapy 29
1.6.2 Stratified medicine and PD-L1 as a biomarker 30
Chapter 2 – Literature Review 35
2.0 Initial findings 35
2.1 Pre-analytics 37
2.1.0 Pre-analytics - Fixation 38
2.1.1 Pre-analytics – Specimen Age 42
2.1.2 Pre-analytics – Specimen types 47
2.1.2.0 Suitability of cytology specimens 48
2.1.2.1 Matched cytology and histology specimens 49
2.1.2.2 Unpaired cytology and histology specimens 52
2.1.3 Pre-analytics – Other 55
2.2 Analytics 57
2.2.0 Analytics –PD-L1 clones, LDTs and inter-pathologist concordance 57
2.2.0.0 Early papers comparing PD-L1 IHC clones and LDTs 58
2.2.0.1 PD-L1 protein abundance and the use of TMAs to assess PD-L1 concordance 60
2.2.0.2 Large studies comparing multiple PD-L1 clones 64
2.2.0.3 Small studies comparing multiple PD-L1 clones 68
2.2.0.4 PD-L1 clone concordance in other studies 69
2.2.1 Analytics- limitations of immunochemistry 71
2.3 Heterogeneity of PD-L1 expression 73
2.3.0 PD-L1 Heterogeneity – intra-tumoural heterogeneity 73

2.3.0.0 Quantitative immunofluorescence to assess PD-L1 intra-tumoural heterogeneity ... 75
2.3.0.1 Using TMAs to assess PD-L1 intra-tumoural heterogeneity ... 76
2.3.0.2 Small biopsies to assess PD-L1 intra-tumoural heterogeneity ... 79
2.3.0.3 Whole sections to assess PD-L1 intra-tumoural heterogeneity ... 82
2.3.1 PD-L1 Heterogeneity– Inter-tumoural and temporal heterogeneity ... 83
2.3.1.0 PD-L1 Inter-tumoural heterogeneity between primary lung and regional lymph node metastases ... 83
2.3.1.1 PD-L1 Inter-tumoural heterogeneity between primary lung and distant metastases 86
2.3.2 Heterogeneity – iatrogenic heterogeneity ... 90
2.4 Literature review conclusion ... 94
Chapter 3 - Materials and methods ... 96
3.0 Materials ... 96
3.0.0 General Materials and Equipment ... 96
3.0.1 Reagents and antibodies ... 96
3.0.2 Software ... 97
3.1 Patient and specimen cohorts ... 97
3.1.0 Liverpool Lung Project (LLP Cohort) ... 97
3.1.1 Clatterbridge Cancer Centre (CCC Cohort) ... 98
3.1.2 Other cohorts ... 99
3.2 Methods ... 100
3.2.0 Microtomy ... 100
3.2.1 Haematoxylin and eosin staining ... 100
3.2.2 PD-L1 SP263 immunohistochemistry ... 101
3.2.3 Digital slide scanning and viewing ... 101
3.2.4 Scoring PD-L1 expression ... 101
3.3 Statistics ... 104
Chapter 4 – The effect of pre-analytical conditions on PD-L1 expression ... 106
4.0 Introduction ... 106
4.0.1 Sampling methods ... 106
4.0.2 Tissue fixation ... 107
4.0.3 Specimen age and storage ... 108

4.1 Methods 110

4.1.0 Review of retrospective cases 110

4.1.1 Use of alcohol or formaldehyde based fixatives 110

4.1.2 Environmental factors 111

4.1.3 Statistics 113

4.2 Results 114

4.2.0 Retrospective case review of PD-L1 expression by sampling method 114

4.2.1 Impact of different fixatives on PD-L1 expression in paired cytology specimens 118

4.2.2 PD-L1 loss in normal conditions 120

4.2.3 PD-L1 loss in accelerated conditions 122

4.2.4 Effect of humidity and temperature on immunoreactivity loss by IHC for PD-L1 and pan-CK 124

4.2.5 Effect of oxygen on PD-L1 expression 126

4.2.6 The impact of desiccant on preventing immunoreactivity loss 128

4.2.7 PD-L1 immunoreactivity loss on specific cell types 131

4.3 Discussion 132

4.3.0 Impact of sampling methods on PD-L1 expression 132

4.3.1 Impact of fixatives on PD-L1 expression 133

4.3.2 Impact of storage conditions on PD-L1 expression 134

4.3.3 Limitations 136

Chapter 5 – Analytics and post-analytics: Digital pathology and PD-L1 interpretation 138

5.0 Introduction 138

5.0.0 Analytics 138

5.0.1 Post-analytics: Interpretation of PD-L1 expression 139

5.0.2 Digital pathology and image analysis 140

5.0.3 Inter-observer concordance of PD-L1 interpretation 140

5.1 Methods 142

5.1.0 Comparing PD-L1 22C3 and SP263 IHC antibody clones 142

5.1.1 Comparing PD-L1 22C3, 28-8, E1L3N and SP142 IHC antibody clones 142

5.1.2 uPath and use of the PD-L1 IHC interpretation algorithm 142

5.1.3 RPA in the LLP cohort 145

5.1.4 Training to use the RPA 146
5.1.5 RLUH Cohort – Intra-observer concordance 146
5.1.6 RLUH Cohort - Inter-observer concordance 146
5.1.7 Eli Lilly Cohort – Inter-observer concordance 147
5.1.8 Statistics 147
5.2 Results 147
5.2.0 22C3 vs SP263 147
5.2.3 Qualitative data from the LLP cohort 153
5.2.4 Intra-observer concordance for scoring PD-L1 156
5.2.8 Qualitative feedback from the RLUH cohort. 159
5.2.9 Training webinar to introduce new users to the RPA 164
5.3 Discussion 166
5.3.0 Variation of PD-L1 IHC clones 166
5.3.1 Digital pathology and image analysis for scoring PD-L1 167
5.3.2 Training to use digital image analysis 168
Chapter 6 – Heterogeneity of PD-L1 expression 170
6.0 Introduction 170
6.0.0 Tumour Heterogeneity 170
6.0.1 Heterogeneity of PD-L1 expression 171
6.0.2 Intra-tumoural and inter-tumoural heterogeneity 172
6.0.3 Temporal and iatrogenic heterogeneity 173
6.1 Methods 174
6.1.0 Selection of cohort 174
6.1.1 Intra-tumoural heterogeneity 176
6.1.2 Digital core biopsies 179
6.1.3 Digital fine needle aspirates 182
6.1.4 Inter-tumoural heterogeneity 183
6.1.5 Heterogeneity and pathological/clinical features 184
6.1.6 Statistics 184
6.2 Results 185
6.2.0 PD-L1 intra-tumoural heterogeneity – descriptives 185

6.2.1 Primary NSCLC - Small-scale heterogeneity and digital core biopsies 185
6.2.2 Primary NSCLC - Medium-scale heterogeneity 191
6.2.3 Primary NSCLC - Large-scale heterogeneity 193
6.2.4 Lymph node (secondary) metastatic deposits – intra-tumoural heterogeneity 194
6.2.5 Inter-tumoural heterogeneity – Primary vs metastases 197
6.2.6 Inter-tumoural heterogeneity – Metastases vs metastases 201
6.2.7 Heterogeneity and pathological/clinical features 201
6.3 Discussion 204
6.3.0 Extensive sampling and PD-L1 heterogeneity 204
6.3.1 Inter-tumoural heterogeneity: hinderer and helper 206
6.3.2 Novel approaches to quantifying heterogeneity 207
Chapter 7 – The tumour microenvironment 209
7.0 Introduction 209
7.0.1 PD-L1 and the TME 209
7.0.2 Types of immune TME 210
7.0.3 T-cells within the TME 211
7.0.4 Defining the TME 214
7.0.5 Multiplex technologies 215
7.1 Methods 216
7.1.0 Selection of cohort 216
7.1.1 TME categorisation by H&E 218
7.1.2 Multiplex immunohistochemistry 218
7.1.3 Multiplex Immunofluorescence (mIF) 221
7.1.4 Validation of image analysis and multiplex IHC approach 223
7.1.5 Association of TME features with PD-L1 expression heterogeneity 225
7.1.6 Classification of TMEs 225
7.1.7 Validation of Multiplex immunofluorescence 230
7.1.8 Statistics 231
7.2 Results 231
7.2.0 Multiplex IHC Data 231
7.2.1 Comparison of image analysis PD-L1 TPS and manual TPS 231

7.2.3 Associations of immune markers with PD-L1 expression heterogeneity using multivariate analysis modelling 233
7.2.4 Categorisation of TILs by H&E 234
7.2.5 Classification of TMEs by multiplex IHC TIL data 235
7.2.6 Classification of TME immune status by multiplex IHC immune markers 236
7.2.7 Classification of PD-L1 expression by multiplex IHC 238
7.2.8 Classification by potential response to PD-1/PD-L1 IMs by multiplex IHC 239
7.2.9 Predicted survival by TME features 241
7.2.10 MIF for PD-L1 assessment 242
7.2.11 mIF compared to multiplex IHC for immune cell markers 243
7.2.12 Classification of TMEs by mIF 248
7.3 Discussion 249
7.3.0 Immune TMEs and their relationship to PD-L1 expression heterogeneity 250
7.3.1 TMEs and their clinical relevance 251
7.3.2 Assessment of multiplex and digital image analysis 252
7.3.3 Other findings 254
Chapter 8 – Predicting response to PD-1/PD-L1 immunomodulatory therapy 256
8.0 Introduction 256
8.0.0. PD-L1 cut-offs for predicting response and toxicity 256
8.0.1 Measuring clinical response 257
8.1 Methods 258
8.1.0 Analysis of CCC clinical outcomes 258
8.1.1 Analysis of survival data in the CCC cohort 259
8.1.3 Pathological analysis of CCC cohort 260
8.1.4 Statistics 265
8.2 Results 265
8.2.0 Demographic and pathological details of the CCC cohort 265
8.2.1 Assessment of PFS and OS in CCC cohort by conventional cut-offs 265
8.2.2 – Exploratory analysis of PD-L1 expression for OS and PFS 270
8.2.3 Relationship of irAEs, PD-L1 expression and treatment response 275
8.2.4 Relationship of ECOG-PS and IM cycles, PD-L1 expression and treatment response…277

8.2.5 Univariate and multivariate linear regression modelling 279
8.2.6 Pathological assessment of CCC cohort 281
8.3 Discussion 283
8.3.0 PD-L1 positive – What cut-off? 283
8.3.1 PD-L1 TPS as a marker of reactive or constitutive expression 285
8.3.2 Using the TME to predict response 286
8.3.3 Predicting toxicity from PD-L1 expression 287
8.3.4 Predicting toxicity from the TME 289
8.3.6 Limitations 289
8.3.5 Conclusion 290
Chapter 9 – The Next Frontier: Improving the prediction of response to PD-1/PD-L1 IMs 291
9.1 Novel use of established biomarkers: PD-L1 expression in the pleura 291
9.1.0 PD-L1 in pleura and pleural fluid as a function of immune escape 291
9.1.1 Specific immune escape functions for specific tissue invasion 294
9.2 Immune TME signatures to predict response to IMs 296
9.3 Predictive power of image analysis 298
9.3.0 Image analysis for different PD-L1 clones 300
9.4 Mass-spectrometry (MS) and the proteome 301
9.5 Circulating biomarkers to predict response 302
9.5.0 Circulating tumour cells and PD-L1 expression 302
9.5.1 Exosomal and soluble PD-L1 303
9.3.2 Other circulating biomarkers 303
9.6 Predicting response in the near future 304
Chapter 10 - Discussion 309
10.0 Strengths and weaknesses of PD-L1 expression by IHC 309
10.0.1 The predictive power of PD-L1 expression 310
10.1 Alternative biomarkers 310
10.1.0 Alternative biomarkers: alternative methods of measuring PD-L1 311
10.1.1 Alternative biomarkers: lessons from single driver mutations 311
10.2 Measuring host immunity 313

10.3 Limitations 314

10.4 Conclusion 315

References 316-358

Appendix 1 359

Appendix 2 360

Acknowledgements

'*Nanos gigantum humeris insidentes*'

Bernard of Chartres

It is with sentiment in mind I make the following acknowledgments, as I could not have completed this thesis, nor enjoyed half the success, without the support of so many people.

To everyone at the LCL: Sharon and Steve, Tom, Naomi, Piya and everyone in the immuno-lab, your training, support and help were invaluable. To everyone from the University of Liverpool and the North-West MRC scheme: Angela and Maria, Professor Pirmohammed, Professor Coupland and Lakkis. This work could not have happened without you all!

My secondment in the USA with Eli Lilly & Co was an excellent period of time, a hugely rewarding experience and productive too. Thanks to Crystal, Dimple, Michael, Andrew, Bradley and all the team at the advanced diagnostic labs. A special thanks to Jeff for not only supporting my work but taking me to basketball games and generally helping me to experience the fun side of Indianapolis and not get too homesick!

To the team at Definiens, many thanks for ongoing and past work, and for inviting me to München to see your HQ – a valuable time and I hope we can produce more work moving forward. A big thank-you to my sponsors, the MRC, UKRI and industrial partners as part of the NW-MRC scheme. I have also been blessed to have the acknowledgement and occasional prize for work presented at various society meetings, and without the support of those from BTOG, PathSoc, APP, PPS, the IASLC and others, as well as the conferences and meetings held by you all, academia would be a far lonelier place.

To all my co-authors on so many projects, past and present (and hopefully in the future too), I cannot name you all and suspect I have missed some, but you know who you are and my thanks is sincere and heartfelt.

Thanks to Ana, who was not only invaluable in helping to collate the CCC cohort data and combine our oncologist and pathologist expertise, but also for giving me advice on the best places to eat in your home city of Barcelona during ESMO 2019.

Many thanks to Natalie, your patience and support in guiding my technical skills in the lab, your ongoing contributions to several pieces of research, your constantly happy and energetic approach to work and life, your interest in all things lung (plus of course your help with the horrendously large

number of FFPE sections and my ‘creative’ sectioning of FFPE blocks) have been invaluable and hugely valued!

A special thanks must go to my most excellent supervisory team. John (Professor Field) your advice support and guidance on a huge range of areas has been invaluable to both my thesis and my development as academic clinician. I am honoured to have played a role in the LLP story. Carles your oncology expertise, enthusiasm and support for clinical-academic work has been invaluable, and I know we will have much to achieve in the future with IMPULSE! Aaron, many thanks for your advice, your frequent calling in from the USA for meetings (even on occasion at 4am!) as well as meeting for a drink in Toronto, showing me around Indianapolis and everything else. Hopefully I will see you in the USA again soon! Mike, you have been a constant guide through the world of academia, industry, lab work, paper writing, legal teams and no end of other new areas, skills and knowledge I have learnt in the past 3 years. It is no understatement to say things would not have been anywhere near as successful without your constant support!

Finally a deep and profound thank-you is reserved for Professor John Gosney. Words cannot express the support you have provided: pathological advice, support with academia, relationships with industry, introductions to esteemed colleagues, collaborations and of course the enjoyment of the social side of clinical and academic work. Without your mentoring, supervision and friendship I would not have made it to this point.

List of abbreviations

ABC	Avidin biotin complex
ADC	Adenocarcinoma
ADSC	Adenosquamous carcinoma
AFA	Alcohol Formalin Acetic acid
AJCC	American Joint Committee on Cancer
Akt (PKB)	Protein Kinase B
ALK	Anaplastic lymphoma kinase
ANOVA	Analysis of variance
AP-1	Activator protein 1
APC	Antigen presenting cell
AQUA	Automated quantitative analysis
AR	Antigen retrieval
ASCO	American Society of Clinical Oncology
ATP	Adenosine triphosphate
BMS	Biomedical Scientist
CAF	Cancer associated fibroblast
CARD9	Caspase recruitment domain-containing protein 9
CAR-T	Chimeric antigen receptor T cells
CC	Cell conditioning
CCC	Clatterbridge Cancer Centre
CD (X)	Cluster of differentiation X (e.g. CD4)
circRNA	Circulating RNA
CK	Cytokeratin
COD	Cause of death
COV	Coefficient of variation
CPS	Combined positive score
CR	Complete response
CSPY	Combined smoking pack years
CT	Computed Tomography
CTCB	Cell-transfer cell blocks
CTCs	Circulating tumour cells
ctDNA	Circulating DNA
CTLA4	Cytotoxic T-lymphocyte-associated protein 4
CXCL	Chemokine ligand
DAB	3, 3'-diaminobenzidine
DC	Dendritic Cell
DCB	Digital core biopsy
DCR	Disease control rate
DDPCR	Digital Droplet Polymerase chain reaction
DFNA	Digital fine needle aspirate
DNA	Deoxyribose nucleic acid
EBUS	Endobronchial ultrasound
ECM	Extracellular matrix

ECOG	Eastern Cooperative Oncology Group
EDTA	Ethylenediaminetetraacetic acid
EGFR	Epidermal growth factor receptor
ELISA	Enzyme linked immunosorbent assay
ER	Oestrogen Receptor
ERK	Extracellular signal-regulated kinase
EUS	Eosophageal ultrasound
FA	Formic acid
FDA	Food and Drug Administration (United States of America)
FFPE	Formalin fixed paraffin embedded
FGF	Fibroblast growth factor
FISH	Fluorescent in situ hybridisation
FNA	Fine needle aspirate
FoxP3	Forkhead box P3
GM-CSF	Granulocyte-macrophage colony-stimulating factor
H&E	Haematoxylin & eosin
H202	Hydrogen peroxide
HCC	Hepatocellular carcinoma
HER2	Human epidermal growth factor receptor 2
HIFs	Hypoxia-inducible factors
HIV	Human immunovirus
HMGB1	High Mobility Group Box 1
HNSCC	Head & Neck Squamous cell carcinoma
HRP	Horseradish peroxidase
IASLC	International association for the study of lung cancer
IC	Immune cell
ICC	Intra-class correlation
ICPS	Immune cell proportion score
IDO1	Indoleamine 2, 3-dioxygenase 1
IF	Immunofluorescence
IFN-γ	Interferon gamma
IHC	Immunohistochemistry
IL	Interleukin
IM	Immunomodulatory therapy
INS	Insufficient
IO	Immuno-oncology
irAE	Immune related adverse event
ISET	Isolation by size of tumour cells
ISH	In situ hybridisation
ITSM	Immunoreceptors tyrosine-based switch motif
IVD	In vitro diagnostic
JAK	Janus kinase
LAG3	Lymphocyte-activation gene 3
LCL	Liverpool clinical laboratories

LC-MS	Liquid chromatography-tandem mass spectrometry
LCS	Liquid Coverslip
LDH	Lactate dehydrogenase
LDT	Laboratory developed test
LIPI	lung immune prognostic index
LLL	Left lower lobe
LLP	Liverpool Lung Project
LN	Lymph node
LUL	Left upper lobe
MAb	Monoclonal antibody
MAPK	Mitogen-activated protein kinase
MDSC	Myeloid derived suppressor cell
MDT	Multidisciplinary team
MHC	Major histocompatibility complex
mIF	Multiplex immunofluorescence
miRNA	Micro RNA
MMP	Matrix metalloproteinases
mRNA	Messenger RNA
MS	Mass spectrometry
MS-Cohort	Merseyside-Salford Cohort
MSI	Microsatellite instability
mTOR	Mechanistic target of rapamycin
NBF	Neutral buffered formalin
NE	Non-epithelium
NFAT	Nuclear factor of activated T-cells
NF-KB	Nuclear factor-κB
NGS	Next generation sequencing
NICE	National Institute for Health and Care Excellence
NK	Natural killer cell
NOS	Not otherwise specified
NPIC	Northern Pathology Imaging Consortium
NPV	Negative Predictive Value
NSCLC	Non-small cell lung carcinoma
OPA	Overall predictive accuracy
ORR	Overall response rate
OS	Overall survival
PCNA	Percutaneous needle aspiration
PCR	Polymerase chain reaction
PD	Progressive disease
PD-1	Programmed cell death protein 1
PD-L1	Programmed death ligand 1
PD-L2	Programmed death ligand 2
PFS	Progression free survival
PI3K	Phosphoinositide 3-kinases

PIP2	Phosphatidylinositol (4,5)-bisphosphate
PIP3	phosphatidylinositol (3,4,5)-trisphosphate
PKC	Protein Kinase C
PPV	Positive predictive value
PR	Progesterone receptor
PR	Partial Response
PTEN	Phosphatase and tensin homolog
PTM	Post-translational modification
QIF	Quantitative immunofluorescence
RBT-PD-L1	Rabbit monoclonal antibody for PD-L1
RECIST	Response evaluation criteria in solid tumours
RLL	Right lower lobe
RLUH	Royal Liverpool University Hospital
RML	Right middle lobe
RNA	Ribose nucleic acid
ROI	Regions of interest
ROS	Reactive oxygen species
RPA	Roche-PD-L1-Algorithm
RPMI	Roswell Park Memorial Institute
RUL	Right upper lobe
SCC	Squamous cell carcinoma
SCLC	Small cell lung cancer
SD	Stable disease
SHP	Src homology 2 domain-containing protein tyrosine phosphatase
SNOMED	Systematised Nomenclature of Medicine Clinical Terms
sPD-L1	Soluble PD-L1
SPSS	Statistical Product and Service Solutions.
SSC	Sodium Chloride Sodium Citrate
STAT	Signal transducer and activator of transcription
TAM	Tumour associated macrophage
TAN	Tumour associated neutrophil
TBNA	Transbronchial needle aspirate
TC	Tumour cell
TCR	T-Cell Receptor
TD-MS	Top-down MS
TE	Tumour epithelium
TFG	Trafficking From ER To Golgi Regulator
TGF-β	Transforming growth factor beta
Th	T-helper cell
TILs	Tumour infiltrating lymphocytes
TIM3	T cell immunoglobulin and mucin domain-containing protein 3
TKI	Tyrosine kinase inhibitor
TMA	Tissue microarray
TME	Tumour microenvrionment

TNF-α	Tumour necrosis factor alpha
TNM	Tumour, nodes, metastases (staging)
Treg	T-regulatory cell
TTF-1	Thyroid transcription factor 1
UC	Urothelial carcinoma
UICC	Union for International Cancer Control
UVA	Ultraviolet A
VEGF-A	Vascular endothelial growth factor A
VISTA	V-domain immunoglobulin suppressor of T cell activation
WHO	World Health Organisation
XCT	Chemotherapy
XRT	Radiotherapy
YWHAE	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein epsilon
ZAP70	Zeta-chain-associated protein kinase 70
ZEB1	Zinc Finger E-Box Binding Homeobox 1

List of tables

Table 1.6.0 PD-L1 IHC clone, scoring approach and 'cut-off' used in various cancer types 30

Table 2.1.0 Results for articles by search term 37

Table 2.1.1 Papers with cytology and histology specimens comparing expression of PD-L1. 54

Table 2.3.0 Papers studying the effect of PD-L1 expression from samples pre- and post-treatment. 92

Table 3.3.0 Statistical tests used 105

Table 4.2.0 PD-L1 expression by SP263 in different NSCLC specimens by sampling methodology .. 115

Table 4.2.1 PD-L1 expression by SP263 in different NSCLC specimens by anatomical location 115

Table 4.2.2 PD-L1 expression by SP263 in 60 paired specimens of EBUS FNAs 119

Table 5.2.0 PD-L1 expression across 1,400 NSCLC specimens (clinical categories)..... 148

Table 5.2.1 PD-L1 expression across 1,400 NSCLC specimens (anatomical location/sampling) 149

Table 5.2.2 PD-L1 expression by 22C3, 28-8, E1L3N or SP142 149

Table 5.2.3 Correlation and scoring agreement of assessing PD-L1 manually and digitally. 151

Table 5.2.4 Predictive power of RPA 153

Table 5.2.5 Intra-pathologist correlation and scoring agreement (manual) 156

Table 5.2.6 Intra-pathologist correlation and scoring agreement (digital) 156

Table 5.2.7 Inter-pathologist correlation and scoring agreement (RPA) 156

Table 5.2.8 Inter-observer correlation and scoring agreement (Positive pixel count) 157

Table 5.2.9 Correlation and scoring agreement in whole sections 158

Table 5.2.10 Predictive power of using the RPA applied to whole sections and annotations 158

Table 5.2.11 Specimen types in the RLUH cohort assessed for PD-L1 expression. 158

Table 5.2.12 – Advice to a new user of uPath 165

Table 6.1.0 Pathological and clinical characteristics of the LLP cohort 175-176

Table 6.2.0 Small-scale heterogeneity of PD-L1 expression 189-190

Table 6.2.1 Summary of the accuracy of DCBs and DFNs (primary) 190

Table 6.2.2 Medium-scale heterogeneity of PD-L1 expression 192

Table 6.2.3 Large-scale heterogeneity of PD-L1 expression 193-194

Table 6.2.4 Intra-tumoural heterogeneity of PD-L1 expression in 30 lymph 195-196

Table 6.2.5 Summary of the accuracy of DCBs and DFNAs (metastases) 196

Table 6.2.6 Inter-tumoural heterogeneity of PD-L1 expression 198-200

Table 6.2.7 Inter-tumoural heterogeneity of PD-L1 expression (metastases) 202

Table 7.1.0 92 cases of NSCLC assessed by multiplex IHC 216-218

Table 7.1.1. The 33 metrics provided by the multiplex IHC images. 221

Table 7.1.2 The 21 metrics provided by the mIF images 222

Table 7.2.0 Correlations and predictive values of image analysis 232

Table 7.2.1. Correlations and predictive values of image analysis of multiplex IHC for 232

Table 7.2.2 PD-L1 expression in NSCLCs by different TMEs (TILs) 234

Table 7.2.3 Comparison of different immune TMEs (TILs) 235

Table 7.2.4 PD-L1 expression in NSCLCs by different TMEs (multiplex IHC) 235

Table 7.2.5. Comparison of different immune TMEs (multiplex IHC) 236

Table 7.2.6 PD-L1 expression in NSCLCs by different TMEs (immune cell markers) 237

Table 7.2.7 Comparison of different immune TMEs (immune cell markers) 238

Table 7.2.8 PD-L1 expression in NSCLCs by different TMEs (IM response prediction) 239

Table 7.2.9 Comparison of different immune TMEs (IM response prediction) 239

Table 7.2.10 NSCLC tumours divided by predicted response to IM therapy 240

Table 7.2.11 Final categories for potential response to PD-1/PD-L1 IMs 240

Table 7.2.12 Comparison of different groups of NSCLCs (IM response) for heterogeneity 240

Table 7.2.13 PD-L1 expression by mIF compared to PD-L1 expression by IHC 243

Table 7.2.14 Correlations and predictive values of image analysis of mIF 245

Table 7.2.15 Average densities for immune cell markers by multiplex IHC or mIF. 245

Table 7.2.16 PD-L1 expression NSCLCs divided into TMEs by mIF. 248

Table 7.2.17 Comparison of different groups of NSCLCs by mIF for heterogeneity 249

Table 8.1.0 Definitions of outcomes and events 259

Table 8.2.0 Demographic and pathological details of the the CCC cohort. 266

Table 8.2.1 2nd-line treated patients divided into 10% incremental groups to compare irAEs 277

Table 8.2.2 2nd-line treated patients divided into 10% dichotomous division to compare irAEs 278

Table 8.2.3 NSCLC cases from the CCC cohort with specimens available for PD-L1 analysis and immune TME subtyping. 282

List of figures

Fig 1.0.0 Oncogenic drivers in NSCLC **3**

Fig 1.0.1 Principles behind immune-checkpoint inhibition. 6

Fig 1.3.0 Immune checkpoint receptors and their ligands 11

Fig 1.4.0 PD-1/PD-L1 axis intra-cellular signalling 19

Fig 1.5.0 TMB in solid tumours 24

Fig 1.6.0 A NSCLC immunolabelled for PD-L1 (SP263) 28

Fig 1.6.1 Survival analysis curves from two seminal PD-1/PD-L1 IM clinical trials 33

Fig 1.6.2 Survival analysis curves from three seminal PD-1/PD-L1 IM clinical trials 34

Fig 2.1.0 Flow chart of the literature review 36

Fig 2.1.1 IHC staining loss in aged tissue 46

Fig 2.2.0 Four studies that compared different PD-L1 IHC clones in NSCLC. 70

Fig 2.3.0 Heterogenous PD-L1 expression in NSCLC 74

Fig 2.3.1 Intra-tumoural PD-L1 expression heterogeneity 76

Fig 3.2.0 Assessment of PD-L1 expression 102

Fig 4.2.0 High PD-L1 expression in four different specimen types 116

Fig 4.2.1 PD-L1 TPS average for lung adenocarcinomas in different specimens. 116

Fig 4.2.2 H&E and PD-L1 expression for pleural and pericardial specimens 117

Fig 4.2.3 Expression of PD-L1 in two pairs of EBUS FNAs 120

Fig 4.2.4 PD-L1 expression in aged tissue and tissue under accelerated conditions 121

Fig 4.2.5 PD-L1 expression in gastric cancer TMAs over 24 months (positivity index) 121

Fig 4.2.5 PD-L1 expression in gastric cancer TMAs over 24 months (CPS) 122

Fig 4.2.7 PD-L1 expression by positive pixel count in NSCLC sections over time 123

Fig 4.2.8 PD-L1 expression by clinical cut-offs in NSCLC sections over time 124

Fig 4.2.9 PD-L1 expression in placenta and tonsil FFPE sections in the acceleration chamber 126

Fig 4.2.10 PD-L1 expression in placenta and tonsil FFPE sections in the acceleration chamber 126

Fig 4.2.11 PD-L1 protein by MS 128

Fig 4.2.12 PD-L1 and pan-CK IHC loss in tonsil and placenta in the acceleration chamber 130

Fig 4.2.13 PD-L1 expression by positive pixel count in placenta and tonsil 131

Fig 4.2.14 Tonsil tissue stained for PD-L1 in sections stored in the acceleration chamber 132

Fig 5.0.1 Challenges in assessing PD-L1 expression to generate a TPS 141

Fig 5.1.0 DP200 user interface 144
Fig 5.1.1 uPath view of NSCLC stained for PD-L1 and H&E. 144
Fig 5.1.2 Applying annotations to NSCLC tissue in uPath. 145
Fig 5.2.0 Examples of PD-L1 expression by SP263 and 22C3 148
Fig 5.2.1 PD-L1 expression by 22C3, 28-8, E1L3N or SP142 150
Fig 5.2.2 Examples of PD-L1 expression by 4 clones in a matched NSCLC specimens 151
Fig 5.2.3 Linear regression for scoring digitally and manually 152
Fig 5.2.4 False scoring of PD-L1 by the RPA 154
Fig 5.2.5 Necrosis and PD-L1 positive TILs in NSCLC sections. 155
Fig 5.2.6 NSCLC tumour stained for PD-L1 with TILs and necrosis 155
Fig 5.2.7 Case assessed for PD-L1 (SP263) using uPath to draw annotations for RPA assistance. 157
Fig 5.2.8 Linear regression for RPA and manual scoring 159
Fig 5.2.9 Samples not suitable for RPA 160
Fig 5.2.10 Samples not requiring RPA 161
Fig 5.2.11 Example of when accurate application of RPA is excessively time consuming 162
Fig 5.2.12 Example of when application of RPA adds benefit. 163
Fig 5.2.13 Benefits of digital pathology. 164
Fig 6.0.1 Intra-tumoural heterogeneity of PD-L1 expression 173
Fig 6.1.0 Creation of a labelled 1cm² area in Qupath 177
Fig 6.1.1 Data visualisation 178
Fig 6.1.2 Medium-scale heterogeneity 179
Fig 6.1.3 Example of a good quality core biopsy of a NSCLC 180
Fig 6.1.4 Simulation of DCBs in primary NSCLC 180
Fig 6.1.5 DCBs data collection. 182
Fig 6.1.6 Simulation of DCBs in a lymph node 183
Fig 6.2.0 Example of small-scale heterogeneity 186
Fig 6.2.1 Patterns of PD-L1 expression 186
Fig 6.2.2 Patterns of PD-L1 expression heterogeneity. 187
Fig 6.2.3 Infiltrating edge of tumour preferentially expressing PD-L1 188
Fig 6.2.4 PD-L1 (SP263) heterogeneity in NSCLC metastatic deposits 195
Fig 6.2.5 Challenges of focal PD-L1 expression in metastatic NSCLC 197

Fig 6.2.6 Examples of dramatic inter-tumoural heterogeneity of PD-L1 expression 200
Fig 6.2.7 Examples of dramatic inter-tumoural heterogeneity of PD-L1 expression (metastases).... 201
Fig 6.2.8 Examples of dramatic intra-tumoural heterogeneity of PD-L1 expression 203
Fig 7.0.0. Somatic mutation frequencies 210
Fig 7.0.1. Different TMEs by the pattern of TILs 212
Fig 7.0.2 H&E of NSCLC with a difficult to classify TME 213
Fig 7.1.0 Low power view of NSCLC tumour with panel of stains performed for multiplex IHC..... 220
Fig 7.1.1 mIF with 5 markers..... 222
Fig 7.1.2. Multiplex IHC with 3 markers..... 223
Fig 7.1.3 Tumour cells and macrophages positive for PD-L1 224
Fig 7.1.4 TME classification by TIL distribution 226
Fig 7.1.5 TME classification with TILs and Granzyme B 227
Fig 7.1.6 TME classification by PD-L1 expression 228
Fig 7.1.7 TME classification to predict response to PD-1/PD-L1 IM therapy..... 229
Fig 7.1.8 Matched images for multiplex IHC and miF 230
Fig 7.2.0. Matched sections of NSCLC tumour stained for PD-L1 by monoplex and triplex assays. .. 233
Fig 7.2.1 Survival curve for NSCLC patients by TME (TILs)..... 241
Fig 7.2.2 Survival curve for PD-L1/CD8+ve High vs PD-L1/CD8+ve Low 242
Fig 7.2.3. Multiplex IHC and mIF for immune markers. 244
Fig 7.2.4. Under-expression of PD-L1 by mIF..... 246
Fig 7.2.5. Over-expression of PD-L1 by mIF. 247
Fig 7.2.6. Aberrant PD-L1 expression in necrotic tissue 248
Fig 8.1.0 Examples of NSCLC biopsies for which immune TMEs can be accurately depicted 261
Fig 8.1.1 Examples of NSCLC biopsies with minimal immune cell presence 262
Fig 8.1.2 Examples of pleural biopsy with striking example of immune exclusion 263
Fig 8.1.3 Example of pleural biopsy samples with infiltrating NSCLC metastases..... 264
Fig 8.1.4 Admixed tumour and immune cells from pleural fluid sample 264
Fig 8.2.0 Survival curve (OS) for patients with tumours having strong or weak PD-L1 expression.... 266
Fig 8.2.1 Survival curve (PFS) for patients with tumours having strong or weak PD-L1 expression .. 267
Fig 8.2.2 Survival curve (OS) for patients with tumours positive (≥25% TPS) or negative for PD-L1 . 267
Fig 8.2.3 Survival curve (PFS) for patients with tumours positive (≥25% TPS) or negative for PD-L1 268

Fig 8.2.4 Survival curve (PFS) for 2nd-line plus patients with PD-L1 positive (≥25% TPS) tumours) ... 269

Fig 8.2.5 OS and HR for patients divided by PD-L1 expression into 10% incremental groups 270

Fig 8.2.6 PFS and HR for patients divided by PD-L1 expression into 10% incremental groups 271

Fig 8.2.7 OS and HR for patients divided by PD-L1 expression by 10% dichotomous division 272

Fig 8.2.8 PFS and HR for patients divided by PD-L1 expression by 10% dichotomous division 273

Fig 8.2.9 Survival curve (OS) for 2nd-line treated patients divided into three PD-L1 categories 274

Fig 8.2.10 Survival curve (OS) for 2nd-line treated patients divided into three PD-L1 categories 275

Fig 8.2.11 Survival curve (OS) for 1st-line treated patients divided into three PD-L1 categories 276

Fig 8.2.12 Survival curve (OS) for 1st-line treated patients divided into three PD-L1 categories 276

Fig 8.2.13 Multivariate analysis predicted values as a function of predicting PFS.....280

Fig 9.1.0 Flow chart of 4 areas for developing future projects..... 292

Fig 9.1.1 Different areas for sampling NSCLC specimens 306

Publications

The following peer-reviewed original publications have been generated as a result of work presented in this thesis:

Haragan A, Liebler DC, *et al*. Accelerated instability testing reveals quantitative mass spectrometry overcomes specimen storage limitations associated with PD-L1 immunohistochemistry. Lab Invest 2020;100:874-886.

Haragan A, Field JK, Davies MPA, Escriu C, Gruver A, Gosney JR. Heterogeneity of PD-L1 expression in non-small cell lung cancer: Implications for specimen sampling in predicting treatment response. Lung Cancer 2019;134:79-84.

Haragan A, Gosney J. Immunohistochemistry for prediction of response to immunotherapy. Diagnostic Histopathology, Diagnostic Histopathology, 2020, Volume 26, Issue 11, 506 – 512.

Gosney JR, Haragan A, Chadwick C, et al. Programmed death ligand 1 expression in EBUS aspirates of non-small cell lung cancer: Is interpretation affected by type of fixation? Cancer Cytopathol 2020;128:100-106.

Liebler D, Haragan A *et al*. Analysis of Immune Checkpoint Drug Targets and Tumor Proteotypes in Non-Small Cell Lung Cancer Scientific Reports 10(1):9805 DOI: 10.1038/s41598-020-66902-0

Kapil A, Haragan A *et al*. Domain adaptation-based deep learning for automated Tumor Cell (TC) scoring and survival analysis on PD-L1 stained tissue images. IEEE transactions of medical imaging. 2021 May 18;PP. doi: 10.1109/TMI.2021.3081396.

Abstracts

Haragan A, Field JK, Davies MPA, Escriu C, Gruver A, Gosney JR PD-L1 Expression Heterogeneity in NSCLC: Accuracy and Reliability of Using Primary Lung Tumour Versus Metastatic Lymph Node Deposits JOURNAL OF PATHOLOGY Volume 248 Page S11-S11 Published 2019

Haragan A, Field JK, Davies MPA, Escriu C, Gruver A, Gosney JR Utilising Heterogeneity: Using a Digital Database of Lung Cancers and Immune Profile to Complement Subjective Assessment JOURNAL OF THORACIC ONCOLOGY Volume 13 (10) Pages: S768-S768 Supplement: S Meeting Abstract: P2.09-19DOI: 10.1016/j.jtho.2018.08.1316

Accelerated Antigen Instability Testing Reveals Quantitative Mass Spectrometry Analysis Overcomes Specimen Limitations Associated with Diagnostic PD-L1 Testing Haragan, AK, Libler D, et al. JOURNAL OF PATHOLOGY Volume: 250 Pages: 13-13 Supplement: 1 Published: MAR 2020

Haragan A, Field JK, Davies MPA, Escriu C, Gruver A, Gosney JR Digital Core Needle-Biopsy to Assess PD-L1 Expression in Non-Small Cell Lung Cancer: Optimal Sampling and Need for Re-Biopsy JOURNAL OF THORACIC ONCOLOGY Volume: 13 Issue: 10 Pages: S534-S535 Supplement: S Meeting Abstract: P1.04-24 DOI: 10.1016/j.jtho.2018.08.739

Haragan A, Elwin F, Gosney JR. High expression of PD-L1 by pulmonary adenocarcinoma metastatic to body cavities; artefact or biology? Lung Cancer, Volume 139, Supplement 1,2020,Pages S57-S58, ISSN 0169-5002, https://doi.org/10.1016/S0169-5002(20)30160-4.

Haragan, A. *et al.* PD-L1 expression in resected non-small cell lung cancer: variation between primary tumour and nodal metastases Lung Cancer, Volume 127, S25, 2019

Sastre AH; Haragan A et al. Looking beyond the assay: Comparison of multiplex chromogenic and fluorescent immunohistochemistry for standardized immune oncology profiling in non-small cell lung carcinoma patients JOURNAL FOR IMMUNOTHERAPY OF CANCER Volume: 7 Supplement: 1 Meeting Abstract: P39

Rognoni L, Haragan A et al. Combining the best of two worlds: Transfer of multiplex immunofluorescence images from non-small cell lung carcinoma patients into pseudo multiplex chromogenic immunohistochemistry images JOURNAL FOR IMMUNOTHERAPY OF CANCERVolume: 7 Supplement: 1 Meeting Abstract: P55

F. Elwin, A. Haragan, J. Gosney, 131 - PD-L1 expression in non-small cell lung cancer according to specimen type; a retrospective study of over 2,000 cases, Lung Cancer, Volume 139, Supplement 1, 2020, Page S57,ISSN 0169-5002, https://doi.org/10.1016/S0169-5002(20)30159-8.

Book Chapters

Gosney J, Haragan A, Laing G, Kerr K. Biomarkers for Immunotherapy. ESMO handbook of Immuno-Oncology 2018. Page Reference: 43-55 ISBN: 978-88-941795-7-6

Gosney J, Haragan A. Difficulties in Assessment of PD-L1 Expression. Problem Solving in Cancer Immunotherapy 2019. Chapter 6. ISBN: 978-09-955954-2-2

Chapter 1 - Introduction

1.0 Lung cancer and treatment options

Lung cancer is the third most common cancer globally, affecting over two million people each year. Lung cancer is the biggest cause of cancer death and kills around 1.8 million people annually, which is more than breast, prostate and colorectal cancers (the first, second and fourth most common cancers respectively) combined.[1] There are over 46,000 new cases of lung cancer and over 35,000 deaths from lung cancer in the UK each year which accounts for nearly a quarter of all UK cancer deaths. Despite these sobering statistics, UK survival has seen an improvement in 1 year survival rate from around 15% in the 1970s to around 35% today, with longer-term survival seeing improvement from around 3% in the 1970s to around 10% today.[2] The vast majority of this increase in survival has occurred in the last two decades since the dawn of stratified medicine and targeted therapeutics, including the use of tyrosine kinase inhibitors (TKIs) for specific genomic alteration in cancer cells, and, more recently, immuno-modulatory (IM) therapy.

Lung cancer is traditionally divided into histological subtypes, with the majority (80-85%) classified as non-small cell lung cancer (NSCLC), of which the most common subtype in the UK is adenocarcinoma (ADC), followed by squamous cell carcinoma (SCC) with a relatively small number of large cell carcinomas or not-otherwise-specified (NSCLC-NOS) depending on specimen type.[2, 3] The majority of successful targeted medicines in lung cancer have been for NSCLC, with few new successful treatments developed for small cell and neuroendocrine tumours, with a corresponding poor prognosis for these diseases.

1.0.0 Development of stratified treatments in advanced NSCLC

Since the 1980s the mainstay of treatment for advanced NSCLC has been platinum based chemotherapy (XCT), but it was only in 2008 that different chemotoxic agents were shown to be associated with variable survival data based on differentiating squamous from non-squamous NSCLC.[4, 5] Although there have been more recent advances in chemotoxic agents, XCT still carries with it significant toxicity and relatively poor overall response, and attempts to use molecular and genetic markers to guide XCT use in NSCLC has been met with minimal success.[6] In 2004, a Phase II clinical trial illustrated the first non-chemotoxic agent as having efficacy in NSCLC: bevacizumab, a monoclonal antibody (MAb) targeted against VEGF-A (vascular endothelial growth factor A) which inhibits angiogenesis within the tumour microenvironment (TME), was used alongside carboplatin (a platinum based chemotoxic agent) and shown to have favourable clinical outcomes.[7] However, patients with SCCs of the lung had increased toxicity and an unfavourable side-effect profile, and as

such the 2006 Phase III clinical trial that earned bevacizumab its license to be used in NSCLC excluded SCC patients,[8] as did other subsequent landmark trials.[9] Therefore the use of bevacizumab was not only a significant landmark in terms of using non-chemotoxic agent for treating advanced NSCLC, but also that it highlighted the importance of stratifying patients by tumour characteristics.

At a similar time the discovery of *EGFR* (epidermal growth factor receptor) overexpression as a major driver of NSCLC survival in a subset of patients resulted in the successful development and use of TKIs to inhibit this activity. Activated *EGFR* acts to increase cell proliferation and survivability via a number of mechanisms including the MAPK activated Ras/ERK (Extracellular signal-regulated kinase) pathways and the PI3K/Akt (Phosphoinositide 3-kinases/Protein Kinase B) pathway, with these becoming constitutively activated in the presence of mutant *EGFR*, and thus leading to significant oncogenic driven tumour cell survival.[10] TKIs targeted against *EGFR* mutants have seen considerable success in treating NSCLC, with the first *EGFR* TKI, gefitinib, receiving its license for use as second line treatment in 2003.[11] However, the two Phase II clinical trials that led to this approval did not stratify patients by the presence of *EGFR* mutants, and whilst these still showed positive results[12, 13] several Phase III clinical trials looking at *EGFR* TKIs in the first and second line setting that also did not stratify by *EGFR* mutant status failed to reach their outcomes.[14-16] Conversely, Phase III studies that did stratify patients by specific *EGFR* mutants showed considerable benefit when treated with *EGFR* TKIs[17, 18] leading to the widespread recommendations and subsequent requirement for *EGFR* mutant testing from 2009 onwards to stratify NSCLC patients for targeted treatment.[11, 19] However, three significant issues limit the effectiveness of *EGFR* TKIs. The first issue is the presence or absence of specific *EGFR* mutants: exon 19 deletions and the missense mutation L858R in exon 21 respond favourably, but other mutations, such as exon 20 mutations are minimally sensitive to TKIs, and other mutations are of no or unknown clinical relevance.[10, 20] The second is that almost all patients commenced on *EGFR* TKIs develop resistance to treatment within 2 years, with secondary mutations such as T790M and C797S, the activation of alternative intra-cellular signalling mechanisms, histological transformation and other processes implicated.[21] Whilst other treatment approaches for resistant tumours include 3rd generation *EGFR* TKIs (e.g. osimertinib and rociletinib) have seen success, tumours eventually develop resistance to these as well.[22] The third issue limiting *EGFR* TKI use is that sensitive *EGFR* mutants only account for around 10-20% of NSCLC patients in Europe, although in certain populations, particularly Asian populations, the prevalence is much higher, often over 50%. [23, 24] Other oncogenic driver mutations besides *EGFR* activation have been well characterised in NSCLC and include *ALK* gene rearrangements, *BRAF* mutants, *ROS1* gene rearrangements and several others (Fig 1.0.0) which are largely mutually exclusive. Although several of these oncogenic drivers have TKIs targeted against them (e.g *ALK* and *ROS1* mutants may be

targeted by lorlatinib) many others, included the commonest *KRAS* mutations, have no effective treatment at present.[25]

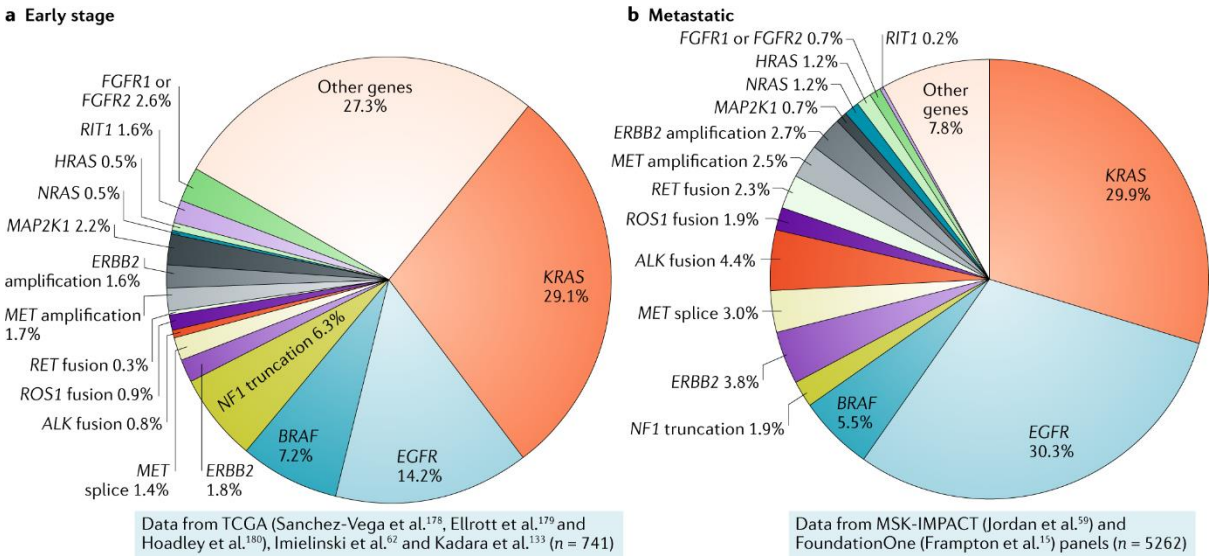


Figure 1.0.0 Oncogenic drivers in NSCLC. **A –** Early stage , **B** – Late stage. Targeted therapy against many of the drivers are approved or are in development and include *EGFR*, *ALK, ROS1, BRAF, RET, MET* and *NTRK*, though effective targeted treatment against many drivers remain elusive. Modified from Skoulidis *et al.* 2019.[25]

Therefore a new diagnosis of NSCLC should be accompanied with an interrogation of the genetic landscape of the tumour which looks for the oncogenic driver mutations that have approved TKIs as a minimum of pathological work-up. Many tumours will not be suitable candidates for TKI treatment, however, and most that are will develop resistance, so there remains a requirement for more treatment options in advanced NSCLC. More recently, the development of IM therapy has seen improvements in clinical outcomes for NSCLC patients whom have no targetable mutation, with long-term survival reported for many patients and occasional reports of exceptional responders surviving in excess of 5 years with metastatic disease,[26-28] making this an exciting prospect for further increasing the repertoire of stratified medicines for NSCLC patients.

1.0.1 History of cancer immunotherapy

The notion that the immune system can be used to successfully target cancer maybe erroneously thought of as a modern concept. Some fairly anecdotal but fascinating reports of tumours spontaneously regressing post-infection have been reported for some 3000 years since the ancient Egyptian physicians noted this phenomena.[29] Erysipelas, a particular form of cellulitis now known to

be caused predominantly by group A β-haemolytic streptococci (or *streptococcus pyogenes*) was noted to be associated with spontaneous regression of tumours by Busch in 1868, in what is the first known example of using an intentional infection to target cancer.[30] In 1882 Fehleisen, whom had independently noted the same phenomena, repeated the experiment and identified *streptococcus pyogenes* as the bacterial agent responsible for this.[31] William Coley noted similar observations in his cohort of sarcoma patients, and went as far a developing 'Coley's toxins' in which he developed various streptococci (the details on which, from his 1909 paper include the use of 1lb of beef used to make a 'bacterial broth') which was first tested on dogs and then humans with sarcoma. Coley reported tumour regression and cure in advanced sarcoma in many hundreds of patients.[32-34] However, unsurprisingly, many of Coley's patients died from sepsis, added to which the technical difficulty of growing the bacteria and the lack of supportive studies that replicated the results, this approach generally fell out of favour. Throughout the next 100 years opinions varied in the general consensus of whether the immune system could be a genuine factor in targeting tumours or not.[35] Furthermore, shortly after Coley's experiments, Alexander Fleming and his contemporaries developed benzylpenicllin, the first broad-spectrum antibiotic, and a better understanding of the role of bacteria and illness was understood, which undoubtedly contributed to the notion of using bacteria to target cancer as unpalatable.[36] In 1967, the first (and to date only) established use of bacteria to treat cancer was developed. Morales, Eidinger and Bruce utilised Bacillus Calmette-Guérin ((BCG), an avirulent form of tuberculosis causing bacteria) to treat non-muscle invasive bladder cancer,[37] which subsequent work has shown to be a result of immune activity targeted against the tumour cells, albeit only successfully in approximately 60% of patients.[38] Despite a few papers in the 1970s still questioning the role of immunity targeting cancer, a large number of papers in the 1980s and 1990s demonstrated beyond doubt that the immune system plays a significant anti-cancer role.[35]

The biggest success in immunotherapy for treating advanced cancer has been targeting immune checkpoints that act as natural regulators of immune cell activity. Tumours can exploit the immunosuppressive activity of immune checkpoints, gaining considerable survival benefit, and the targeting of these immune checkpoints by MAbs is the preeminent form of IM therapy in cancers to date (Fig 1.0.1). In 1996, inhibition of CTLA-4 (cytotoxic T-lymphocyte-associated protein 4) which is chiefly characterised by its ability to inhibit activated cytotoxic T-cells, improved immune mediated anti-tumoural response in animal models,[39] and in 2011 anti-CTLA-4 MAbs (ipilimumab) became the first checkpoint inhibitor approved to treat cancer (stage IV melanoma) by the FDA (USA Food and Drug Administration).[40] CTLA-4 inhibition was trialled in various other tumours, but saw little initial success in NSCLC. The development of antibodies against another immune-checkpoint, PD-1

(progammed-cell-death-protein-1) followed, with one such IM therapy (pembrolizumab) gaining approval to use in advanced melanoma in 2014 by the FDA, followed by another (nivolumab) in the same year.[41, 42] In 2015, both anti PD-1 agents were approved for use in advanced NSCLC.[43, 44] In 2016, MAbs targeted against the main ligand for PD-1, PD-L1 (programmed death ligand one) (atezolizumab) was approved in NSCLC and urothelial cancers.[45] IM therapy targeting this interaction of PD-1 and PD-L1, the 'PD-1/PD-L1 axis' has seen a rapid expansion in their approval, with trials for virtually every tumour type in progress. Approval in NSCLC at various stages has been granted, as well as a wide variety of other cancers including gastric, breast, head and neck and urothelial carcinomas.[46-54] The use of PD-1/PD-L1 IM therapy in the context of NSCLC is broad, including single-agent use or combined with XCT and/or CTLA-4 IM therapy, and utilised at various stages of NSCLC, including more recently efforts to use it in the neoadjuvant setting.[55, 56] Many other immune related therapies being developed including CAR-T (Chimeric antigen receptor T cells), TCR (T-cell receptor) gene-modified T-cell therapy, oncolytic viruses and tumour vaccines, but this thesis is concerned chiefly with the inhibition of the PD-1/PD-L1 immune checkpoint process in the context of NSCLC.

1.1 PD-1 and PD-L1; a brief introduction

1.1.0 PD-1

PD-1 (also known as CD279) was first described in 1992 as a transmembrane receptor with a mature form length of 268 amino acids (50-55kda (kilodaltons)) with four extracellular N-glycosylation sites expressed on many immune cells, including T-cells, B-cells, NK-cells (natural killer), macrophages and DCs (dendritic cells), and is particularly upregulated on tumour specific T-cells.[57, 58] PD-1 acts as an inhibitor of both the adaptive and innate immune systems and has a diverse role, but its function is best understood and characterised within the context of T-cells. PD-1 is selectively upregulated in activated T-cells in response to ongoing antigen exposure.[59] Its main function is to act as a co-inhibitory signal when an APC (antigen presenting cell) presents antigens to a T-cell. When bound by a suitable ligand (e.g. PD-L1) PD-1 intracellular tyrosine is phosphorylated and activated, which results in the recruitment of SHP-1 (Src homology 2 domain-containing protein tyrosine phosphatase-1) and SHP-2 to the C-terminal ITSM, (immunoreceptors tyrosine-based switch motif), which is the site of PD-1 intra-cellular activity. These phosphatases act to counter the stimulatory signals being sent from CD28, via inhibition of ZAP70 (Zeta-chain-associated protein kinase 70) and CD3δ which in turn inhibits the PI3K/AKT and RAS/ERK signalling pathways, which results in the decreased activation of transcription factors such as AP-1 (Activator protein 1), NFAT (Nuclear factor of activated T-cells) and NF-κB (Nuclear factor-κB), which drive T-cell activation, proliferation, effector functions and survival.[60, 61]

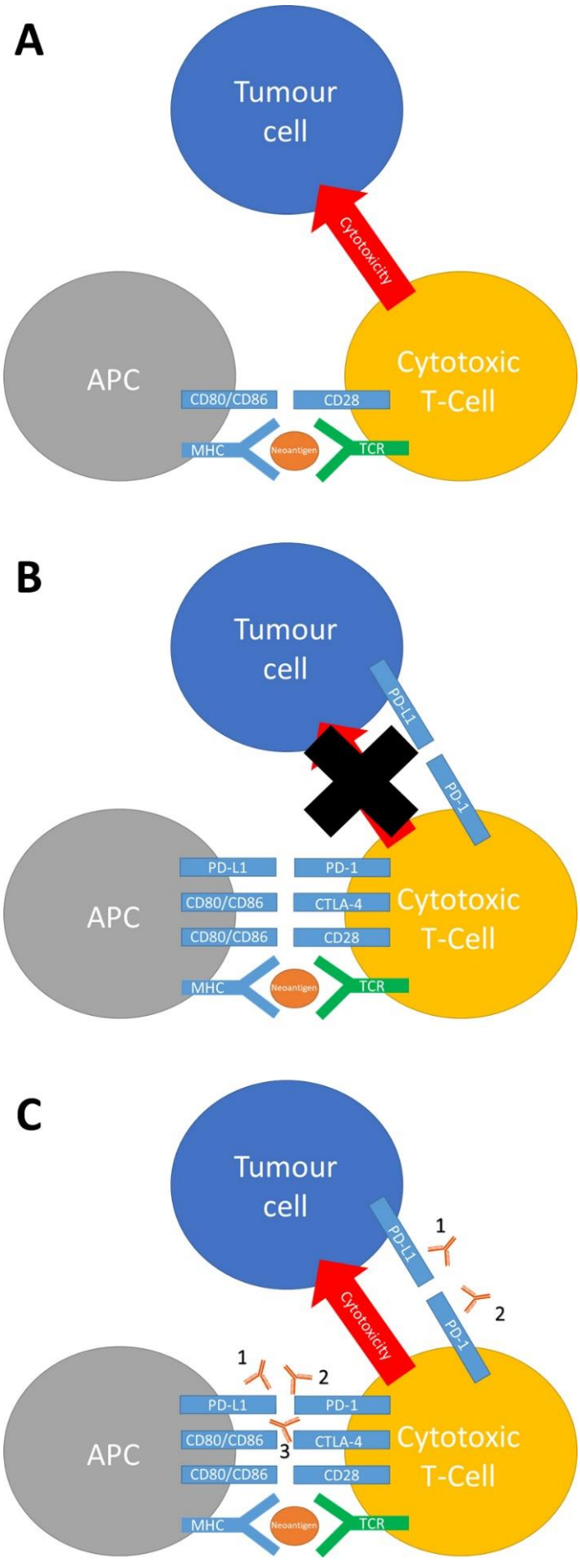


Figure 1.0.1 Principles behind immune-checkpoint inhibition.

A – APCs present antigens to cytotoxic T-cells via class II MHCs. The T-cell requires co-stimulation via interaction of CD80/CD86 on the APC with CD28 on the T-cell, which results in cytotoxic activity against the tumour cell.
B – Regulation of cytotoxic activity includes the expression of immune checkpoints on T-cells including CTLA-4, which also binds to CD80/CD86 but with a higher affinity than CD28 and acts as a co-inhibitor, and PD-1, which is bound by PD-L1 expressed by both APCs and tumour cells directly, which also acts as an co-inhibitor, resulting in decreased cytotoxic activity and immune escape by the tumour cell.
C – Monoclonal antibodies targeted against the immune checkpoints include (1) PD-L1 (e.g. atezolizumab) (2) PD-1 (e.g. nivolumab, pembrolizumab) and (3) CTLA-4 (e.g. ipilimumab) which prevent co-inhibition of the T-cell, resulting in reversal of immune escape and renewed cytotoxic activity against the tumour cell.

APC, antigen presenting cell; MHC, major histocompatibility complexes; TCR, T-cell receptor; PD-1, programmed cell death protein 1; PD-L1, programmed death ligand 1; CTLA-4, cytotoxic T-lymphocyte-associated protein 4)

PD-1 can also activate PTEN (Phosphatase and tensin homolog), thereby further inhibiting PI3K/AKT pathway[62] and inhibit PKCδ (Protein Kinase C-δ) signalling further reducing NF-κB activity.[60, 61, 63-65] These are summarised in Fig 1.4.0.

Ultimately, the activation of PD-1 results in the inhibition, dysfunction, and eventual cell death of T-cells, and as PD-1 is upregulated specifically in response to antigens this can be a beneficial mechanism (e.g. in autoimmune disorders or with ineffective immune responses) but also harmful, particularly in the context of ongoing antigenic stimulation from cancer cells resulting in less effective immune responses.

1.1.1 PD-L1 and PD-L2

PD-1 has two main ligands: PD-L1 and PD-L2. PD-L2 is less well characterised and has received less attention than PD-L1, although it appears to play a similar role in generating immunosuppressive activity via PD-1 and has a distinct expression pattern. It may be a distinct predictor of response to anti-PD-1 therapy, but its role in within cancers is thought to be much less significant than PD-L1.[66-70] PD-L1 (also known as B7-H1) is a 290 amino acid transmembrane glycoprotein that was renamed to PD-L1 when it was discovered to interact with PD-1.[71, 72] The classic mechanism of action described for PD-L1 is to bind to PD-1 expressed on activated T-cells in order to inhibit their cytotoxic, and therefore anti-tumoural activity. Professional APCs are critical to this process as T-cells cannot recognise soluble antigens and require tumour neoantigens to be presented to them via class II MHCs which, alongside CD28 mediated co-stimulation, results in the activation of naïve T-cells and subsequent T-cell proliferation and differentiation, cytokine production and beginnings of the T-cell mediated adaptive immune response.[73, 74] As part of normal constitutive activity, the regulatory processes involved that act to regulate self-tolerance and excessive activity of T-cells include the activity of immune checkpoints such as CTLA4 and the PD-1/PD-L1 axis.[75] The exploitation of these immune checkpoints by tumour cells and the subsequent targeting of these is the fundamental basis for PD-1/PD-L1 IM therapy. (Fig 1.0.1). This thesis will largely be concerned with this 'classic' mechanism of action, although alternative pathways and variants are considered as appropriate throughout. For the avoidance of confusion, 'PD-L1' used throughout this work refers to the mature, 290 amino-acid length, membrane bound glycoprotein, unless otherwise specified.

PD-L1 has been shown to be expressed in several normal tissue types, such as epithelial and immune cells within the tonsil, endothelial cells including the myocardial endothelium, cortical thymic epithelial cells, and syncyciotrophoblasts, and has been shown to have an immuno-regulatory function in some of these tissues.[70, 76-78] PD-L1 is also expressed widely and heterogeneously in many solid tumours, including NSCLC, gastric, breast, melanoma, mesothelioma and other cancers.[73] The

expression of PD-L1 is used to determine treatment to PD-1/PD-L1 IM therapy, but to understand the role of the PD-1/PD-L1 axis in cancer, a more comprehensive review of the relationship between cancer cells and immune cells must first be understood.

1.2 Cancer and the immune system

1.2.0 Cancer cells and immune cells

The relationship between the immune system and cancer cells is an intimate and inexorably linked one, even at the earliest stages of tumour development. Certain immunoincompetent groups are more at risk of developing specific cancers (e.g Kaposi's sarcoma in immunodeficient patients[79]) and around 25% of cancers have an origin related to infection (e.g. Hepatitis C is associated with hepatocellular carcinoma) or other inflammatory processes (e.g. prostatitis is associated with prostate cancer).[80, 81]

The early detection and eradication of cancerous and pre-cancerous cells by the immune system, or immunosurveillance, is an extension of the immune system's ability to detect and eradicate abnormal cells in the non-malignant setting, such as viral or bacterial infections. The notion that the immune system could target cancer cells was suggested by Ehrlich in 1909[82] and more formally conceived as immunosurvellance by Burnett and Thomas in the 1950s. [83, 84] However, this work was mostly conducted in animal models, and various immuno-deficient mice models failed to show conclusive evidence of higher rates of tumours and it wasn't until the 1990s that definitive data could be generated that showed immunocompromised mice were indeed associated with increased rates of malignancy.[85-87] In 2002 Schreiber's group described the concept of immunoediting based on these observations and is thought of as the selective growth of tumour cells due to their innate ability to overcome immune responses targeted against them.[88] Therefore the immune response targeted against tumour cells is a chief factor in determining the final cancer cell populations present; susceptible tumour cells are eradicated, but tumour cells with innate protection against the immune response become the dominant cell population, and thus immunity conveys both pro-tumoural and anti-tumoural activity.[89, 90]

1.2.1 Immunoediting

The immunoediting theory involves 3 phases of immune response to the tumour: elimination, equilibrium and immune escape.[88, 89] Although direct evidence is limited, elimination is considered the earliest response to cancer cells – where spontaneous destruction and regression of tumours occur as a result of T-cell activity and related mechanisms destroying pre-cancerous cells before they are able to develop more advanced immune defences.[91, 92] As a result, tumours maybe

spontaneously eradicated without a clinical manifestation of the disease.[93] During the equilibrium phase, this dominance is lost, the immune system is unable to destroy or eradicate the tumour – but it still has capacity to prevent tumour growth, aggressive invasion and metastases. Again there is little direct evidence of this phase, but there is evidence that cancer cells can become "dormant", thereby temporarily evading immune cell scrutiny, which allows them to reactivate at a later date – and indeed tumours in patients can remain quiescent for months and even years with an immune insult the potential triggering factor for tumour growth and aggressive behaviour.[94-96]

The final phase, immune escape, is the phase most patients are encountered, and the phase at which most research is undertaken, and is when the immune response is insufficient to limit or control the tumour growth, including the invasion or metastatic spread of cancer cells. An effective immune response against tumours requires the expression of neoantigens on tumour cells, and a conducive TME, the regulation of which plays a major role in immune escape. Neoantigens are abnormal surface proteins expressed on tumour cells that have not been encountered by the host immune system and therefore not avoided by self-tolerance mechanisms, and identify these cells as 'foreign' to the adaptive immune process. A lack of neoantigens maybe due to low mutational burden tumours,[97] selective presentation of neoantigens by the tumour cells[98] or modulation of neoantigen expression secondary to IM therapy[99] all of which may lead to immune escape. Modulation of the TME by recruiting immunosuppressive cells to the TME can increase immune self-tolerance of tumour cells and decrease the effectiveness of cytotoxic elements acting directly against the tumour cells.[100] Another mechanism of immune escape is the expression of immune-checkpoints within the TME, which can directly inhibit the cytotoxic T-cells within the TME, and is the predominant pathway by which PD-1/PD-L1 is believed to exert a tumour survival benefit (Fig 1.4.0), with several other checkpoints such as TIM-3 (T cell immunoglobulin and mucin domain-containing protein 3), IDO1 (Indoleamine 2, 3-dioxygenase 1), LAG-3 (Lymphocyte-activation gene 3), and CTLA4 also described, many of which are the target of ongoing clinical trials (Fig 1.3.0).[75, 101] The modulation of the TME to achieve immune escape also involves regulation of cytokines, apoptotic pathways and other immunosuppressive pathways, and PD-1 and PD-L1 appear to mediate many of these. Therefore an overview of the various elements of the TME is necessary to fully appreciate the role of PD-1 and PD-L1 in cancer immune escape.

1.3 The Tumour microenvironment

1.3.0 Components of the TME

The TME is a complex structure that is composed of tumour cells, pre-malignant cells, non-tumour cells as well as extra-cellular matrix (ECM) and vasculature. Communication is a dynamic process

between all of these components and includes the production and expression of growth factors, cytokines, chemokines and exosomes. The focus of this thesis in regards to cells modulating the TME is predominantly that of the interaction of immune cells and tumour cells, however virtually every other cell type may play an important role in modulating the immune function of the TME. For example, normal fibroblasts have typically anti-tumoural activity by expressing anti-tumour cytokines (e.g. TNF-α (Tumour-necrosis factor α)) and suppressing tumour growth through modulation of the ECM,[102] whereas cancer associated fibroblasts (CAFs) (a heterogenous group of fibroblasts of probable multiple cell type origin) secrete growth factors and chemokines that stimulate tumour cell growth and survival and encourage the migration of other cells to the TME and are thus typically regarded as pro-tumoural in activity.[103] Furthermore, CAFs secrete numerous immunosuppressive cytokines (e.g. TGF-β (Transforming growth factor β), help manufacture collagen and matrix metalloproteinases (MMPs) that inhibit effective lymphocyte infiltration, and a sub-population of CAFs have been shown to express immune checkpoints including PD-L1 thereby contributing significantly to an immunosuppressive pro-tumoural TME.[104] Pericytes have been shown to regulate the TME by helping immune cells infiltrate the TME and adipocytes expressing PD-L1 may further contribute to immune escape.[105, 106]

Of particular relevance are the immune cells present within the TME: including T-cells, B-cells, NKs, DCs, MDSCs (myeloid derived suppressor cells), macrophages and neutrophils. As with many other aspects of the TME, these cells types can be divided into subcategories which typically play either a pro-inflammatory anti-tumoural role, or an immunosuppressive pro-tumoural role.

1.3.1 T-cells in the TME

T-cells play a diverse role in adaptive immunity. They can be classified into various subtypes, with anti-tumour T-cell subtypes including cytotoxic (CD8+ve) T-cells, Thelper (Th) (CD4+ve) cells and pro-tumoural T-cell subtypes including regulatory T-cells (Tregs) (FoxP3+ve). Th cells represents a diverse category of T-cells classically categorised as Th1, Th2, Th17 and Treg. Th1 play a role in immune surveillance and regulating intracellular pathogens and are associated with autoimmune disorders, Th2 help eliminate extracellular parasites and play a role in allergies and asthma, Th17 target extracellular bacteria and fungi and are also associated with autoimmune disorders, and Tregs play a role in immune tolerance and the regulation of immune responses, and are associated with immune suppression of TMEs.[107, 108] For the purposes of clarity, throughout this thesis CD4+ve T-cells will refer to non-Treg Th cells specifically. CD4+ve T-cells play a major role in assisting CD8+ve T-cells, both of which are involved in initial tumour development[91] and can help destroy tumour cells via

cytokine mediated pathways (e.g. IFN-γ (Interferon-γ) and TNF-α) and the activation of DC mediated cell death.[109]

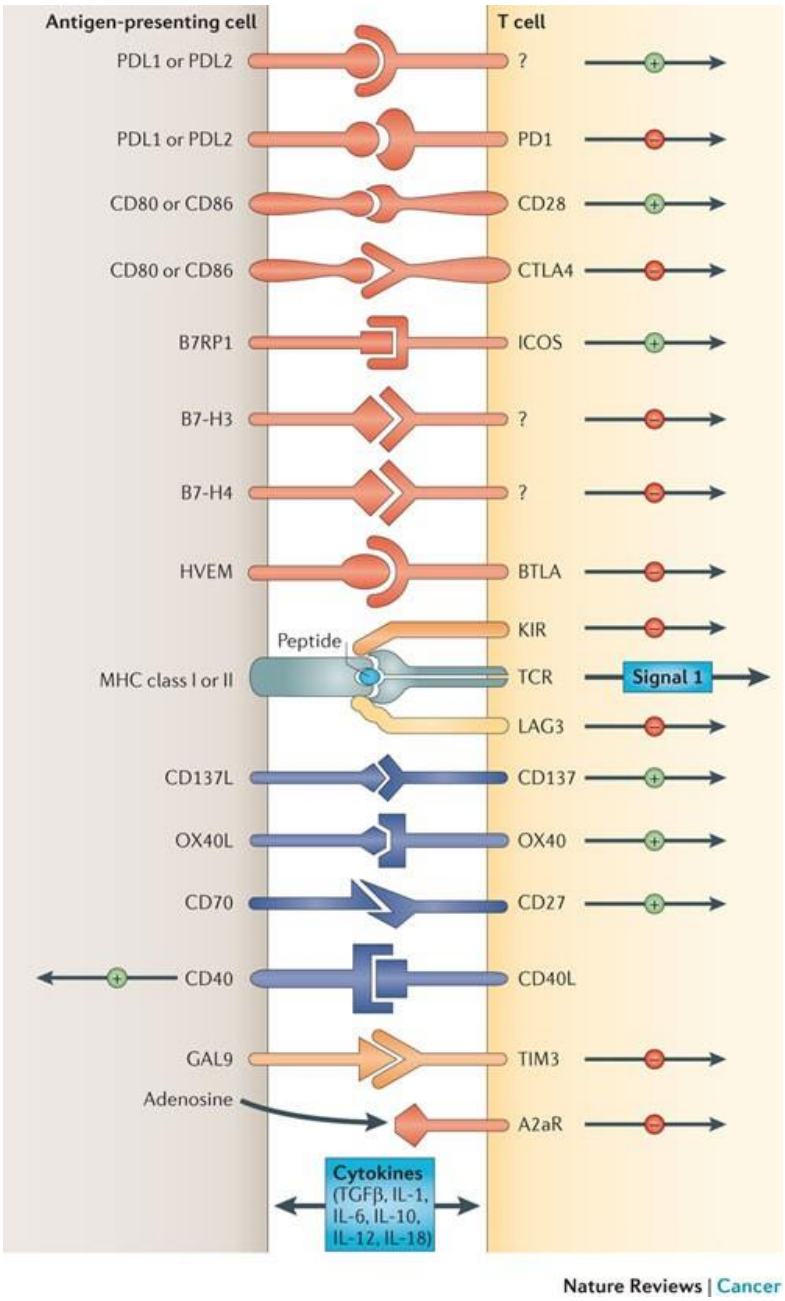


Fig 1.3.0 Immune checkpoint receptors and their ligands. Modified from Pardoll 2012.[75]

CD4+ve T-cells within the TME have been shown to upregulate expression of PD-1 (and other immune checkpoints such as TIM3 and LAG3) and decrease cytotoxic cytokine expression,[110]

suggesting a contributing role of T-cell exhaustion (described later) in CD4+ve T-cells to immune escape. Despite this, CD4+ve T-cells and the role they play within the context of TMEs and PD-1/PD-L1 mediated escape has generally received less attention than other T-cell types.

Tregs mediate immunosuppressive activity within the TME via a number of mechanisms, including the expression of CTLA4, production of immunosuppressive cytokines (e.g. IL-10 (Interleukin-10), TGF-β), the direct destruction of T-cells via granzyme mediated cytotoxicity, the competitive consumption of IL-2 and the conversion of ATP (Adenosine triphosphate) to adenosine, thereby reducing ATP-driven T-cell cytotoxicity which ultimately results in the suppression of CD8+ve and CD4+ve T-cells activity.[108, 111-115]Despite these immunosuppressive features acting in contrast to CD4+ve and CD8+ve T-cells, their activation is still dependent on TCR and CD28 mediated signalling, and the fact they can also express PD-1 implies this will result in Treg inhibition and eventual exhaustion. Indeed the blockade of PD-1, which typically results in increased immune and anti-tumoural activity, can also increase Treg activity and enhance its immunosuppressive functions.[108, 116]

The most studied T-cell in the context of PD-1/PD-L1 mediated immune escape targeted by PD-1/PD-L1 IM therapy are the cytotoxic CD8+ve T-cells. The increasing knowledge of the role that PD-1/PD-L1 plays outside CD8+ve T-cells can go some way to explaining much of the variable response to IM therapy, but the major mechanism of action is still believed to mediated by these cells. As such, cytotoxic T-cells are a major focus of this thesis, and the steps required to traffic CD8+ve T-cells to the TME and their activity within the TME is discussed in more detail in Chapter 7. In brief, within the TME CD8+ve T-cells are activated by neoantigen presentation via class II MHCs on APCs with co-stimulation by CD28 mediated signalling and extracellular cytokine activity. Once activated, CD8+ve T-cells can exert cytotoxic activity on targeted tumour cells via several pathways: the release of lytic granules (e.g. perforin that forms pores in target cells, and various granzymes that activate apoptotic pathways within the target cells) in a Ca2+ mediated fashion, Fas signalling in an Ca2+ independent fashion that activates caspase mediated apoptosis and the release of cytotoxic cytokines (e.g. IFN-γ, TNF-α).[117, 118] The presence of co-stimulatory mediated PD-1 on CD8+ve T-cells, and the binding of this by PD-L1 within the TME to inhibit cytotoxic activity is the cornerstone rationale behind PD-1/PD-L1 IM therapy.

It is also worth mentioning that the activity of T-cells, the regulatory nature of PD-1/PD-L1 and the impact of IM therapy on these areas is not limited to cells within the TME. Indeed the role of PD-1 in preventing auto-immune disorders is well described, and dysregulation in this pathway is implicated in many such diseases, including Type 1 Diabetes Mellitus, multiple sclerosis, rheumatoid arthritis

and others. [119, 120] Perhaps unsurprisingly therefore, autoimmune toxicities and subsequent irAEs (immune related adverse events) that result from treatment by PD-1/PD-L1 IM is a major clinical challenge, as their impact on immune checkpoint inhibition is not limited to the TME, and virtually every organ in the body maybe affected by this. [121]

1.3.2 Other immune cells within the TME

Two important immune cell types for CD8+ve mediated cytotoxicity are DCs and TAMs, as a result of their professional APC behaviour. DCs play a pivotal role in linking innate and adaptive immune systems together and are the most effective professional APCs and play a critical role in presenting neoantigens to T-cells in the TME.[122] As such they are major targets of tumour cell mediated immune escape via the tumour cell's ability to negatively regulating DCs' efficiency and functionality as APCs.[123] Various immunosuppressive cytokines (e.g. TGF-β, IL-13, IL-10 GM-CSF (Granulocyte-macrophage colony-stimulating factor)) within the TME can directly inhibit DC activity or convert DCs to immunosuppressive immune cells such as M2 TAMs and MDSCs.[123-125] Tregs can also inhibit DCs by the expression of CTLA4, resulting in increased DC expression of PD-1 and by the expression of inhibitory cytokines IL-10 and TGF-β.[126, 127] Expression of immune checkpoints, including PD-L1, appear to inhibit DC function, and in turn the blockade of PD-L1 can improve DC mediated anti-tumoural activity.[123, 128]

Macrophages are classically split into polarised subtypes: M1 which have pro-inflammatory behaviour and are therefore associated with anti-tumour activity, and M2 which have tissue repair and cell proliferation behaviours, and are therefore associated with tumour promoting activity.[129] M1 are typically activated by anti-tumoural cytokines such as IFN-γ and TNF-α mediated pathways whereas M2 are activated by immunosuppressive cytokines (e.g. IL-10, IL-13).[130, 131] TAMs can be composed of macrophages polarising in either direction, but are mostly considered to resemble pro-tumoural M2 macrophages. As a result, TAMs can promote proliferation and metastases of tumour cells, promote angiogenesis and inhibit anti-tumour T-cell activity.[130, 132] The plasticity of macrophages is such that they can be polarised in a particular direction depending on other features and TAMs may therefore play both anti-tumoural and pro-tumoural roles within the TME, with this plasticity a possible target in regulating cancer progression.[130, 133] As macrophages also play a role as professional APCs, the expression of PD-L1 in macrophages contributes to tumour immune escape, provides macrophage protection against T-cell cytotoxic activity[134] and may play a role in macrophage proliferation and activation[135] with an overall trend towards M2 phenotype polarisation.[136] PD-1 is also expressed by macrophages within the TME and appears to inhibit phagocytosis of macrophages and also exert M2 phenotype polarisation.[133] Despite their role as

APCs presenting neoantigens to T-cells, TAMs, and in particular PD-1/PD-L1 expressing TAMs, overall appear to generate an immunosuppressive TME.

Other immune cells within the TME also interact with PD-1/PD-L1 axis. B-cells are traditionally described as key regulators of adaptive immunity by the production of antibodies once terminally differentiated to plasma cells. However they also play a diverse role within immune responses by functioning as APCs, producing cytokines and chemokines and regulating the activation and expression of T-cells.[137] B-cells are composed of multiple subtypes but typically play an immunosuppressive, pro-tumoural role within the TME, with specific subtypes (e.g. CD40 +ve) potential targets for clinical trials.[138] An important classification of B-cell subtypes are B-regulatory cells (Bregs) that play a role in inflammation, auto-immune disease and cancer[139] and exert an immunosuppressive activity by expression of immunosuppressive cytokines (e.g. IL-10, TGF-β) and the upregulation of FoxP3+ve Tregs and CTLA4 expression.[140] The role of Bregs specifically in NSCLC is not well characterised[139] but PD-L1 expression has been shown to be higher on Bregs versus other B-cell subtypes, and that PD-L1 expression on Bregs may play a key role in suppressing non-Treg expressing PD-1 T-cells and increasing Tregs, thereby further contributing to immunosuppressive pro-tumoural activity.[141, 142]

Neutrophils are a key component of innate immunity with phagocytic activity and a role in producing reactive oxygen species (ROS) and toxic neutrophil granules. Their role in infectious disease and auto-immunity is well characterised, but their role is less well understood in the context of the TME. Tumour infiltrating neutrophils (TANs) within the TME play a predominantly pro-tumoural role by promoting tumour cell proliferation and invasion, angiogenesis and inducing T-cell apoptosis.[143] However, TANs can also play a anti-tumoural role via cytotoxic activity by production of ROS and the recruitment and activation of CD4+ve and CD8+ve T-cells and may have a similar 'polarisation' of N1 and N2 subtypes similar to macrophages.[144, 145] In regards to PD-1/PD-L1 interaction, TANs have been shown to express PD-L1[146] and recently that blocking PD-1/PD-L1 interaction enhances anti-tumoural cytotoxic activity of TANs.[147]

NKs are part of the innate immune system and can regulate other parts of innate immunity as well as adaptive immune cells. They produce a large number of cytotoxic cytokines such as IFN-γ, TNF-α, and GM-CSF. Self-tolerance of NKs is mediated by self-antigens presented via class I MHCs on host cells. Loss of class I MHC in tumour cells prevents inhibition of NKs and results in direct and indirect cytotoxic activity, and therefore suggests a predominantly anti-tumoural role of NKs.[148] NKs have been shown to express PD-1 in normal patients and in tumour patients[149] and helps mediate CD8+ve cytotoxic activity within the TME.[150] The blockade of an inhibitory receptor expressed on both NKs

and T-cells (NKG2A) may have increased efficacy alongside blockade of PD-1/PD-L1 in increasing NKs cytotoxic anti-tumoural behaviour.[151]

MDSCs are key innate immune cells and play a predominantly pro-tumoural role by exerting significant immunosuppressive activity on innate and adaptive immune responses by multiple mechanisms including expression of immunosuppressive cytokines (e.g. IL-10 and TGF-β) and depletion of amino-acids required for effective T-cell function.[152, 153] MDSCs have also been shown to express PD-L1, the blockade of which has been shown to inhibit PD-L1 mediated inhibition of T-cells and reduced immunosuppressive cytokine production.[154, 155] Furthermore, the presence of MDSCs in NSCLC has been shown to be associated with poorer response to PD-1/PD-L1 IM therapy.[156, 157]

1.3.3 Cytokine activity within the TME

The ability of immune cells to impact the TME is frequently via the production of various cytokines. Their specific relationship to PD-1/PD-L1 is discussed later, but, as with immune cells, cytokine activity is complex and can have both pro-tumoural and anti-tumoural activity depending on other features within the TME. For example, IFN-γ (a type II interferon cytokine) produced by T-cells and NK cells typically acts in an anti-tumoural fashion by inducing tumour cell cycle arrest and apoptosis by direct mechanisms or by modulating the immune aspect of the TME such as by inhibiting immunosuppressive Tregs, MDSCs and M2 macrophages. However, IFN-γ can also play a role in pro-tumoural activity by encouraging tumourigenesis and angiogenesis and increasing tumour immune tolerance, including the upregulation of PD-L1 expression in tumour cells and APCs.[158-160] TNF-α is produced by CD4+ve T-cells, NK, macrophages and other cells and is a cytotoxic molecule induced by the MAPK, NF-Kb and apoptosis pathways. Despite the strong cytotoxic activity it typically portrays, TNF-α can also increase the numbers of MDSCs and neutrophils in the TME which can result in increased tumour growth by angiogenesis and amplification of inflammation.[60, 161] Many interleukins are upregulated within the TME and have a diverse range of activity, for example, IL-1β is induced by a number of pathways including MAPK, mTOR (Mechanistic target of rapamycin) and NF-KB, and is expressed by NK, macrophages, T-cells and fibroblasts. It promotes inflammation and macrophage differentiation towards M1 and differentiation of B cells to plasma cells, and increases IFN-γ levels, cytotoxic T-cells and T-cell polarisation to Th1. Despite this IL-1β can also increase immunosuppressive M2 macrophages and increase VEGF and FGF (fibroblast growth factor) that support angiogenesis and metastasis, and indeed IL-1β is thought to play a key role in NSCLC tumourigenesis, with its presence an indicator of poorer survival for these patients.[162-164] TGF-β has also been shown to have both anti-tumoural and pro-tumoural activity, and is a potent tumour-suppressor in pre-malignant cells, but in later stages promotes tumour metastases.[165]

There are many more cytokines active within a typical TME, but these illustrate that the role of any given factor is complex, with both potential pro-tumoural and anti-tumoural activity. TMEs may therefore be thought of as being generally immunosuppressive and pro-tumoral or cytotoxic and anti-tumoural, but is realistically composed of a dynamic mixture of these properties. Therefore effectors and regulators of immune-mediated tumour responses must be looked at in the broader context of other factors within then TME, rather than by considering each in isolation. As the components of the TME are dynamic and change over time, a single tumour sample can provide an accurate snapshot of the tumour-immune relationship, but it is important to remember the previous and future activity of the TME not captured by a single specimen.

1.4 The role of PD-1/PD-L1 in mediating tumour immune escape

1.4.0 Alternative mechanisms of action

In addition to the conventional mechanism of action described above (Fig 1.0.1), PD-L1 has also been shown to modulate immune response to cancer by a number of other pathways, including PD-1 independent pathways, most of which are pro-tumoural in effect. One of these is PD-L1's ability to directly alter mechanisms within tumour cells, so called 'tumour intrinsic pathways', which includes inducing anti-apoptotic survival in tumour cells in a 'reverse signal' pathway from PD-1 stimulation,[166] acting as a direct defence against IFN-γ via the STAT3 pathway in tumour cells, which can be reversed by PD-L1 blockade[167] and inducing mTOR mediated survival, inhibiting autophagy and increasing glycolysis in a PD-1 independent fashion.[168-170] PD-L1 can also induce immunosuppressive features within the TME outside of direct T-cell inhibition. For example, PD-L1 expression on macrophages appears to be associated with M1/M2 polarization, altered cytokine activity and macrophage antigen expression, that may have an overall immunosuppressive effect including inducing T-cell anergy[133] and PD-L1 mediated signalling on T-cells that results in the indirect recruitment of MDSCs which also increase the immunosuppressive nature of the TME.[171] Furthermore, PD-L1 has splice variants, soluble forms, variants with homodimerisation abilities as well as a large number of post-translational modifications, the full role of which has not been well characterised.[172-174]

PD-1 can also exert immunosuppressive activity outside of the classic PD-1/PD-L1 axis. PD-1 may cause anergy of T-cells without PD-L1 via the SHP2 mediated activity on PI3K and ERK pathways via distinct Ca2+ mediated mechanisms.[175]

Therefore, although different tumours may share PD-L1 expression as their primary method of immune escape – the direct mechanisms and downstream effects by which it evades the immune

response are not necessarily the same. Indeed even within the same tumour multiple mechanisms of PD-L1 mediated immune escape processes maybe active.

The basic principle, however, is that PD-L1 and PD-1 can play an integral role in mediating tumour immune escape, predominantly through the 'classic mechanism' of direct inhibition of cytotoxic CD8+ve T-cells, resulting in their dysfunction and eventual cell death, but also by evoking a more immunosuppressive TME through tumour intrinsic pathways and the regulation of various cytokines and immune cells. Given the diverse range of immunosuppressive activity that PD-1 and PD-L1 can achieve, the multiple pathways of regulating PD-1 and PD-L1 expression will now be explored.

1.4.1 Regulation of PD-1 expression

The regulation of PD-1 expression is not well characterised. It is known to be very low in naïve T-cells but rapidly upregulated when activated. This process is in part regulated by TGF-β[176] as well as various interleukins (e.g. IL-2, IL-6, IL-12 and IL-21).[177, 178] Binding of the TCR or activity of various cytokines may regulate PD-1 expression via transcriptional factors (e.g. NFATc1 (Nuclear factor of activated T-cells C1), AP-1, NF-κB and STATs (Signal transducer and activator of transcription).[60, 179] PD-1 expression can also be effected by epigenetic regulation (e.g. histone modification, miRNAs (micro-ribose nucleic acid) and methylation) and protein regulation (e.g. FBXO38, TOX, FUT8), but overall remains a poorly characterised area.[60, 179]

1.4.2 Regulation of PD-L1 expression – genetic and epigenetic factors

PD-L1 expression is regulated by several cell survival pathways known to play a role in cancer survival. (Fig 1.4.0) Activation of the MAPK pathway (a protein kinase signalling cascade that catalyses several pathways, including the Raf/MEK/ERK pathway indicated in driving many tumours including *KRAS* and *EGFR* mutant oncodriven NSCLCs) upregulates PD-L1 expression, and MEK inhibition downregulates PD-L1 expression.[180-182] The PI3K/Akt pathway (an intra-cellular signalling pathway that plays a role in cell survival, growth and proliferation, in both T-cells and in tumours, and can also be activated by *EGFR* mutants) has also been implicated in upregulating PD-L1 expression, possibly as a result of PTEN inactivation.[183, 184] The JAK (Janus kinase)/STAT pathway (also an intra-cellular signalling pathway responsible for growth factors and cytokine production indication in a number of tumours, including *EGFR* driven NSCLC) has also been shown to potentially be able to upregulate PD-L1.[184, 185] These three pathways can be activated by a large number of upstream signals, suggesting that PD-L1 upregulation via these pathways is a normal function in modulating immune response. However, they also provide potential mechanisms for how tumour cells, such as *EGFR* mutant driven tumours, can upregulate PD-L1 in a manner that may not be

reactive (IE the upregulation is a result of aberrant activation of cell signalling pathways, rather than as a result of selective tumour cell survival from PD-L1 mediated immune escape) which may also explain why *EGFR* mutant patients are generally quite resistant to PD-1/PD-L1 IMs even in the context of PD-L1 positivity, and why PD-L1 expression that is constitutively expressed in this fashion is a poor candidate for response to IM therapy.[186, 187]

miRNAs can also play a role in regulating PD-L1. Typically miRNAs act to regulate proteins by binding to mRNA (messenger RNA) and promoting its degradation or inhibiting its translation.[188] Direct binding of various miRNAs to PD-L1 mRNA has been shown to occur in a variety of cancers, including NSCLC.[60, 189, 190] miRNAs can also inhibit PTEN which acts to suppress PD-L1, resulting in the indirect promotion of PD-L1[191] and can induce PD-L1 via STAT3 in NSCLC.[192] miRNAs also play a role in IFN-γ mediated PD-L1 expression; when IFN-γ upregulates PD-L1 it also regulates various miRNAs, including miR-513 and miR-155. IFN-γ upregulates miR-155 and downregulates miR-513 that increases the overall amount of PD-L1 expression.[193, 194] miR-200 has also been shown to downregulate PD-L1 expression levels, and suppression of miR-200 by ZEB1 appears to play a role in NSCLC progression and metastases.[189]. These processes are summarised in Fig 1.4.0

1.4.3 Regulation of PD-L1 expression – cytokines and environmental factors

IFN-γ is generally considered to be the main cytokine responsible for upregulating PD-L1 expression. IFN-γ induces PD-L1 expression through the JAK/STAT3 and PI3K-AKT signalling pathways, leading to immune escape,[136] although it is interesting to note that there is an overlap, though nonetheless distinct pathway of PD-L1 upregulation by IFN-γ in different tumour types, including, for example, JAK2/STAT1/IFR-1 in gastric cancer[195] and MEK/ERK signalling pathway in myeloma.[196] However IFN-γ expression is associated with both pro-tumoural and anti-tumoural activity: high levels of IFN-γ expression are associated with anti-tumoural activity as its limits angiogenesis, prevents metastases and limits expression of immunosuppressive cytokines,[197] but treatment with IFN-γ conveys a worse survival in melanoma[198] and ovarian cancer[199] with no success seen in NSCLC patients beyond some early clinical trials.[200, 201] Therefore the expression of IFN-γ alone is not enough to be considered anti-tumoural or pro-tumoural, and although it can help to elicit immune escape via PD-L1 upregulation, IFN-γ expression requires context to properly interpret. TNF-α is also involved in PD-L1 regulation. TNF-α can induce inflammation, apoptosis, necrotic cell death and impair tumour cell proliferation,[60, 161] but can also promote tumour cell proliferation in a number of tumours, including skin and ovarian as well as the proliferation and metastases of NSCLC[202, 203] and contributes to immune escape via immunomodulatory pathways including the upregulation of PD-L1 expression, by mechanisms that are both distinct [204] and the same as IFN-γ.[194]

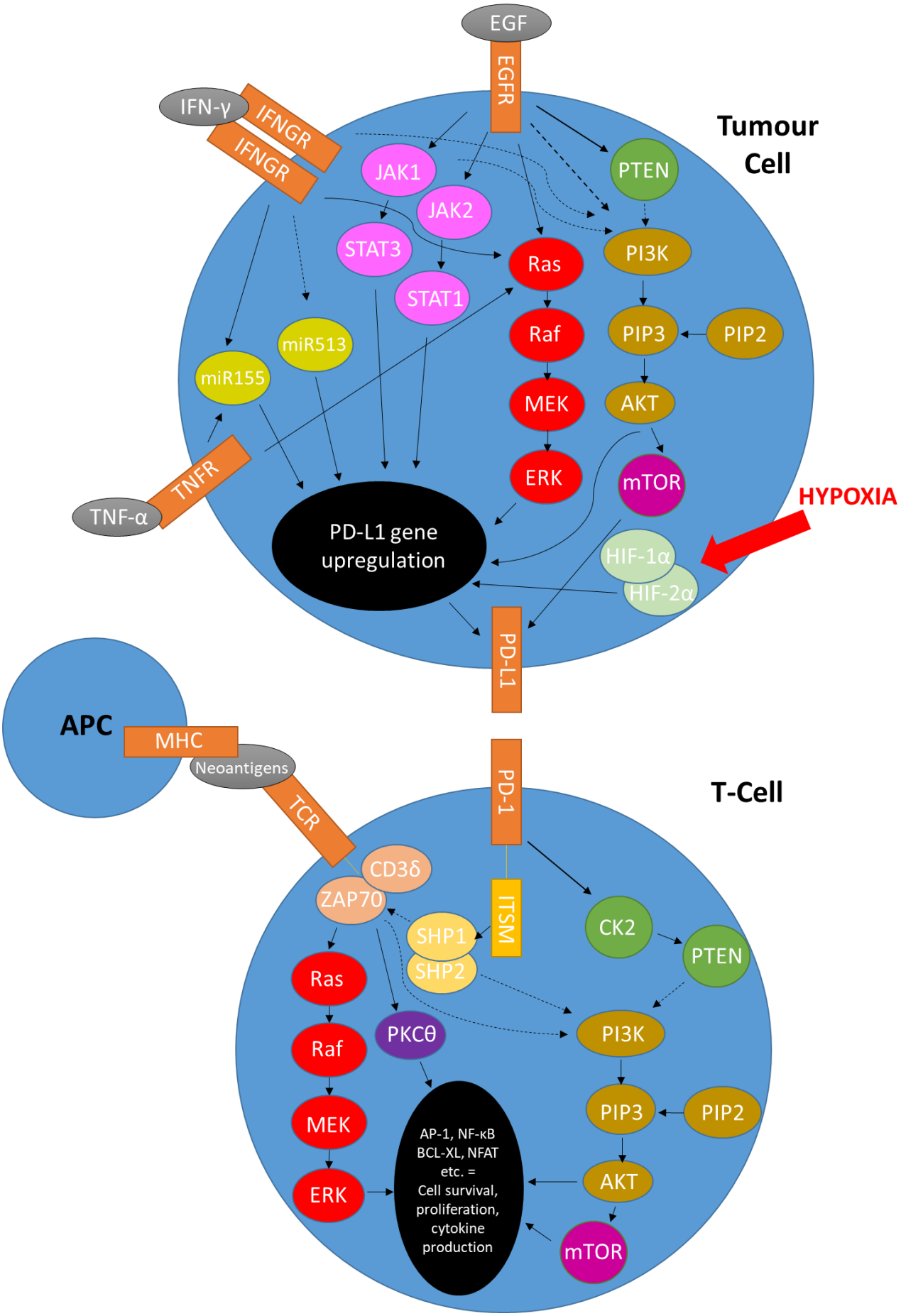

EGF
EGFR
IFN-γ
IFNGR
IFNGR
Tumour Cell
JAK1
JAK2
STAT3
STAT1
PTEN
PI3K
Ras
Raf
MEK
ERK
PIP3
PIP2
AKT
mTOR
miR513
miR155
TNFR
TNF-α
HYPOXIA
HIF-1α
HIF-2α
PD-L1 gene upregulation
PD-L1
APC
MHC
Neoantigens
TCR
PD-1
T-Cell
CD3δ
ZAP70
ITSM
SHP1
SHP2
CK2
PTEN
Ras
Raf
MEK
ERK
PKCθ
PI3K
PIP3
PIP2
AKT
mTOR
AP-1, NF-κB
BCL-XL, NFAT
etc. =
Cell survival,
proliferation,
cytokine
production

Fig 1.4.0 PD-1/PD-L1 axis intra-cellular signalling **A** – Regulatory pathways that increase PD-L1 expression by increased gene expression, translation or stabilisation of protein. (Interleukins can use similar pathways – IL-17 uses PI3K/Akt pathway, IL-6, IL-17, IL-25, IL-27 uses JAK/STAT signalling etc. to upregulate PD-L1 expression (not shown)). **B** – Downstream signalling of PD-1 when bound by ligand resulting in increased cell survival and cytotoxic activity of activated CD8+ve T-cell. Full-line arrows indicate increased activity of target, dotted arrows indicate reduced activity of target.

Various interleukins also play an important role in the regulation of PD-L1 expression. Interleukins play a diverse role in immunity, cell survival and anti-tumoural activity, and their behaviour can vary depending on context, as already described for IL-1β. However, most are thought of as having predominantly anti-inflammatory activity (e.g. IL-10, IL-13) or pro-inflammatory activity (e.g. IL-1β, IL-6, IL-23).[205] Immunosuppressive interleukins such as IL-27 and IL-10 can induce PD-L1 in T-cells and tumour cells, including in NSCLC,[206, 207] as can pro-inflammatory interleukins such as IL-6, IL-12 and IL-17.[208-210] Many of these pathways are not fully understood, but some overlap with IFN-γ and TNF-α pathways, and others have distinct pathways, but again the overall regulation of PD-L1 by these pathways is best understood in the context of other cytokine activity.

TGF-β can both upregulate and downregulate PD-L1 depending on context, although it typically acts to upregulate PD-L1 in NSCLC specifically.[60] Hypoxia-inducible factors (HIFs) regulate protein expression in hypoxic conditions,[211] and has been shown to upregulate PD-L1 in a various immune and cancer cells in these conditions.[154, 212] Although there is no direct evidence of HIFs regulating PD-L1 in NSCLC *per se*, patients receiving both an anti-angiogenic agent (bevacizumab) and anti-PD-L1 therapy (atezolizumab) with XCT in non-squamous NSCLC had superior outcomes to patients only receiving either bevacizumab with XCT or atezolizumab with XCT[213] suggesting HIFs plays a clinically significant role.

Virtually all of the environmental factors within the TME that are indicated in regulating reactive PD-L1 expression can have pro-tumoural or anti-tumoural activity depending on other variables. The mere presence alone of any of these remains insufficient to ascertain if it is acting in an overtly pro-tumoural or anti-tumoural fashion by the encouragement or suppression of immune-escape, and this reinforces the need to look at as much detail of the TME and immune features in relationship to each other as possible.

The expression of PD-1 and PD-L1 is a frequent occurrence in NSCLC, and despite the complexity and variation of the TME within and between tumours, there remains a significant proportion of patients whom respond favourably to PD-1/PD-L1 IMs. Many patients do not respond to IM therapy, however, even in the context of their tumours expressing PD-L1. The limitations of detecting PD-L1

via current clinical methods (and therefore resulting in possible false-negatives and false-positives of PD-L1 expression) will be discussed later, but an overview of the potential mechanisms to resistance to PD-1/PD-L1 IM therapy in the context of assumed genuine expression will now follow.

1.5 PD-1/PD-L1 inhibitor resistance

Primary resistance to PD-1/PD-L1 IMs is when there is no response or minimal response following the commencement of therapy, whereas secondary resistance is described in tumours that initially respond to treatment but fail to do so over time. Mechanisms of resistance may contribute to either or both, as a result of the dynamic nature of the TME.

1.5.0 Primary resistance – PD-L1 heterogeneity and constitutive expression

PD-L1 expression is clearly an important factor for deciding PD-1/PD-L1 IM therapy, but PD-L1 negative tumours can respond to treatment, and PD-L1 positive tumours may not.[46, 49] One explanation is that of tumour heterogeneity, in which various components of a tumour maybe highly variable over time and space. This is discussed in greater detail in Chapter 6 (Tumour heterogeneity), but in short, PD-L1 expression is not necessarily uniformly expressed across a tumour or between metastatic sites. Sampling with biopsies or needle aspirates will by necessity only include a small part of the overall tumour burden, and so for a highly heterogenous marker like PD-L1, any given sample may not be wholly representative of the disease as a whole, and therefore may mislead attempts to guide IM therapy decisions. PD-L1 expression may also be constitutive, rather than reactive, driven by pathways such as the MAPK, PI3K and JAK pathways, and in the context of tumours with no immune response, or tumours that are utilising other mechanisms of immune escape, may play no role in cell survival, and therefore the use of PD-1/PD-L1 IMs would be of limited value. This is discussed in greater detail in Chapter 7 (The tumour microenvironment), but as above, innate genetic and reactive immune regulatory processes of PD-L1 expression both play key roles but maybe quite distinct in predicting response to therapy. In addition, as has been described, PD-L1 can exert immunosuppressive activity in a PD-1 independent fashion, which would particularly limit the effectiveness of PD-1 IM therapy. Therefore although PD-L1 expression maybe genuinely associated with mediating immune escape, multiple mechanisms of PD-L1 mediated immune escape maybe present within a single tumour, not all of which would necessarily respond to PD-1/PD-L1 IM therapy.

1.5.1 Primary resistance – TILs and T-cell dysfunction

The presence, density and type of TILs (tumour infiltrating lymphocytes) is an important factor in immune escape and probable response to IM therapy. An immune response must be present,

sufficiently robust and of the correct type to be ideal candidates for IM therapy. In addition, shared properties of TILs can be both intuitively favourable and unfavourable for IM therapy; for example, high levels of PD-1 expression would imply the PD-1/PD-L1 axis is the main mechanism of immune escape, but high PD-1 expression is also associated with T-cells that are exhausted,[214] and so this may be a favourable or unfavourable marker of response to treatment, depending on the status of the T-cells.

T-cell anergy is a key process in regulating peripheral tolerance and can be induced by a number of mechanisms. It is a relatively rapid process occurring at the time of antigen exposure to the T-cell and results in the T-cell becoming functionally inactive. This anergic state can occur by several mechanisms including insufficient co-stimulation at the time of antigen presentation, binding in the presence of immunosuppressive cytokines (e.g. TGF-β, IL-10), the presence of immunosuppressive immune cells (e.g. MDSCs, TAMs) or the activity of co-inhibitory signals (e.g. PD-1, CTLA4).[215, 216] This is particularly relevant within the context of cancer as tumour cells typically act as poor APCs, expressing few co-stimulatory factors and upregulate immune checkpoint expression and indeed tumours inducing T-cell anergy has been previously demonstrated.[217, 218] T-cells in an anergic state are still capable of becoming effector T-cells, but this process is likely mediated by various factors such as the presence of cytokines (e.g. IL-2 and IL-15),[219-221] the absence of which may be a barrier to reversing anergy. PD-1/PD-L1 IM therapy may therefore be limited in efficacy in the context of anergic T-cells, though this remains an under characterised area.

Unlike T-cell anergy, which is a rapid process occurring at the earlier stages of tumour development, T-cell exhaustion is a mechanism of T-cell dysfunction that results from long-term exposure to antigens. T-cell exhaustion has been best characterised in CD8+ve cytotoxic T-cells and was first described in chronic viral infections such as hepatitis C and HIV but has now been indicated in playing a critical role within cancers as a major route for lack of immune response within the TME.[214, 222, 223] T-cells are initially activated and gain their cytotoxic and effector functions but, as the insult persists and continuous antigen stimulation remains, they become unable to act as effector cells, gradually losing functions, such as cytotoxic activity, proliferation, cytokine/chemokine production, and, at very late stages, die from apoptosis. Immune checkpoints including PD-1 are upregulated, transcription of genes coding for cytokine receptors are downregulated, genes coding for migratory and adhesive processes are altered and there are profound metabolic changes all resulting in ineffective T-cell function. [214, 223, 224]. T-cell exhaustion is further induced by immunosuppressive immune cells (e.g. MDSCs, Tregs, TAMs), cytokines (e.g. IL-10, TGF-β) and immune checkpoints including PD-1 and CTLA4.[214] T-cell exhaustion as a result of chronic viral infections has been shown to be regulated by PD-1 and IL-10 mediated pathways independently of each other, with the

blockade of either associated with recovery of exhausted T-cells to active effector status[222] and the combined blockade of PD-1 with other immune checkpoints such as TIM3 and LAG3 also associated with recovering exhausted T-cells secondary to chronic viral infections.[225, 226] The same reversal of T-cell exhaustion as a result of PD-1/PD-L1 IM therapy in cancer has not been consistently observed, with various mechanisms including the upregulation of multiple immune checkpoints, the downregulation of CD28, and PD-1 mediated metabolic changes all potential causes for T-cell exhaustion irreversible by PD-1/PD-L1 IM therapy,[227, 228] though it has been demonstrated that blockade of multiple immune checkpoints maybe more effective than IM monotherapy in cancer[229, 230] and may be superior at reversing T-cell exhaustion. [231] There is no specific marker for T-cell exhaustion, although the upregulation of immune checkpoints including PD-1, LAG3, TIM3 and CTLA4 have been shown to be correlated with more severe exhaustion states,[225] which may partially explain the rationale for improved response with dual IM blockade. In any case, the presence of severely exhausted T-cells within the TME maybe a key factor for PD-1/PD-L1 IM resistance even in the context of PD-L1 expression and TIL presence.

1.5.2 Primary resistance – TMB/Neoantigenicity

Tumours that have a higher rate of mutations (tumour mutational burden; TMB) are typically associated with the greatest environmental insult as a precursor to malignancy (Fig 1.5.0), for example malignant melanoma is associated with ultraviolet radiation exposure, and SCC of the lung is associated with smoking.[232] These mutations are essentially random, and contribute to many factors of oncogenesis and ongoing tumour survival and aggression, including driver mutations (*EGFR*/*ALK*), metastases and so forth. The probability of a tumour expressing neoantigens increases with a higher TMB, though this is not a perfect correlation.(Fig 1.5.0) At present there is no direct marker of neoantigen rate, so measurement of TMB maybe used as an imperfect surrogate. The use of this as a biomarker is discussed later, but there is evidence that there is a higher TIL rate of CD8+ T-cells in tumours with higher TMB,[233] and indeed there is some evidence that a higher TMB rate confers superior response to anti-PD-1 treatment in NSCLC[234] with a synergetic predictive ability when combined with PD-L1 status,[235] although more recent data has shown this not to be wholly consistent.[236, 237] However, higher TMB tumours (e.g. melanoma and NSCLC) are generally more immunogenic and generally see the best response rates to immunotherapy, including anti PD-1/PD-L1 IM, whereas the low TMB tumours (e.g. pancreas and prostate cancer) are less immunogenic and generally respond more poorly to immunotherapy.[238] Thus neoantigen levels, or markers of "non-self", are important in both initiating and maintaining an immune response, and the bigger this immune response is, the more profound response to anti PD-1/PD-L1 treatment may be. Whilst NSCLC have high TMBs on average, many will be relatively low (Fig 1.5.0) and as this does not

necessarily translate to high neoantigenicity, any therefore immune response against the tumour maybe muted.

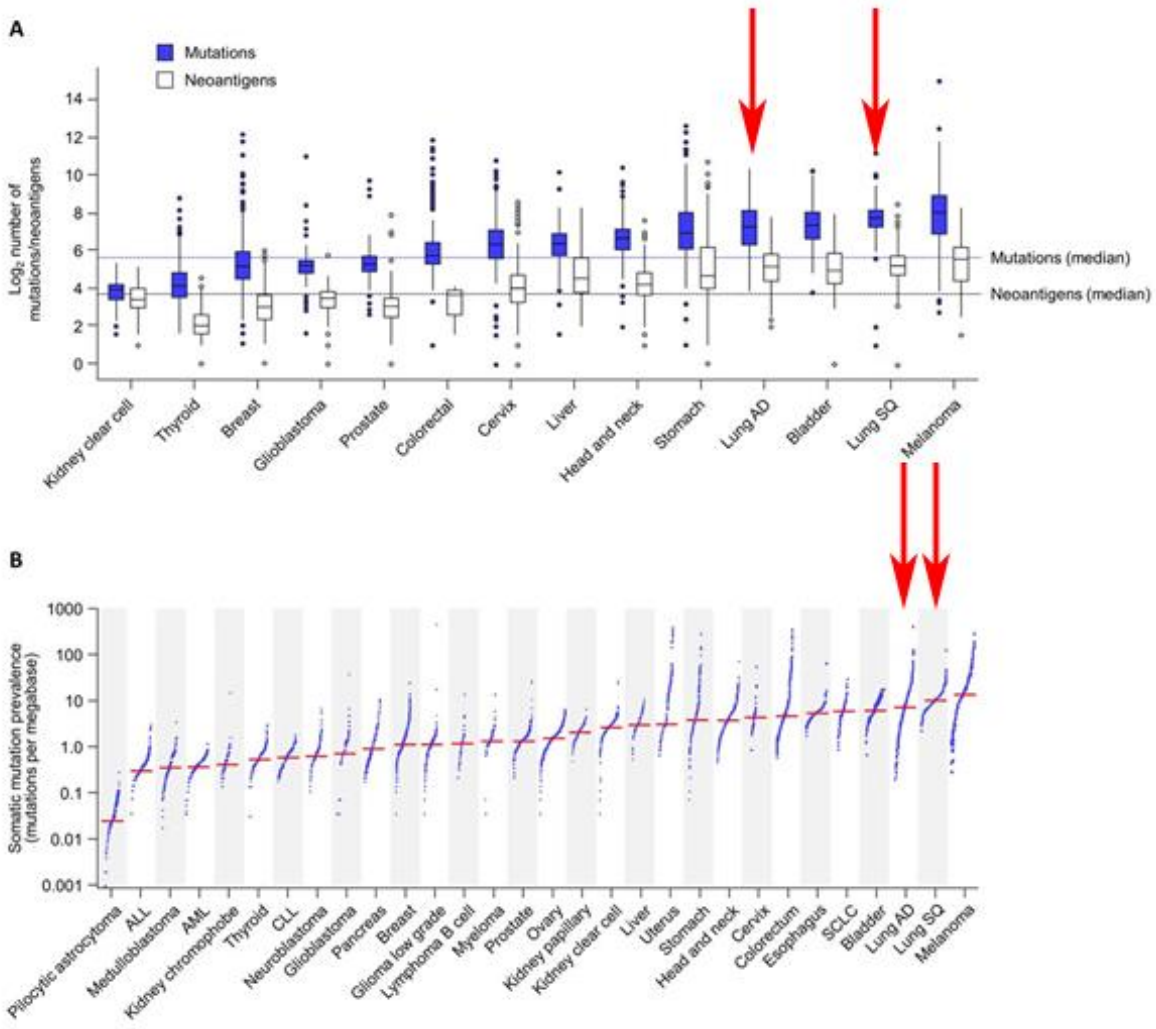


Fig 1.5.0 TMB in solid tumours. **A**- TMB and predicted neoantigen levels for several solid tumour types. A general but imperfect correlation between overall TMB and neoantigens are seen. Lung ADC and SCC are highlighted via red arrows. **B** – TMB variation across 30 tumour classes. Whilst NSCLC (red arrows) is on average at the high end, many cases for NSCLC, particularly for ADC, fall between 0.1 and 1 mutations /mbp, and therefore have a relatively low TMB. Modified from Buttner *et al.* 2018.[266]

1.5.3 Primary resistance – other inhibitory mechanisms

Resistance to IM therapy may come from epigenetic silencing of genes responsible for stimulating anti-tumoural response. For example, silencing of the chemokines CXCL9 and CXCL10 that stimulate Th1 via IFN-G has been implicated as mechanisms of resistance to IM therapy, and success in using

epigenetic silencers in improving T-cell effectiveness, inhibiting tumour progression and enhancing PD-L1 therapy has also been seen.[239] The loss of tumour suppressor genes may also contribute to resistance; for example loss of *PTEN* has been associated with PD-L1 resistance probably through the downregulation of CD8+ve T-cells and upregulation of immunosuppressive cytokines such as VEGF and CCL2.[240, 241] Other immune checkpoint inhibitors may be expressed in TMEs at the same time as PD-L1 such as TIM3, LAG3 and CTLA4.[242] These alternative immune checkpoint molecules are not routinely looked for in clinical specimens, but may be active alternative routes of immune escape by tumour cells. T-cells can express adenosine A2A receptors, that act as immune checkpoint proteins, and a high level of adenosine within hypoxic regions of the TME would activate these and result in another pathway to immune suppression.[243]

Beyond these mechanisms, other pathways of resistance are likely to exist, and within a tumour several immune inhibitory mechanisms that may or may not include the PD-1/PD-L1 axis might be active in parallel. Furthermore, the blockade of PD-1/PD-L1 in the context of multiple immune escape mechanisms may lead to the tumour favouring a different pathway, thus resulting in secondary resistances.

1.5.4 Secondary resistance

Secondary resistance is likely to occur as a result of changes within tumour cells and other cells within the TME, particularly the immune cells. Immunoediting resulting in a change of tumour cell populations may occur by actively downregulating neoantigen expression or by the "natural selection" process of allowing the most immunogenic cells to die, leaving the least immunogenic, and therefore cancer cells least likely to respond to IM therapy, as the dominant cell class within the TME. Direct evidence for this is minimal, but one study demonstrated in NSCLC patients that neoantigen loss in response to combined PD-1 and CTLA-4 IM therapy through both such mechanisms: the elimination of tumour subclones, and the deletion of chromosomal regions that contained truncal alterations.[99] Other mechanisms that have been suggested include the upregulation of other immune-escape mechanisms in place of PD-1, such as TIM3,[244] and the mutation of genes involved in responding to immune mechanisms such as *JAK1/JAK2* mutations that result in a lack of response to T-cell mediated IFN-G release.[245] As PD-L1 and PD-1 can both act to generate immune escape by mechanisms not described in the classic PD-1/PD-L1 axis, primary resistance maybe a predominance of immune escape mediated by these alternative pathways, and secondary resistance caused by an upregulation of their impact.

Most of the mechanisms outlined as potential reasons for resistance come with several questions however: are these mechanisms applicable to all tumour types? Do different immune-checkpoint

inhibitors share some, all or none of these features? Does PD-1/PD-L1 specific resistance have multiple mechanisms of resistance that are seen even within the same tumour?

There are additional unknown questions such as the longer term changes on the immune system's ability to mount responses to tumours caused by PD-1/PD-L1 IM therapy. For example, Pauken *et al.* 2016 found that, although PD-1 blockade could reverse some activity of exhausted T-cells to effector cells, the ability of these T-cells to become T memory cells was lost. This is perhaps particularly relevant in patients with recurring tumours, and it is currently unknown if this would result in resistance to PD-1/PD-L1 IM therapy in the setting of a recurrent tumour after previously successful treatment by IM therapy.[246]

However, whilst immune escape mechanisms outside the classic axis exist, and the possible pathways to resistance are many, the fact remains that many NSCLC patients do respond to PD-1/PD-L1 IM therapy, and this suggests the 'classic' PD-1/PD-L1 axis is a major mechanism of immune escape for many tumours. An important next step is to be able to better predict which tumours predominantly use this method of immune escape, and therefore which patients will best respond to PD-1/PD-L1 therapy.

To date, the only clinically validated predictive biomarker for assessing response to PD-1/PD-L1 IM therapy is PD-L1 expression by immunohistochemistry (IHC). This thesis will explore in greater details many limitations of this approach, but IHC is the *de facto* method of analysing protein expression in the majority of tissue samples used clinically, due to the relatively inexpensive costs, ease and speed of IHC.[247] A brief overview of the strengths and weakness of IHC with particular regards to PD-L1 will be given here.

1.6 Predicting response to PD-1/PD-L1 IM therapy

1.6.0 Immunohistochemistry: a brief overview of a typical pathway.

The use of immunostaining to highlight features of tissue began in the 1930s and was gradually developed throughout the 20th century as a highly effective, specific and relatively cheap tool now used in clinical pathology laboratories globally.[247, 248] IHC is the practice of using antibodies that bind to a specific epitope region of a protein of interest (primary antibodies) combined with a dye that will allow visualisation of this event, often another antibody conjugated to a chromophore (secondary antibody). A typical IHC workflow is presented here: tissue selected for IHC is prepared by producing FFPE (formalin fixed paraffin embedded) samples in which tissue is 'fixed' within formalin (a formaldehyde based fixative, typically '10% neutral-buffered formalin' (NBF) that contains approximately 4% formaldehyde.), a process that prevents the degradation of tissue by

decomposition or autolysis, and preserves the cellular architecture and composition of tissue, and then dehydrated before the addition of paraffin, which embeds the tissue into a paraffin wax block. FFPE blocks can then be sectioned, in which a microtome is used to produce thin (usually 2-4μm) slices of tissue that can be floated onto water before being placed on glass slides. (More detail is given in Chapter 3 (Materials and Methods) and Chapter 4 (The effect of pre-analytical conditions on PD-L1 expression)). A by-product of formalin fixation is cross-linking, in which nucleic acid based or amino-acid based intra-cellular components are bound together, resulting in the formation of DNA and/or protein crosslinks, which can result in 'epitope masking'; a process that results in antibodies unable to correctly bind to their respective proteins. More detail is given in Chapter 4, but 'antigen retrieval' (AR) is the process of breaking crosslinks so that epitope regions are no longer masked, usually by heat or enzymatic methods. Next, blocking of proteins within the tissue may be required to prevent non-specific binding of the antibodies (and therefore non-specific staining) which is particularly important if using the avidin-biotin complexes (ABCs) frequently used for IHC assays, as biotin is a frequently found in many human tissues, although alternative approaches, such as the Roche Ventana Optiview (used for the PD-L1 clones SP263 and SP142) uses a nonendogenous hapten based approach that minimises this issue and reduces non-specific staining.[249] Blocking of endogenous enzymes may also be required, for example endogenous peroxidase can interfere with assays using horseradish peroxidase (HRP)-conjugated antibodies and will need to be blocked (e.g. SP263 uses 3% hydrogen peroxide). The next step involves selecting a primary antibody to bind to the protein of interest. For PD-L1, many primary antibodies used for IHC are available, including SP263, SP142, 22C3, 73-10, 28-8, E1L3N and others, each with distinct epitope regions and staining patterns, the focus of which is discussed in more detail in Chapter 2 (Literature review). Generally these are monoclonal rabbit antibodies which means they are highly specific, prone to less variation and highly reproducible, but requires adherence to a stricter set of assay protocols and are more sensitive to artefactual loss of staining than polyclonal antibodies. The secondary antibody, that binds to the primary antibody and provides the visualisation component must also be selected, with the commonest secondary antibody being the chromophore DAB (3, 3'-diaminobenzidine) which binds to HRP to produce a brown stain easily distinguishable from the blue of haematoxylin counterstain on microscopy (Fig 1.6.0). HRP/DAB is stable, permanent, resistant to many environmental variables and can be visualised under a standard brightfield microscope, although it can produce background staining even with blocking steps in place and is not an inherently easy to quantify stain.

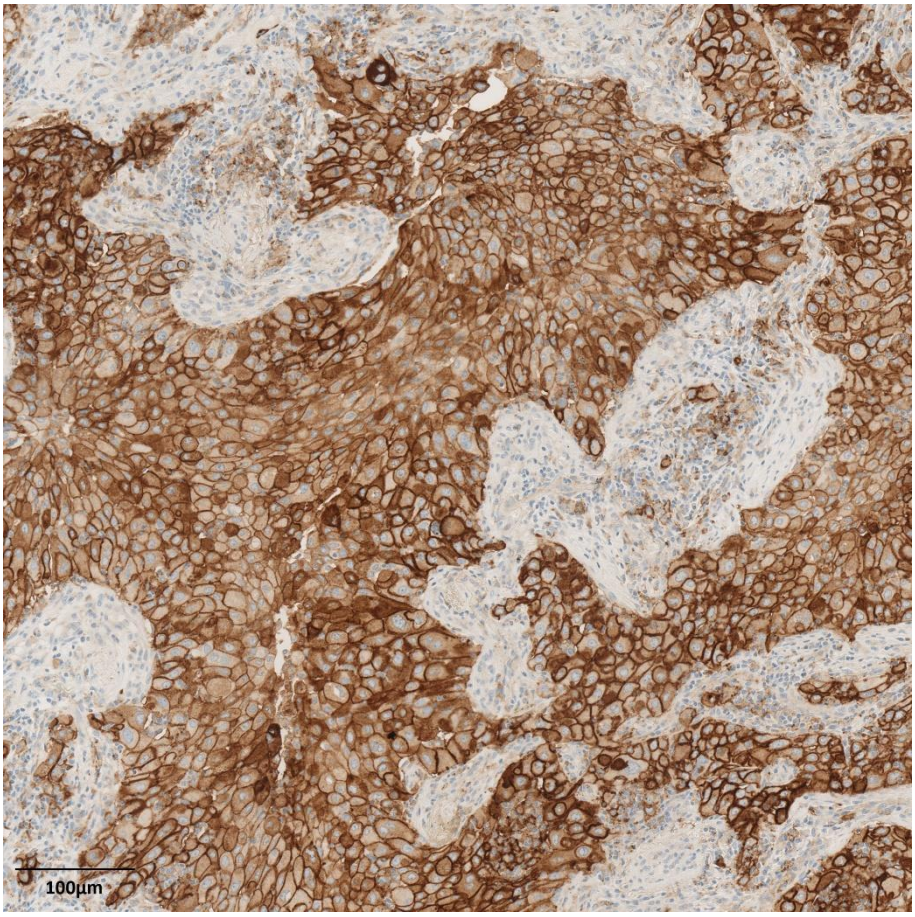


Fig 1.6.0 A NSCLC immunolabelled for PD-L1 (SP263) showing its typical location on tumour cell membranes by the brown chromophore DAB. There is weaker cytoplasmic expression, which is ignored in terms of scoring if unaccompanied by membranous staining. Some immune cells are also stained. The blue counterstain of haematoxylin allows for the cellular component of the DAB staining to be identified.

Several steps can be taken at this stage in order to amplify the signal from the secondary antibody, which can improve the visualisation of the antibody but makes it harder to standardise quantification, particularly as this is not always performed and can vary between assays (e.g. SP263 does not require amplification, but SP142 does). Finally, a counterstain is applied to the tissue, (for HRP/DAB this is typically haematoxylin), so that the specific cellular location of the DAB staining can be identified. 'Controls' are pieces of tissue with known expression levels that are required to determine if a successful reaction has occurred: a positive control will provide evidence the assay has been successfully completed, and a negative control that there is not excess background and non-specific staining, as well as to allow for the comparison of different sections stained in different runs to ensure consistency of the assay. An example of PD-L1 IHC is shown in Fig 1.6.0.

The practice of using IHC to detect PD-L1 means it can be applied in the majority of clinical labs worldwide, although variation between platforms and antibodies is an area of significant concern. In addition, the nature of the tightly regulated IVD (*in-vitro* diagnostic) assays required for PD-L1 IHC means individual labs are less able to make changes to the protocol. Whilst this can ensure consistency, it also means tissue is susceptible to pre-analytical variation that can have a significant impact on the final result. The assessment of PD-L1 IHC is discussed in more detail in Chapter 3, but it is important to note that unlike many stains, the presence or absence of DAB staining in the correct cellular components is not sufficient to produce a clinical report; the quantity of staining is also necessary, and pre-analytical and analytical variation are key factors that can dramatically alter the quantification of PD-L1 expression. IHC is not an assay inherently designed for quantification, and as has been discussed, differences in assays, including AR and amplification can dramatically alter chromophore levels, rendering comparisons between clones challenging. This is discussed more fully in Chapter 2 and Chapter 5 (Analytics and post-analytics: Digital pathology and PD-L1 interpretation).

Other techniques for assessing PD-L1 expression levels include immunofluorescence (the use of a fluorophore instead of a chromophore), multiplex IHC and mass-spectrometry, all of which may be able to better quantify PD-L1, either by inherent advantages of the technique or though use with digital image analysis, as well as to provide simultaneous details about other relevant proteins from the same piece of tissue. Some of these (non-clinically validated) techniques are discussed later in this thesis.

1.6.1 PD-L1 IHC to guide IM therapy

Many of the indications for the use of PD-1/PD-L1 IMs is based on the predictive power of PD-L1 expression by IHC. The complex nature of PD-L1 IHC is illustrated by the myriad ways in which a tumour is defined as 'PD-L1 positive', its inconsistent predictive behaviour, and the wide variety of assays and clones that can be used, hence PD-L1 IHC's variable status as a companion or complementary diagnostic. (Table 1.6.0). PD-L1 expression levels may rely only on scoring PD-L1 positive tumour cells (e.g. to generate a *tumour proportion score* (TPS)) or both tumour cells and immune cells together (to generate a *combined positive score* (CPS)) or the tumour cell and immune cell components are scored separately (such as TC/IC scoring for SP142). Furthermore, the percentage cut-off for any of these to determine positive or strong positive tumours varies between clones and tumours (Table 1.6.0, [11, 250-253]). For NSCLC, however, a TPS is generally used (with the exception of the SP142 clone), and throughout this thesis a percentage TPS will be given to describe PD-L1 expression unless otherwise specified.

Clone	Cancer with associated PD-L1 scoring level				
	NSCLC	***Gastric***	***Cervical***	***Urothelial***	***HNSCC***
22C3	≥1% ≥50% TPS	≥1% CPS	≥1% CPS	≥10% CPS	≥1% CPS
	Urothelial	***Breast***	***NSCLC***		
SP142	≥5% IC	≥1% IC	≥50% TC or ≥10% IC; ≥1% TC or ≥1% IC		
	NSCLC	***Urothelial***			
SP263	≥1% ≥25% TPS	≥25% TC; ICP >1% + IC ≥25%; ICP = 1% + IC = 100%			
	NS-NSCLC	***HNSCC***	***Urothelial***	***Melanoma***	
28-8	≥1% ≥5% ≥10% TPS	≥1% TPS	≥1% TPS	≥1% ≥5% TPS	

Table 1.6.0 PD-L1 IHC clone, scoring approach and 'cut-off' for classifying 'PD-L1 positive' by tumour type according to FDA *in-vitro* diagnostics approvals.

A brief overview of some of the current approvals for PD-1/PD-L1 IM therapy in NSCLC and their relationship to PD-L1 IHC illustrates the complexity and variability of the situation: indications that require the use of PD-L1 IHC as a companion diagnostic (in which a specific assay and result is required for prescription of a specific therapy in a specific clinical context) in NSCLC include pembrolizumab for 1st line metastatic treatment (TPS ≥50%)[48, 254] or in 2nd/3rd line metastatic treatment (TPS ≥1%).[255] Use of PD-L1 IHC testing as a complementary diagnostic (in which an assay can be used to help guide therapeutic decisions) includes nivolumab for 2nd/3rd line metastatic non-squamous NSCLC (TPS ≥1%)[47] or to guide a number of other PD-1/PD-L1 IM indications even if increased benefit for PD-L1 positive patients is not seen, such as pembrolizumab as 1st line therapy combined with XCT in metastatic NSCLC,[256, 257] nivolumab as 1st line therapy combined with ipilimumab in stage 4/recurrent NSCLC[258] or 2nd/3rd line metastatic squamous NSCLC,[46] atezolizumab in metastatic NSCLC as 1st line combination therapy with XCT[259] or bevacizumab,[213] or as 2nd/3rd line metastatic NSCLC monotherapy.[49, 260] Therefore PD-L1 IHC as a companion diagnostic has a relatively small number of uses, whereas PD-L1 IHC as a complementary diagnostic has a broader array of applications. This is a constantly developing field both in NSCLC and in other tumour types, but the use of PD-L1 IHC in the clinical context is now firmly established, with new indications a frequent occurrence.

1.6.2 Stratified medicine and PD-L1 as a biomarker

Stratified medicine, targeted therapies and predictive biomarkers have revolutionised the treatment landscape of many diseases, including advanced NSCLC. The process is far from straightforward, however. In an ideal situation a specific therapeutic agent would be developed alongside a predictive biomarker designed to be used as a companion diagnostic and these would then be approved at the same time, such as when PD-L1 IHC (22C3 clone) was approved as a companion diagnostic for prescribing pembrolizumab for patients with advanced NSCLC in 2015.[254] This gave a clear remit for the use of a PD-L1 IHC assay and prescription of a specific PD-1 IM therapy. Drug development and predictive assays are not always so clearly linked; for example, in the same year, the 28-8 clone was also approved as a complementary diagnostic for nivolumab,[47] and atezolizumab was subsequently approved for use regardless of PD-L1 expression levels.[49] Hence the current situation is one in that PD-L1 positivity is either required, used as a guide, or not required at all, for prescribing PD-1/PD-L1 IMs, depending on specific circumstances.

It is to be expected that the approach to designing and using companion and complementary diagnostics focusing on a specific target will change over time however. For example, the well characterised use of scoring HER2 (human epidermal growth factor receptor 2) for prescribing trastuzumab in breast cancer was initially based on expression by IHC, but companion diagnostics for this indication have diversified and now include ISH (*in-situ* hybridisation) and broad NGS (next-generation sequencing) based panels.[261, 262] Change in the use of PD-L1 IHC has also been seen as clinical trial design has evolved: in the early clinical trials for PD-1 inhibitors, PD-L1 expression by IHC either wasn't performed, was only performed retrospectively and/or was only performed on a subcohort of the treated patients.[46, 263-265] Contemporary trials generally all use PD-L1 IHC to assess PD1/PD-L1 IMs, however, and in clinical practice it is recommended to perform PD-L1 IHC as a reflex test in several tumour types, including NSCLC,[266] indicating a shift away from testing for PD-L1 only in specific circumstances, and a general acceptance of testing all NSCLC patients.

Despite this, the use of PD-L1 IHC to predict treatment is not a simple or consistent situation. In the Checkmate 057 trial an association between PD-L1 expression and favourable outcome was noted for previously treated patients with advanced non-squamous NSCLC treated with nivolumab,[47] using a cut-off of 1% TPS (with increasing responses using ≥5% and ≥10% cut-offs). The Keynote-010 trial treated advanced NSCLC in the second line setting and recruited only patients of ≥1% TPS, and found them to have a superior PFS (progression free survival) with pembrolizumab[255] and in the Keynote-024 trial patients with advanced NSCLC whom had a PD-L1 TPS of ≥50% had superior PFS when treated with pembrolizumab in the first line setting.[48] This initial data seemed to point in the

direction that PD-L1 positivity was therefore a useful predictive biomarker for PD-1 IM therapy, and higher levels of expression seemed to convey a more favourable outcome. Even for the Checkmate-017 trial, which saw benefit in advanced squamous-NSCLC treated with nivolumab but saw no stratification by PD-L1 expression, it is fair to point out not all patients had tissue available to test, and what was tested was a mixture of fresh and archived tissue, and so was possibly an outlier in terms of the benefits of PD-L1 expression.[46]

Unfortunately, not all trials provided such clear data for the benefit of stratifying patients by PD-L1 expression: Checkmate 026 looked at treating advanced NSCLC patients with ≥1% PD-L1 expression with nivolumab in the first line setting but failed to reach its target outcomes.[267] This failure was blamed on various reasons, including a cohort more likely to respond favourably to XCT (a subsequent statement paper cited reasons including that the cohort generally had small tumour burdens and was composed of a high proportion of women) and that there was a relatively low proportion of patients with high PD-L1 expression (≥50%).[268] The OAK trial found benefit to all advanced NSCLC patients treated with atezolizumab in the second line setting, and whilst higher PD-L1 expression was associated with superior response, benefit over XCT was still significantly better, albeit to a lesser degree, for the PD-L1 negative cohort.[49] (Fig 1.6.1)

Taken together, these results seem initially to be incompatible: several studies show the benefit of PD-1/PD-L1 IM therapy only in PD-L1 expressers, but other studies show the benefit is there for all patients regardless of PD-L1 status. However, several factors could explain these apparent discrepancies, including the unknown variation of efficacy between the PD-1/PD-L1 IMs or the different patient cohorts between studies. Other variations between clinical trials includes the use of different PD-LC IHC clones to study expression, the use of different 'cut-off's to define positivity, the use of different specimen types between and within studies and unknown variables such as variation in pre-analytical factors.

The difference between predictive biomarkers that share the same target has been shown to be clinically relevant in other settings with known variation in detection rates between technologies even for matched patient samples. For example there maybe variation between matched samples for *ALK* translocation by detection by IHC or FISH.[269]

PD-L1 IHC adds further complexity by the sheer number of IHC clones available, the different platforms the assays can be performed on, the differences in interpretation, and the heterogeneity of tumours. It is also interesting to note that for many of these studies there are a small number of 'rapid-progressors' in the IM treated arm, with survival curves showing an initial cross over, in which at the earliest stage XCT patients perform better than IM therapy, implying that, for some patients,

IM therapy will produce worse outcomes, even if PD-L1 positive, even in the context of a cohort that overall has benefit from IM therapy.[49, 258, 270] (Fig 1.6.2). This suggests the expression of PD-L1 alone is not a discriminatory feature for predicting rapid progressors and illustrates a lack of understanding in how best to classify PD-L1 expression.

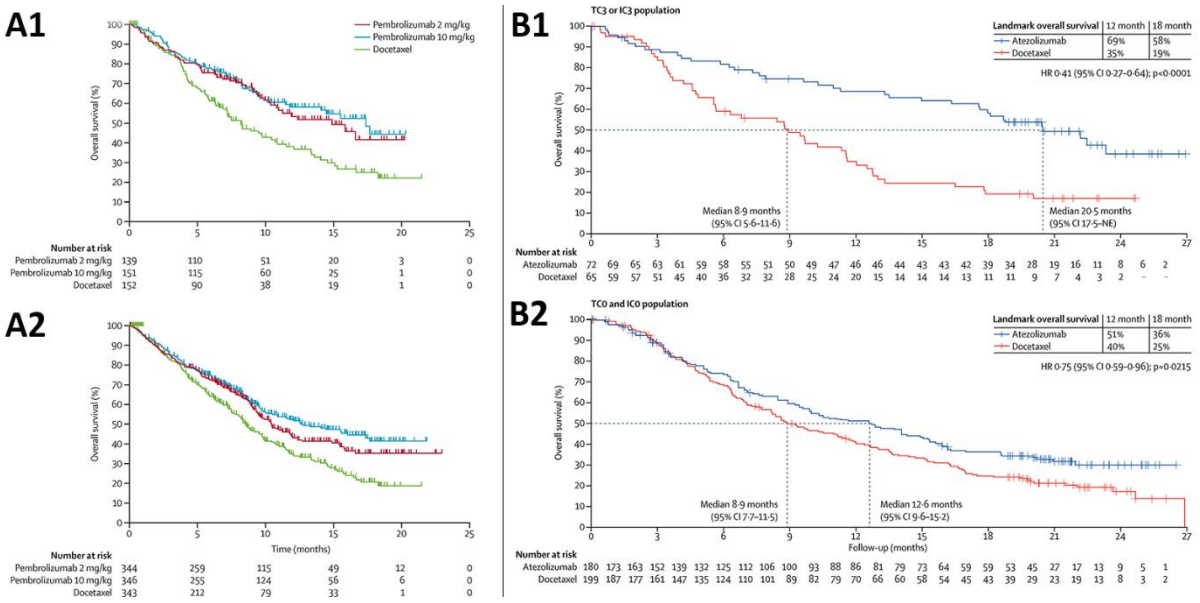


Fig 1.6.1 Survival analysis curves showing overall survival (OS) in two seminal PD-1/PD-L1 IM trials **A** – Keynote 010 **B** – OAK. **A1** + **B1** show high PD-L1 expressers with significant benefit from IM therapy over XCT. **A2** shows that all comers (≥1% TPS) and **B2** shows that PD-L1 negative patients both still benefit from IM therapy over XCT, albeit to a lesser degree than high PD-L1 expressers. Modified from Herbst *et al.* 2016[255] and Rittmeyer *et al.*2017.[49]

The results of clinical trial data summarised in Fig 1.6.1 and Fig 1.6.2 illustrates that PD-L1 IHC provides important biological information, but does not translate easily to consistent clinical guidance. The optimisation and maximal accuracy of PD-L1 IHC to achieve consistent results is the first step to improving prediction of response to PD-1/PD-L1 IM therapy. In so doing, not only will better treatment guidance be available, but effective stratification of patients reduces waste, minimises unnecessary toxic exposure and improves cost effectiveness of treatment. The second step is a consideration of which factors within the TME could prove to be powerful additional biomarkers to use in conjunction with PD-L1 IHC. It is clear the complexity of the TME means that any single cell type, cytokine or immune marker, including PD-L1, is not necessarily consistently associated with pro-tumoural or anti-tumoural activity, and therefore no single marker will be a reliable predictor of response to IM therapy.

In order to do so, this thesis will first consider what is already known about PD-L1 IHC and the potential reasons for its variable performance as a predictive biomarker by performing a comprehensive literature review on this topic.

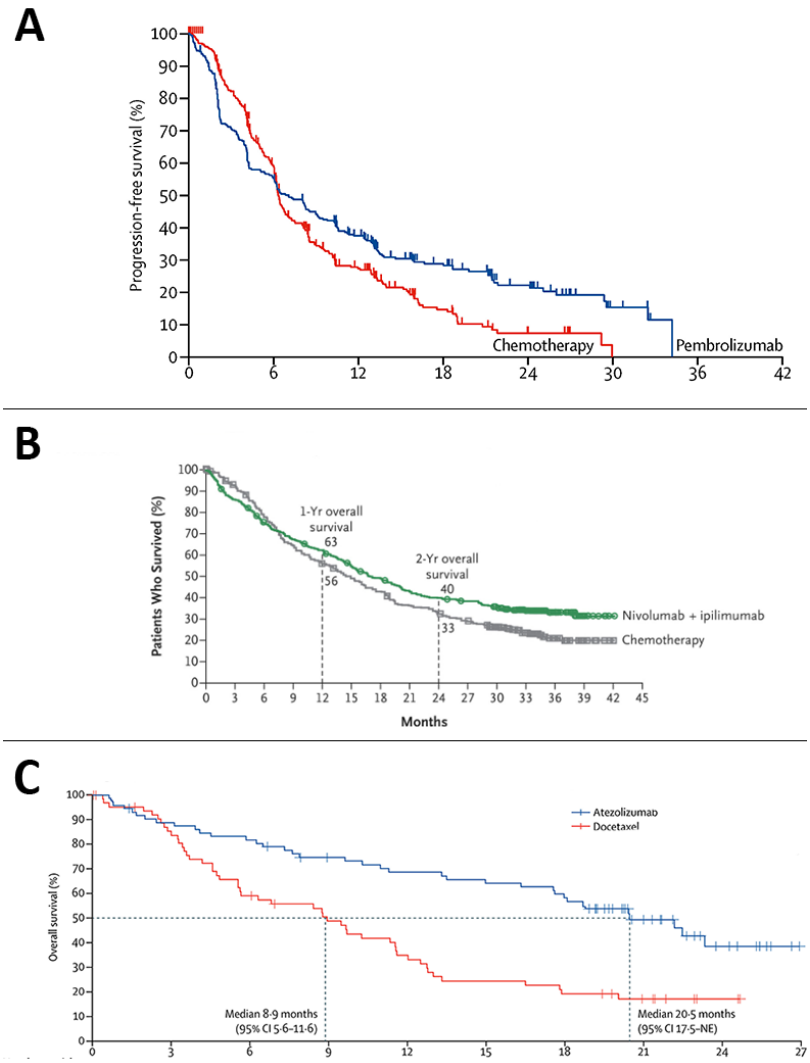


Fig 1.6.2 Survival analysis curves showing overall survival (OS) in three trials **A** - Keynote 042 **B** - Checkmate 227 **C** – OAK. All three trials show PD-L1 positive patients have superior OS when treated with IM therapy over XCT, but also show an early-phase 'cross-over' of survival curves in which IM therapy conveys worse survival for a subcohort of PD-L1 positive patients.

Chapter 2 – Literature Review

The purpose of this literature review is to summarise the current knowledge base on the limitations of PD-L1 expression by IHC as a predictive biomarker for PD-1/PD-L1 IM therapy, with a focus on its use in NSCLC specifically. The data and findings in other tumour types may yield additional interesting information, and so relevant papers have been included as appropriate. I have split the literature review into different sections that reflects how this thesis has approached the problem of overcoming the limitations of PD-L1 expression by IHC as a predictive biomarker. Firstly, the impact of pre-analytics is considered: if handling or usage of tissue results in a false change of PD-L1 expression, and thus is not a true reflection of the genuine biology occurring within the selected sample. Secondly, the analytics and interpretation of PD-L1 IHC is explored: if changes in PD-L1 expression vary as a result of using different antibody clones or assays to detect PD-L1, and what the variation between and within the pathologists performing the interpretation is. Finally, heterogeneity is looked at: what is the variation of PD-L1 expression within a sample, between samples, and over time? By focusing on these three areas to ascertain our knowledge of when variation is artefactual, when "variation" is related to interpretation, and when variation is genuine, this literature review will highlight areas of general consensus and areas where our current knowledge is lacking.

The literature review was initially performed in September 2017, but this rapidly developing area has led to a significant increase in the general body of data available, and reflects our increasing knowledge in this area. Thus a repeated literature review was performed in August 2019. The search process and terms are shown in Fig 2.1.0 and Table 2.1.0 respectively, but to illustrate this point, only 61 papers were included in the initial search, with over double that number included in the final review. The nature of this rapidly developing field is such that simply searching for "PD-L1" in SCOPUS returns 507 hits for a ten year period of 2000 to 2010, 1,369 hits in a five year period of 2011 to 2015, but over 3,000 hits in 2019 alone. Therefore this literature review accepts it may not contain the most recent of developments, but for the large part accurately reflects the overall body of knowledge pertaining to the use of PD-L1 expression as a biomarker, and, in some cases, highlights where a general consensus has been agreed, and where areas of contention still require debate.

2.0 Initial findings

Two databases, SCOPUS and Pubmed were searched with the term "PD-L1" using Boolean operators and 21 additional terms to generate 18 unique searches on the 2nd August 2019 (Table 2.1.0). The intent was to include original research articles only. A total of 2,474 hits were returned, with 149

articles finally selected for the in-depth literature review. Reviews, abstract/poster only returns, articles in non-English and articles dealing with non-NSCLC were excluded, although 28 non-NSCLC papers of particular interest, seminal findings or when there are no equivalent NSCLC papers were included. This process has been summarised in Fig 2.1.0.

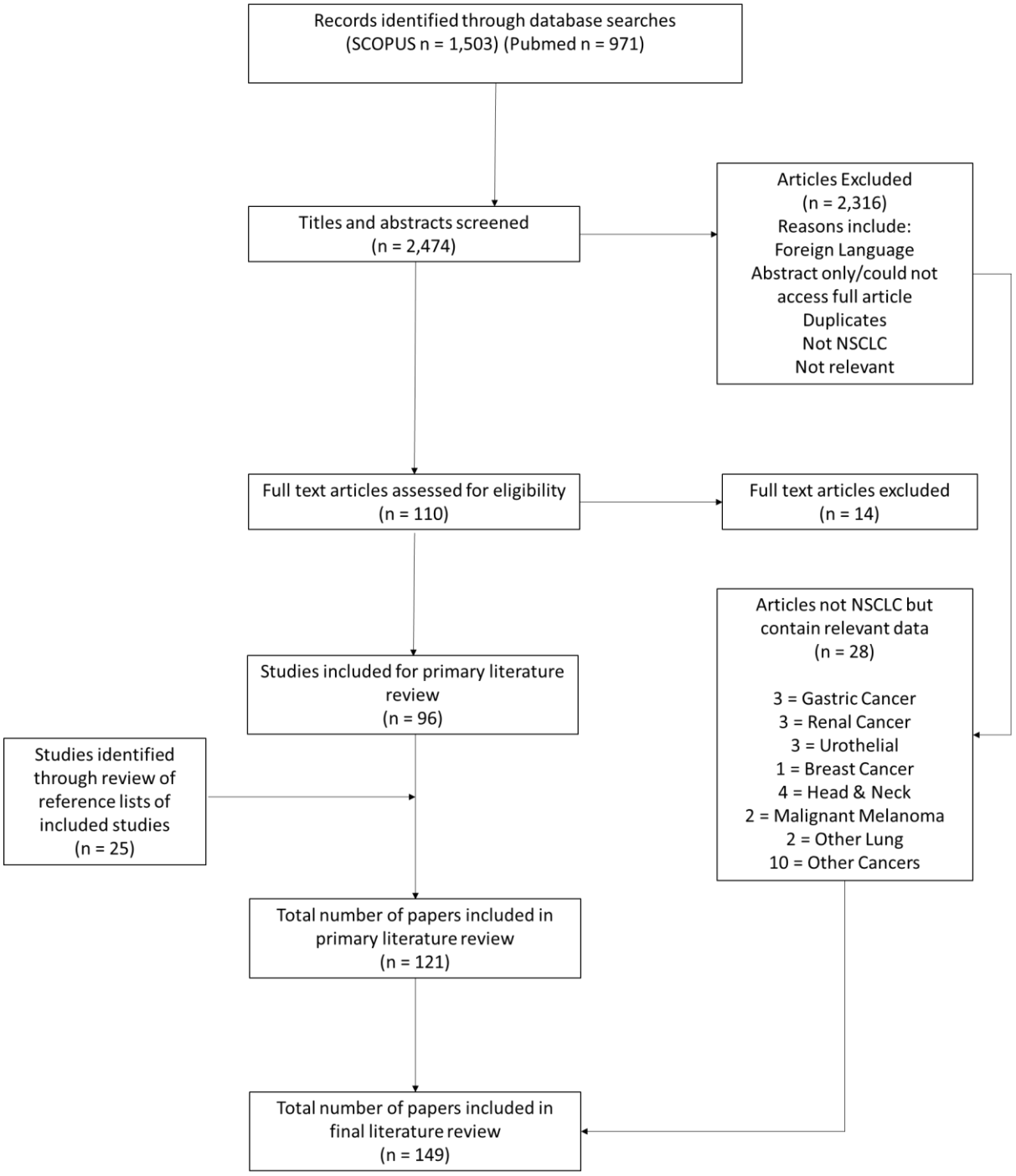


Fig 2.1.0 Flow chart of the literature review process illustrating total number of return articles from initial search and rationale for inclusion/exclusion of papers.

Search Term*	Results	Excluded	Included
"PD-L1" AND:			
"pre-analytical"	6	3	3
"fixation"	15	8	7
"specimen age"	78	66	12
"sample age"	85	4	8
"storage"	6	4	2
"Decalcification"	1	0	1
"EDTA"	4	3	1
"cold" AND "Ischaemia" AND "time"**	1	0	1
"PD-L1" AND:			
"IHC comparison"	125	75	50
"Clones comparison"	29	24	5
"Post-translational modification"	7	4	3
"Immunohistochemistry" AND "Soluble"	23	15	8
"PD-L1" AND:			
"Heterogeneity"	206	158	48
"Concordance"	117	69	48
"Paired"	99	88	11
"Cytology"	264	241	23
"EBUS"	17	5	2
"PD-L1" AND "XCT" AND "immunohistochemistry" AND "expression" AND "lung"	378	368	10
"PD-L1" AND "Radiotherapy" AND "immunohistochemistry" AND "expression" AND "lung"	128	122	6
		Total (with duplicates)	249
		Grand total	149

Table 2.1.0 Results for articles by search term with numbers included and excluded. Grand total removes duplicate papers that appeared in multiple searches.

*These are for initial search returns and therefore some articles appear in multiple searches.
**American and British spelling terms are both used where appropriate; for example ischaemia and ischemia were both searched for in this context.
EBUS: endobronchial ultrasound; EDTA: Ethylenediaminetetraacetic acid

2.1 Pre-analytics

7 unique terms were used for the pre-analytical part of the literature review (Table 2.1.0). It is interesting to note the IASLC (International association for the study of lung cancer) guidelines for pre-analytical conditions include suggestions for 10 parameters, with a chief focus on fixation and

storage/specimen age. Additional parameters include cold ischaemia time and decalcification, though there is little hard data to evidence these suggestions in the context of PD-L1 specifically; indeed searching for these terms with PD-L1 return very small numbers of relevant articles (Table 2.1.0). This reflects a general issue with PD-L1 IHC in that best practice for IHC has been applied from knowledge from generic immunochemistry data, but it is only with hindsight that the particularly delicate nature of PD-L1 IHC, and the uniquely critical nature of the quantification of PD-L1 expression suggests that there are gaps in our knowledge about when specimens should be considered inadequate or inappropriate due to pre-analytical conditions. Manufacturers of the PD-L1 clones have been clear to make the point that they have not been validated under certain conditions, and there is a general lack of information of the effect of certain pre-analytical parameters on PD-L1 expression. None the less, PD-L1 analysis is performed on tissues that, in a global context, have been subjected to a massive variation in pre-analytical conditions, and the clinical implications of this is generally unknown.

Most of the focus and data for pre-analytical conditions effecting PD-L1 expression, and questions surrounding the use of specific samples, has been around the age of the sample, and questions pertaining to fixation, and these are therefore the chief factors considered in this part of the literature review.

2.1.0 Pre-analytics - Fixation

Possibly the earliest paper that mentions the potential effect of fixation on PD-L1 expression is Thompson *et al.* 2005,[271] whom compared a cohort of FFPE renal cell carcinomas stained for PD-L1 (5H1 clone) with previous work performed on fresh-frozen specimens of renal cell carcinomas in 2004 and 2005. In their fresh-frozen cohorts, they found 37% of patients were positive for PD-L1, but only 24% in the FFPE cohort, with the cross-linking properties of formalin blamed for this discrepancy. It is also possible that differences in patient population or problems pertaining to heterogeneity or the 5H1 clone could be partially to blame, but it was not until major work in the NSCLC area came to light that this area was majorly addressed again in the literature. Manufacturers' manuals for the clinically validated clones Dako-Agilent 22c3 and 28-8 and Roche-Ventana SP263 and SP142 all explicitly state they are for use in FFPE tissue, and typically make no mention of either fresh-frozen tissue or other fixative methods.[250-253] Fresh-frozen is not typical for routine clinical tissue samples, and the effects of using formalin on PD-L1 expression has been considered in subsequent papers. Differences between histology and cytology specimens is considered in more depth later in this review, but the question of the use of cytology specimens is

key due to the prevalence of non-formalin based fixatives used in cytology specimens, though again this is rarely touched upon specifically in the literature.

Rebelatto *et al.* 2016 discuss an optimised approach to the specific use of SP263 clone in patients samples of NSCLC and HNSCC (Head & Neck squamous cell carcinoma) as well as cell line ADCs.[272] In addition to other findings, they recommend that alcohol based fixatives, such as 95% alcohol and AFA (Alcohol Formalin Acetic acid), as well as Glyoxal based Prefer fixative, should be avoided as they cause a loss of PD-L1 staining intensity, but 10% NBF, zinc formalin and Z5 (a zinc based fixative) do not. It should be noted, however, that although this paper is a comprehensive overview of the effects of pre-analytical parameters that may affect PD-L1 expression, the verification studies in this study used a single-cut off point of 25% to divide samples into either "high" or "low/negative", (in keeping with the Phase 1 clinical trials using SP263 at the time [243].) However, this single cut-off is problematic in that significant differences in PD-L1 expression, including changes around the more frequently used clinical cut-offs of 1% and 50% will be missed in assessing the effect of pre-analytical variation. Indeed this is a common limiting factor in studies that compare PD-L1 expression; very typically only one or two cut-offs are used to determine if changes are present, without the recording of specific percentage changes, and therefore large changes either side of the 'cut-off' are not included (for example, a change from 30% to 100% may not be recorded if a ≥25% 'cut-off' threshold is used) and small percentage changes around the cut-off (for example, a change from 20% to 25%) may falsely emphasise the impact of certain parameters, particularly as smaller changes maybe more susceptible to variation from inter-observer interpretation or tumour heterogeneity. Notwithstanding this limitation, the fact that non-formalin based fixatives appear to decrease PD-L1 expression, is an interesting finding, and one that appears to confirm the validity of formalin as the *de facto* fixative of choice.

Wang *et al.* 2018 looked at four separate fixative methods in 1,419 NSCLC cytology, small biopsy and surgical specimens. 10% NBF, Cytolyt (Methanol based fixative), alcohol followed by NBF, or TissuFix (formaldehyde and alcohol mix) were used and the specimens were scored for PD-L1 TPS using the 22C3 clone, categorising samples into negative, weakly positive or strongly positive (using <1%, 1-49% and ≥50% TPS cut-offs respectively).[273] Overall they found no difference in PD-L1 expression between the different fixatives types. They note that in previous studies the effect of methanol on reducing immunohistochemical staining intensity was largely limited to nuclear stains (ER, TTF-1 etc.) and largely unaffected in membranous stains, of which the relevant scoring of PD-L1 is an example of. Their results echo this, in that membranous PD-L1 expression is unaffected. They conclude that 10% NBF remains their fixative of choice.

Jain *et al.* 2018 compared fixatives via matched cytology specimens and small biopsies in 26 NSCLC patients; using 10% NBF for the small biopsies and CytoRich Red (Formalin and alcohol mix fixative) for the cytology specimens. PD-L1 expression was assessed using the SP263 clone, and the same protocol for histology and cytology specimens was used.[274] Specimens were split into either "high" or "low/negative" using a 25% TPS cut-off point. 23 of 26 cases (88.4%) were concordant for PD-L1 expression, regardless of the fixative used. Although there was some concern around false negatives and aberrant nuclear staining, the overall conclusion was that alcohol-based fixatives are generally suitable for using in tissues for PD-L1 analysis.

Lloyd *et al.* 2019 provide a comprehensive comparison of different fixatives' effect on PD-L1 expression by comparing 5 different preservative/fixative methods in human ADC cell lines of known high PD-L1 expressing phenotype.[275] ThinPrep PreservCyt, ThinPrep CytoLyt (both methanol-based), normal saline, RPMI (Roswell Park Memorial Institute) medium (both preservatives rather than fixatives) and 10% NBF were used to fix the cells before cell block preparation and staining for PD-L1 with the 28-8 clone. Lloyd's group used a slightly more involved scoring process than is usual, in which cells are placed into a category (1-5) that account for location and completeness of staining as well as cell viability, with each individual cell within a sample counted and placed into one of these 5 categories by three independent pathologists. Some changes in cell viability were noted (e.g. CytoLyt had significantly more non-viable cells than formalin). No false positives were noted in the negative cell line, regardless of fixative used, and the majority of cells in the known strongly expressing cell lines demonstrated some form of PD-L1 expression, though this did include cytoplasmic or other staining patterns in some instances (96.3-99.8% of cells). Interestingly inter-pathologist concordance had different variations depending on fixative type – for example weak intensity scoring was most disparate in PreservCyt, but the largest discrepancy in 10% NBF and CytoLyt was complete membranous staining. PreservCyt demonstrated the largest amount of weak staining cells, (34.2% vs 16.1%). The overall conclusion was that different fixatives can affect morphology and staining quality, and thus PD-L1 interpretation should be adjusted in view of this. They conclude formalin based fixatives are acceptable, but CytoLyt preparations should probably be avoided. In addition, though the cell blocks were generally created using the plasma-thromboplastin method, a duplicate group of specimens were fixed with PreservCyt using the Cellient cell block preparation, which in their hands provided the most consistent scoring between pathologists, provided the best morphology and most amount of membranous staining, leading them to suggest this may be the preferred option in cytology samples. However, this platform is not practically available in all labs, and these results would ideally need to be duplicated using patient samples, rather than cell lines, and with other PD-L1 clones.

Taken together these papers suggest that fixative types can have an effect on the expression of PD-L1, but that the industry standard of formalin based fixatives, in keeping with the manufacturers' instructions, is probably the best way to ensure consistent staining, even in the context of cytology based samples. However, they also demonstrate a recurring theme in wider attempts to answer questions around what effects PD-L1 expression: the clones, types of specimen and scoring methods all vary between studies; which means making firm conclusions based on these various studies is problematic. It also demonstrates the relatively small numbers of studies attempting to answer these questions.

The question of the effect of fixation timings has also been addressed by a small number of studies. Forest *et al.* 2019 looked at a small number of NSCLC samples (excluding negative and heterogeneously expressing samples) as well as placental tissue.[276] Tissues samples post-surgery (cold ischaemia <15 minutes) were placed into 10% NBF at 6 different time points, from immediate to 24 hours delay. A total of 5 NSCLCs were included, and the 22C3 and EL13N clone were used to assess PD-L1 expression, with the percentage of placental villi scored, and for the NSCLC samples the percentage of positive tumour cells given, as well as the intensities of staining. In this modestly sized study, with samples selected for homogenous expression, there was no difference seen in delay of fixation.

Kawachi *et al.* 2019 addressed the question of fixation duration in 10% NBF, and looked at 55 NSCLC samples fixed for either 12-24 hours or 96-120 hours, and then stained for PD-L1 using both 22C3 and SP263 clones.[277] This group compared the mean TPS between the fixative groups, and found no difference to PD-L1 expression regardless of length of duration. However, although the differences were small (1.4% to 2.5%) and non-significant, a mean TPS change means little without the context of clinical guideline cut-off categories (the difference between 0 and 1% is maybe clinically relevant, for example). This finding was echoed by Kai *et al.* 2019 whom looked at 4 different NBF fixation times (ranging from 6 hours to 1 week) using the 22C3 clone (as well as Her-2 expression), albeit in 32 gastric cancer patients.[278] Taken together these results imply fixation time does not significantly affect PD-L1 expression, but these small sized studies, using different interpretation approaches, would need larger follow-up studies before the question of fixation time can be entirely dismissed, particularly when prolonged fixation is known to alter other IHC stains.[279]

These papers represent the majority of published data on the effects of different fixatives and fixation time on PD-L1 expression, and generally imply that, as long as a formalin-based fixative is used, this area of pre-analytics does not repeatedly cause major concerns for the effect it may have on PD-L1 analysis by IHC.

As noted, papers that only use dichotomous cut-offs for PD-L1 expression may still miss important data or over-emphasise small changes, but conversely only recording mean TPS changes without reference to clinical cut-offs may underrepresent the clinical impact of differences. The inconsistency of reporting PD-L1, as well as the myriad other variations between studies does make comparisons challenging, and as will be demonstrated, this is a common theme when looking at PD-L1 IHC.

2.1.1 Pre-analytics – Specimen Age

An important question in the suitability of specimens for PD-L1 expression analysis is the age of the specimens. This is important for two reasons: the first is the clinical and biological relevance of an older specimen in regards to how representative it is of the patient's current disease status, and the second is the possibility of older tissue falsely representing the status of PD-L1 by artefactual under- or over- staining. The first will be discussed later in this thesis, but the second concern will be considered here in detail.

The first major phase II/III clinical trials for the use of nivolumab (Checkmate 063, Checkmate 017, Checkmate 057 etc.)[46, 47, 280] looked at PD-L1 expression retrospectively, whilst trials at a similar time for pembrolizumab (Keynote 001, Keynote 010, Keynote 024 etc.)[48, 254, 281] incorporated the use of PD-L1 expression to stratify patients much earlier in the clinical trial pathway. In both cases, however, and in several subsequent trials, it seems apparent that PD-L1 expression was tested on a variety of different samples, including both fresh samples and archived samples from different patients, and it was only with some of the retrospective analysis that much thought appeared to be given to the age of the samples having a potential effect on PD-L1 expression, with Herbst *et al.* 2016 producing an abstract at ASCO's (American Society of Clinical Oncology) annual meeting that showed they found minimal difference in outcome for patients stratified for treatment using either fresh or archived tissue in the Keynote 010 cohort, using either the 1% or 50% cut-off value.[281] However this data was not followed up on in a publication, and no specific numbers were given. They do state that for the 50% cut-off value, 40% of archival and 45% of fresh samples demonstrated this level of positivity, but based on the study sample size (1,034), and the number of samples acquired fresh (578) or from archive (456), this is in fact a difference of 260 patients in the fresh group and 182 patients in the archived group, which is a relative decrease between fresh and archived of 30%; a not inconsiderable quantity of patients not eligible for the 50% cut-off depending on if they used fresh or archived samples. This abstract does have the benefit of being backed up by clinical outcome data, but is one of several pieces of data that showed varied and conflicting reports on the age of the specimen and the effect it has on PD-L1 expression.

Manufacturers generally recommend FFPE blocks are sectioned and stained for PD-L1 as soon as possible, and generally within 1-6 months for most applications[250-253] and the IASLC guidelines recommend using blocks no older than 3 years of age,[282] but the data for this is mixed. Calles *et al.* 2015 first suggested that blocks older than 3 years old should be avoided based on their cohort of 114 *KRAS*-mutant NSCLC patients.[283] They used a non-clinically validated clone (9A11) with a cut-off of 5% of tumour cells deeming a sample 'positive'. Based on this, they found FFPE blocks older than 3 years of age had lower numbers of patients positive for PD-L1 than from blocks newer that 3 years (39% of cases positive in the newer samples, dropping to 15% of cases in 3 years or older). This was echoed by Midha *et al.* 2017, another ASCO meeting abstract, that looked at data from the ATLANTIC trial, which used the SP263 clone to assess PD-L1 expression on 1546 NSCLC patients' samples, using a 25% cut-off for "positive".[284] They also found a large drop off of positive rates after 3 years (around 30% of cases positive in the newer samples, dropping off to 13% of cases in 3 years or older samples).

This loss of PD-L1 expression in older specimens is seen in other studies at different time points: Hata *et al.* 2017 found a drop-off in specimens older than 12 months (54.2% vs 26.4%), (though they could not confirm this in multivariate analysis) from 96 samples from 77 NSCLC patients using the 28-8 clone.[285] They used a semi-quantitative method to generate an H-score based on number of positive cells and intensity, (ranging from 0 to 300, 300 being 100% of cells positive at level "3" intensity), with a H score of ≥1 taken to be positive, ≥5 moderately and ≥10 strongly positive, with a final confirmation by pathologists. This loss was seen even with the observation that re-biopsied patients may have PD-L1 expression go up or down over time. Boothman *et al.* 2019 looked at 1,590 patients' samples from the ATLANTIC trial using the SP263 clone, and split these samples according to age (4 categories, from ≤3 months to >3 years) and used a 25% cut-off to deem samples as 'positive'. They found samples newer than 3 months were significantly more likely to express PD-L1 at higher levels than older samples, with positive rates decreasing with age (34.3% vs 30.3% vs 29.4% vs 20.2% across categories increasing with age, p=0.039).[286] Interestingly, however, they also studied 123 patients with paired samples that were newer and older than 3 months; in this sub-cohort the majority had no change (74%) but 6.5% of patients had a decrease of PD-L1 expression over time and 19.5% of patients had an increase over time. The effects of treatment did not seem to be behind these changes, but it is uncertain if this is genuine temporal heterogeneity or an artefact of older samples.

Giunchi *et al.* 2018 looked at 58 NSCLC patients and created TMAs from their samples. They used the SP142 clone and scored both the tumour cells and TILs to give a combined positive score of 1+ (<5% of cells), 2+ (5-50%) and 3+ (>50%) and compared samples from 2014 to ones from 2015 (~4 and ~3

years old at time of publication respectively).[287] They found a significant reduction in PD-L1 expression in the older samples for tumour cells (25.9% vs 15% p=0.018) though the numbers are small (15 cases vs 9 cases) and this was not seen for TILs. Their conclusion is that archived tissue is suitable for PD-L1 analysis, but caution should be used for tissues older than 1 year, though this conclusion is confusing as the article is published in 2018, some 3 years after their most contemporary tissue. The problems with SP142 in the context of NSCLC are discussed later in this thesis, but suffice to say it is not the preferred clone and so studies using this must be interpreted with caution.

Despite these papers showing generally consistent evidence that older tissue sees a loss of PD-L1 expression, not all published data does.

Kim *et al.* 2017 looked at multiple clones in 97 resected NSCLC specimens with matched TMAs using 22C3, SP263, SP142 and the research clone E1L3N. The major findings of this paper are discussed later. They used raw percentages to generate TPSs in increments of 5%, as well as dichotomous divisions using cut-offs of 1%, 5%, 10%, 25% and 50%. The study did not look at age of tissue as a primary outcome but they do mention that their tissues (collected over a 2 year period of 2015-2016) did not differ in terms of PD-L1 expression based on age of the FFPE block; an observation consistent for all clones.[288] Nakamura *et al.* 2018 looked at the question of aged tissue specifically, and used specimens from lung cancer patients with a wide variety of sampling techniques (surgery, bronchoscopy biopsies, CT guided biopsies, lymph node biopsies and pleural biopsies) stained for PD-L1 with both 22C3 and 28-8 clones, and categorised into non-expressing (<1%) low-expressing (1-49% for 22C3, 1-10% for 28-8) and high expressing (≥50% for 22C3, and ≥10% for 28-8) PD-L1 tumours.[289] The tumours were sub classified by age using a cut-off of 6 months or newer ("fresh") or older than 6 months ("archived"), and they found no significant difference in PD-L1 expression between fresh or archived tissue for either clone, though it should be noted the numbers for fresh (n=109) were much larger than archived (n=28). Wang *et al.* 2018 looked at 1326 specimens of NSCLC from surgical resections, small biopsies or cytology cell blocks (composed of EBUS-FNAs (Fine needle aspirates), pleural/pericardial fluids, EUS (endoscopic ultrasound) FNAs, FNAs and bronchoalveolar lavage).[290] PD-L1 expression was looked for using the 22C3 clone and scored to generate a TPS and categorised as negative, weak positive or strongly positive (<1%, 1-49% and ≥50% TPS respectively). When considering age as a factor specifically, they used a cut-off of 50% and split samples into <1 year, 1-2 years and >2 years and found no significant difference between age categories for PD-L1 expression.

Taken together, these papers initially seem to refute the conclusion that age is a factor for affecting PD-L1 expression in FFPE blocks. However, it should be noted that these papers all deal with tissue or cut-offs less than 3 years of age, and in the majority of cases that see a sharp drop off in PD-L1 expression, 3 years or greater seems to be a consistent finding. Indeed, Gagne *et al.* 2019 looked specifically at addressing the value of the 3 year cut-off based on its inclusion within the IASLC guidelines.[291] Using 1249 NSCLC patient specimens of various types, they stained for PD-L1 using the 22C3 clone to generate a TPS and categorised into negative, weak positive or strongly positive (<1%, 1-49% and ≥50% TPS) categories. Tissue specimens were split by age into 4 categories (<1 month, 1-12 months, 1-3 years and >3 years) and they found a significant reduction in PD-L1 expression based on the weak and strongly positive categories for specimens older than 3 years (p=0.0031), and unlike previous studies, they found this to be a consistent finding in multivariate analysis that considered various demographic and pathological characteristics.

Finally, if FFPE blocks are generally robust for most immunochemical stains, unstained sections are generally accepted as being more fragile.[292, 293] This phenomena means that tissue destined for immunochemical analysis is generally stained as soon as possible after sectioning, and indeed manufacturers of the PD-L1 clones recommend staining sections as soon as possible.[250-253] Not all IHC assays are as equally susceptible to loss when using aged tissue,[294] but oestrogen receptor (ER) and progesterone receptor (PR) IHC, which, similar to PD-L1 IHC, have a scoring element including staining intensity and quantity, and are used for guiding treatment decisions in cancer patients, have been shown to be vulnerable to this (Fig 2.1.1)[5, 295]. Given the variability of PD-L1 expression by IHC and the quantitative element of scoring, it seems reasonable to suggest that PD-L1 IHC maybe vulnerable to this loss as well, and if so it could have important clinical ramifications. This has not been widely studied, but Sato *et al.* 2018 looked at resected NSCLC specimens from 12 patients which had 5 serial sections stored and stained for PD-L1 using the 28-8 clone from time zero to 8 weeks.[296] This small study found that older sections had a stepwise loss of PD-L1 expression, and that sections should therefore be stained as soon as they are ready. Hendry *et al.* 2017 found samples sectioned fresh compared to the same samples sectioned 4 years previously had increased levels of PD-L1 expression intensity and significantly more cases reaching the 50% cut-off threshold for both the 22C3 clone (8.2% vs 3.7% p=0.019) and the SP263 clone (13.3% vs 3.3%, p<0.001) in TMAs (tissue microarrays) constructed from 368 NSCLC patients.[297] However, Rebelatto *et al.* 2016 found there was no difference in PD-L1 (SP263) expression in tissue sectioned and tested at 0 days to those sectioned and tested at different time points for up to one year, though as discussed previously this study used a single cut-off of 25% TPS and this alone is not the best indicator of change in PD-L1 expression.[272]

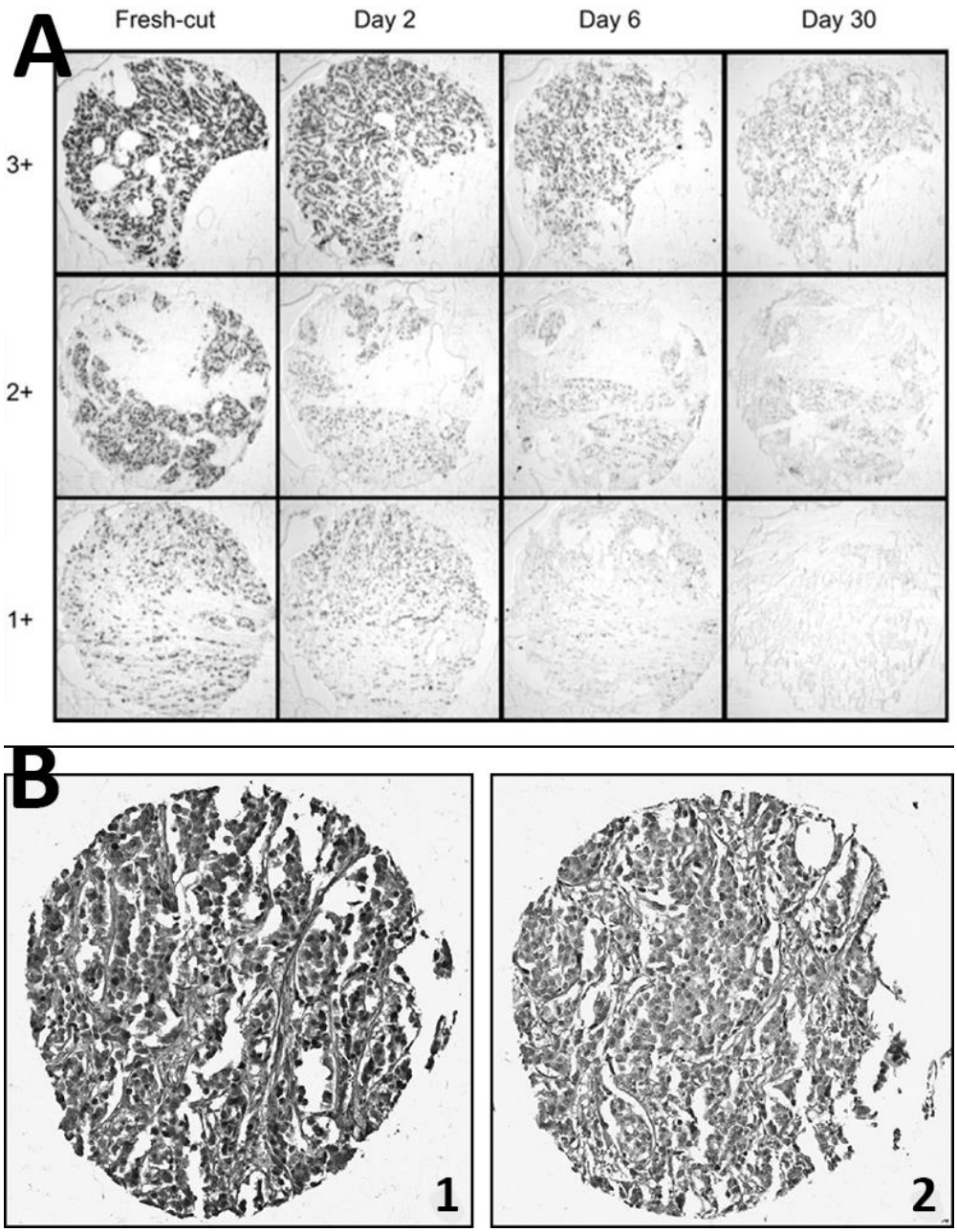


Fig 2.1.1 Composite image of two studies that have looked at predictive IHC stains used to guide treatment decisions in breast cancer and demonstrate the susceptibility of staining loss in older tissue sections. **A –** Oestrogen receptor expression from tissue freshly cut to 30 day old tissue, modified from DiVito *et al.* 2004.[296] **B** – Progesterone receptor expression in **B1** - Freshly cut section, **B2** – 6 month old section. Modified from Fergenbaum *et al.* 2004.[5]

Notwithstanding the typical problems with attempting to synthesise a conclusion based on various papers looking at a specific aspect of PD-L1 expression (differing clones, differing specimen types, lack of multivariate analysis etc.) the general body of data suggests that FFPE blocks are fairly robust for PD-L1 expression provided they are not older than 3 years of age, but samples from before that

time are at significant risk of having falsely reduced PD-L1 expression. Tissue sections are probably quite susceptible to PD-L1 expression loss over time, but there is little direct evidence of this.

Genuine biological change of PD-L1 expression over time (temporal heterogeneity) is discussed more fully later in this literature review, but it is worth mentioning here that artefactual change in PD-L1 expression in older specimens is only partially the problem. There is also the ethical dilemma of choosing to subject patients to sampling in order to gain contemporary tumour tissues, which involves invasive procedures carrying a non-zero morbidity and mortality risk, against using older, easily accessible, but less clinically relevant tissue. Whilst previous work has attempted to laydown guidelines in this area,[298] there is no simple solution, and it is worth bearing in mind that whilst tissue less than 3 years old may have no artefactual change in PD-L1 expression, it may bear little resemblance to a tumour that has metastasised and evolved over the same time period.

2.1.2 Pre-analytics – Specimen types

An important aspect of pre-analytics is how a specimen is retrieved from the patient, and how variations in the acquisition, storage and transport of such specimens might affect the eventual outcome. As most specimens validated via the major clinical trials were either biopsies or surgical specimens, 'histology' specimens are the recommended tissue of choice in terms of providing relevant clinical information. However, a considerable number of initial diagnostic specimens are 'cytology' specimens, that is, specimens in which the architecture of tissue is lost, and the diagnosis is made on cellular characteristics alone. Indeed, over a third of patients have only cytology specimens available for diagnosis in advanced NSCLC,[299] and it may be the most contemporary specimen available. Thus the question of the validity of cytology specimens is a critical one, but particularly challenging in that cytology specimens typically contain less material, represent a variety of tumour sites (primary, regional lymph nodes, pleural fluids etc.) and variation in terms of collection (e.g. smaller gauge needles than core biopsy), storage (often non-formalin based fixatives) and processing (typically cell blocks are produced for IHC), and indeed a difference in how they may be interpreted. Direct comparisons of fixatives has been discussed above, and the question of inter-tumoural heterogeneity between sites of tumour follows in this review, (though it should be noted that the use of cytology and histology specimens is a potentially confounding factor in a number of studies), but a consideration of cytology specimens in general, including direct comparisons between cytology and histology specimens, will be considered here.

2.1.2.0 Suitability of cytology specimens

Stoy *et al.* 2017 looked at the suitability of cytology specimens in 22 patients whom underwent EBUS-TBNA (transbronchial needle aspiration) or CT (computer-tomography)-guided aspirations of regional lymph nodes (LNs) (54.5%) or primary lung lesions (45.5%), with specimens placed in Cytolyt to prepare cell blocks for PD-L1, stained with 28-8. 2 cases (9%) were inadequate, but there was no difference in success rate between sampling methods or by needle gauge (5 sizes ranging from 19-guage to 25-gauge).[300] 30% of the samples were PD-L1 positive (by 1% cut-off), which is perhaps a little low, but 2 cases had matched histology samples from the same location, which were found to be concordant (though the authors do not specify TPS) and one with histology from another location (primary vs regional LN) which was found to be different (90% at the primary lesion, 5% via the cytology LN), although the authors do note the histology specimen was taken 37 days previously, and prior to treatment. The authors conclude that cytology specimens are largely adequate for PD-L1, but provide only minimal data for concordance with histology specimens. Biswas *et al.* 2018 looked at 50 retrospective cases of NSCLC whom underwent EBUS TBNA and assessed the adequacy of the samples for PD-L1 by 22C3 as well as for *ALK* FISH and NGS for driver-mutations. 82% of the samples were adequate for all three testing methodologies, with a further 4% suitable for PD-L1/NGS or PD-L1 alone, and 8% insufficient for testing at all, suggesting that cell blocks derived from EBUS-TBNAs generally provide adequate material for both PD-L1 and other critical predictive profiling.[301] Arriola *et al.* 2018 looked at a variety of cytology techniques and compared them to concurrent core biopsies using 22C3. Techniques included conventional cell blocks, cell-transfer cell blocks (CTCBs) from smears, and direct smears from a variety of primary and secondary sites.[302] Generally speaking cell blocks and direct smears performed well, with good levels of concordance to their matched histology (80% and 94.4% respectively) but CTCBs performed poorly, being only adequate in 70% of cases and with poor concordance (62% for ≥1% TPS, dropping to 17% for ≥50% TPS) with the authors concluding that smears and cell blocks are adequate substrates, but unconventional cytology approaches are not reliable for PD-L1 IHC. In addition, though direct smears provided good concordance, they were the most challenging to interpret with 22% deemed 'challenging' due to a variety of factors. Noll *et al.* 2018 also looked at direct smears and cell blocks compared to histology samples in 41 cases.[303] The specimens were from a variety of sites and not all were concurrent, but using 22C3 they showed a generally high level of agreement, superior for direct smears, when compared to the matched histology core biopsy (cell block 84% concordance, direct smear 97% concordance), crossing a ≥50% clinically relevant TPS group in 6 cases and 1 case respectively.

A number of other papers included cytology specimens as part of a study comparing various PD-L1 clones. Tsao *et al.* 2018 which reported the Blueprint study (discussed in more detail later) commented on the use of cytology specimens, and compared the ICC (Intraclass correlation coefficient) between multiple pathologists interpreting PD-L1 by various clones, and found generally good results for TPS scoring on both glass and digital slides (ICC 0.78 and 0.85 for all clones combined respectively) which was only slightly lower than histology specimens (ICC 0.89 and 0.93 for the same).[304] Villaruz *et al.* 2019 also compared different PD-L1 clones, and included a number of cytology specimens (grouped with small biopsies) but did specifically mention that for both 22C3 and SP263 the cytological nature of the specimen made no difference to PD-L1 expression rates.[305] Kim *et al.* 2019 compared 3 different PD-L1 clones in various NSCLC samples, which included 8 (18%) cases of cytology from either EBUS-TBNA or FNAs, but were included with other histology samples as 'small biopsies' and were typically compared to surgical specimens for PD-L1 expression analysis.[306] Overall concordance between small biopsies and surgical specimens was reasonable for all three clones at various TPS cut-offs (Concordance + Kappa: 22c3 73-96%, K 0.323-0.649; SP263, 72-91%, K 0.143-0.587; SP14265-80%, K 0.146-0.468), but due to various sampling sites and grouping it is difficult to interpret these results.

These studies provide evidence that sampling cytology specimens *per se* provide suitable quantity and quality for PD-L1 IHC, but the impact the different sampling techniques may have on PD-L1 expression requires a more direct comparison to histology samples.

2.1.2.1 Matched cytology and histology specimens

The most direct and robust method of comparing differences between cytology and histology specimens is by sampling the same tumour site by both methods to produce matched specimens, and comparing the resultant PD-L1 expression. Studies using matched specimens vary from use of predominantly matched specimens to comparing only a few, and these will be considered here.

Skov & Skov 2017 looked at 86 paired cytology and histology specimens of various types stained with both 22C3 and 28-8.[307] As discussed later 22C3 and 28-8 cannot be considered the same, but they are similar, as the authors demonstrated (R^2 of 0.95). Comparing cytology to matched histology specimens they found reasonable concordance for both clones at various cut-offs (R^2 ranges from 0.87 to 0.89 depending on cut-offs and clones). Although the confounding factors in this study include the wide range of sampling techniques, varying locations of specimen origin, and an unspecified number of cases waiting up to a month between section cutting and staining, it does however illustrate that PD-L1 expression across two clones is consistent in cytology specimens, despite these issues. Heymann *et al.* 2017 looked at 214 specimens from 188 patients and included

surgical resections, biopsies and various cytology specimens from different sites.[308] PD-L1 via 22C3 was scored for TPS, but only a TPS of ≥50% was considered positive. In regards to cytology, the average PD-L1 TPS was non-significantly higher than histology (39% vs 25% p=0.083), but there were only 4 cases of matched histology and cytology specimens with sufficient tissue, those these all showed agreement. The authors also looked at various sampling techniques compared to each other: for matched cases overall concordance was good (n=23, 91%), but the study is limited by a mix of sample types, often with many years between sampling, or treatment intervals, meaning the effects of inter-tumoural, iatrogenic and temporal heterogeneity are difficult to separate from variation due to sampling technique. Sakakibara *et al.* 2017 looked at a variety of samples taken from NSCLC patients, primarily focusing on EBUS-TBNA of regional lymph nodes compared to bronchial biopsies of primary tissue, resections or resected whole regional lymph nodes.[309] Various comparisons between each specimen type were made, the highest correlation being between regional lymph nodes sampled as cytology versus histology ($R^2 = 0.93$), followed by cytology of lymph nodes versus primary lung biopsy or resection (both $R^2 = 0.75$), which in fact had a higher correlation than lung resection versus lung biopsy ($R^2 = 0.52$). Unfortunately, although 97 specimens were included in total, only 27 had matched pairs, and in addition a research clone (EPR1161) was used. Despite these issues, this paper demonstrates a high level of adequacy for cytology specimens (98.9%) and between 0.75-0.93 R^2 concordance between cytology and histology regardless of sample site or location, although no clinical categories were used, and the characteristics of this clone are relatively unknown.

Ilie *et al.* 2018 compared 70 matched samples of histology and cytology with the 22C3 clone on 2 different LDTs (laboratory developed tests) by 2 independent pathologists.[310] ICCs were generally high regardless of the LDT or assay used when comparing histology to cytology (0.884-0.898), though cytology specimens were a mixture of bronchial washings and pleural fluid, with concordance generally poorer for pleural fluid (ICC 0.815-0.835) than bronchial washings (0.950-0.963). When using clinical cut-offs the concordances were similarly good (≥1% cut-off = 97% concordance, ≥50% cut-off 96% concordance). They also found the LDTs compared well to the validated platform, though slightly better for histology (ICC 0.999-1) than for cytology (ICC 0.936-0.947). Xu *et al.* 2018 compared 52 matched pairs of histology and cytology from EBUS or CT-guided procedures (they do not specify procedure or location, but do mention 'some' cases were from mediastinal lymph nodes, thus the question of primary or secondary tumour is not known) taken within a month of each other and stained for 22C3. Overall concordance was good: 81% of cases showed concordance using a 1% TPS cut-off (Cohen's Kappa 0.423) but when split by tumour type concordance was noticeably poorer for SCC over ADC (79% vs 83%, Cohen's Kappa 0.35 vs 0.54).

Russell-Goldman *et al.* 2018 looked at various lung cancers in 46 patients with paired cytology and histology, of various specimen types and various primary and secondary sites.[311] However the majority of cytology specimens were primary FNAs and taken concurrently from the same site as the histology specimens, and as a result found concordance generally good by E1L3N (Spearman's 0.78), though much poorer for immune cell scoring (Spearman's 0.23) and with lower inter-observer agreement (ICC 0.47-0.67 vs 0.96). Hernandez *et al.* 2019 looked at 52 paired cytology and histology specimens from a variety of primary and secondary locations and by various sampling methods, mostly concurrent, and, curiously, used a ≥50 tumour cell threshold for adequacy (when adequacy is generally taken at ≥100 tumour cells)[266, 282], though in fact 28% of their cases had fewer than 100 tumour cells.[312] Despite this, they illustrated reasonable concordance between histology and cytology specimens by 22C3 (67% concordance, Cohen's Kappa 0.51) which increased if only matched samples from primary tumour were considered (83% concordance Cohen's Kappa 0.74). The authors noted no significant difference in patients treated vs untreated, but the effects of treatment on PD-L1 expression is discussed in more detail later.

Grosu *et al.* 2019 looked purely at pleural effusions compared to histology samples in 82 matched pairs by 22C3 and found generally good concordance by both overall TPS correlation (0.79) and by clinical categories of negative, weak or strong (<1%, 1-49% and ≥50% TPS) positive (0.76), despite variation in time between samples and XCT delivered in some patients between sampling.[313] Generally, the pleural fluids were found to be more positive than the histology specimens by both cut-offs (≥1% TPS 56% vs 60%, ≥50% TPS 20.7% vs 24.4%). Sakata *et al.* 2018 compared 61 pairs of tissue samples from patients who underwent primary NSCLC resection and had a follow-up EBUS TBNA within 1 year.[314] Any PD-L1 positivity defined as ≥1% TPS was present for 48% of patients, but overall concordance between the specimens was generally high, (≥1% TPS - 87% concordance, ≥50% TPS 82% concordance) The majority of the EBUS specimens were from regional lymph nodes (69%) and showed little difference in concordance (86% and 83% concordance for the same cut-offs.)

The majority of cytology specimens in these papers included needle aspirates processed to form cell blocks prior to IHC, but a small number of papers using matched specimens also included direct smears. Munari *et al.* 2019 is a paper in a series of studies by this group using the same main cohort, from which they extracted 53 pairs of resections and FNAs as direct smears for histology and cytology comparisons.[315] At ≥50% TPS levels generally good concordance was found (90.6%) but poorer for ≥1% TPS (81.1%) with a Pearson's correlation of 0.9 if treated as a continuous variable. 50 cases also had matched core biopsies, which showed good correlation with the resections (92%) and only slightly worse when compared to the FNA smears (88%). However, the paper largely focused on the ≥50% cut-off, and did not provide complete data on 1% cut-offs, and also included some mixed

findings around 10%. Overall the authors conclude that TPS of 10-49% may warrant further sampling, but that smears are a reasonable substrate for PD-L1 analysis. Jain *et al.* 2018 looked at 26 matched primary core biopsies versus primary cytology specimens (bronchial brushings/washings with 5 cases also having direct smears).[274] The overall concordance between cytology and histology was 88%, but they use only a single cut-off of ≥25% TPS using SP263. The authors also note the interpretative difficulties of using direct smears, but this paper is at least limited to only primary tumours with concurrent collection of both samples, minimising the impact of heterogeneity.

Capizzi *et al.* 2018 looked at 50 paired direct smears and biopsies, with samples coming from both lymph nodes and primary tumour, though these were of unspecified numbers and there was no split analysis.[316] They compared three PD-L1 clones on the histology samples but used a duplex stain of SP263 and CD68 to make the cytological comparison and found a reasonable number of cases were concordant when split into negative, weak or strong positive (90%, K 0.364), increasing if using a 50% TPS cut-off only (K 0.626). Lozano *et al.* 2019 compared direct smears to a combination of cytology cell blocks and histology specimens (grouped together as FFPE specimens) from a variety of primary and secondary sites.[317] Despite the fact the authors did not separate out the cell blocks from the histology samples and a variety of sites were included, overall concordance between direct smears and FFPE tissue was good using 1%, 1-49% and 50% TPS groupings (97.3% concordance). A subset of samples (57) were stained for both 22C3 and SP263, and found there was minimal change between the clones for concordance of direct smears to FFPE (98.2% and 96.5% respectively). However, the authors again noted the difficulty in interpreting the direct smears for PD-L1.

Taken together, these papers suggest that cytology smear samples maybe suitable for PD-L1 analysis, though they generally seem to be challenging to interpret, which, as discussed later is a problem for inter-pathologist concordance, which may negate the usefulness of these specimens for PD-l1 IHC.

2.1.2.2 Unpaired cytology and histology specimens

A number of papers made indirect or non-matched comparisons of cytology to non-cytology specimens for assessing the use of these for PD-L1 expression. Evans *et al.* 2018 looked at over 10,000 cases of NSCLC with 22C3.[318] Cytology specimens were more likely to express PD-L1 at ≥50% TPS cut-off versus biopsies or resections (39.2% vs 29% and 22.5% respectively, $p<0.001$), however this significance is lost when comparing sampling methods in only pleural or lymph node samples ($p=0.086$, $p=1$ respectively) with the authors concluding that overall, the use of cytology specimens is suitable. This paper highlights the specific issue of comparing different sampling methods across multiple sites, and the potential confounding factor of heterogeneous expression, as well as the

issue of not using matched tissue, but does none the less provide a large dataset that shows that cytology specimens are no different to histology specimens, although with a tendency to slightly overestimate PD-L1 expression. Torous *et al.* 2018 looked at 232 non-matched specimens including 94 cytology specimens from a variety of sites including TBNAs, pleural fluids and FNAs.[319] Overall they found no significant difference between specimens based on sampling methodology, by 1% or 50% cut-offs, or by increasing 10% TPS increments (p=0.57), with similar rates of cases falling into weak-positive (21.3% vs 25.3% cytology and histology respectively) and strong-positive (35.1% and 34.8% cytology and histology respectively). Mei *et al.* 2019 looked at 265 patients with lung cancer (predominantly NSCLC) which included 100 cytology specimens, of which 96% were adequate for PD-L1 analysis by 22C3.[320] There were similar rates of PD-L1 positive cases for both 1% and 50% TPS cut-offs comparing cytology to histology (56% vs 52% and 31% vs 28.5% respectively). Samples were unmatched and came from various sites but the comparable rates of positive samples led the authors to conclude the suitability of cytology specimens, though again cytology was slightly higher for PD-L1 expression on average. Wang *et al.* 2018 also found a higher rate of PD-L1 expression in cytology specimens.[290] They looked at 1419 specimens that included 371 cytology specimens stained by 22C3, with no difference in adequacy noted between sampling methods. Cytology specimens significantly expressed higher rates of positivity than both surgical specimens and histology specimens at ≥50% TPS cut-off (42% vs 29% and 36% respectively), but included multiple sites including pleural fluids, EBUS, EUS and FNAs. There was no significant difference between biopsies and surgical specimens, however. They also included 27 matched samples, which showed significant correlation ($r=0.925$, $p<0.001$).

Higher average TPSs in cytology specimens in unmatched studies may reflect the inclusion of various biological sites. Conversely, as summarised in Table 2.1.1, cytology specimens may favour trends towards higher or lower TPS scores in any given study, but overall there is no definite trend. As with all aspects of understanding the limitations of PD-L1, many of these papers include samples of varying age, fixation, sampling sites, treatment effects and so forth, and thus the individual effect of 'cytology' is hard even to define, let alone compare. However, the general body of evidence would suggest that, with correct handling of other factors, the acquisition and use of cytology specimens should not be a barrier for PD-L1 analysis. This is perhaps further reinforced with the widespread acceptance and use of other IHC assays in cytology specimens [321] with evidence they are suitable for other predictive assays, such as ER, PR and HER2 expression in metastatic breast cancer. [322] However, it is worth remembering there are inherent limitations to cytology specimens, particularly samples from lymph node aspirates, that are relevant to PD-L1 assessment.

Paper	Cancer	Specimens (n)	PD-L1 Clone	Scoring	TPS for 'positive'	Pathologists	Difference	Age of specimens	Fixation	Type of specimens	Matched	Adequate (≥100 cells)	Cytology TPS vs Histology TPS	PD-L1 positive cases	Matched sample findings
Heymann et al 2017	NSCLC	214	22C3	TPS	≥50%	Multiple	Consensus	Not recorded	FFPE/Cytolyt	Resection, Biopsy, EBUS FNA, PF	23 pairs	Cyto - 90%, Bx - 96%, Resection - 99%	Higher	26%	Concordance: 93%
Sakakibara et al 2017	Lung Cancer	97	EPR1161	TPS	Any TC	2	Consensus	2013-2014	FFPE	Resection, Biopsy, EBUS FNA	27 pairs	96 (98.9%) (same for TBB)	Similar	NR	Pearson's r (0.75-0.93)
Stoy et al 2018	NSCLC		28-8	TPS	Any, (intensity recorded)	NR	NR	2016-2017	Cytolyt	EBUS -FNA, CT-guide FNA	3 pairs	20 (90.9%)	Similar	30%	Cytology generally suitable (3 matched pairs all have TPS <1%).
Skov Skov 2017	Lung Cancer	172	22C3, 28-8	TPS	Various	1	n/a	2012-2015	FFPE	Resection, biopsy, EBUS FNA, CT FNA	86 pairs	81% (100% histo)	Lower	Histology: 47-50%; Cytology 44-47%	Concordance: 22c3 - ≥1% TPS 85%, ≥50% TPS - 94%; 28-8 ≥1% TPS 87%, ≥5% TPS 95%, ≥10% TPS 90%
Ilie et al 2018	NSCLC	140	22c3 (2 LDTs)	TPS	1%, 50%	2	ICC (0.997 - 1)	2014-2016	FFPE or Novacyt	Biopsy, PF, BW	70 pairs	NR	Similar	NR	ICC (0.884-.0889) ≥1% TPS 3% discordanct, ≥50% TPS 4% discordant
Xu et al 2018	NSCLC	104	22c3	TPS	1%, 50%	3	Consensus	2012	FFPE	Biopsy, EBUS FNA	52 pairs	NR	Higher	Histology, ADC 57%, SCC 55%; Cytology ADC 74%, Cytology 62%	Concordance: ADC - 83% (K 0.54), SCC - 79% (K - 0.35) Combined - 81% concordance K (0.423)
Arriola et al 2018	NSCLC	60	22C3	TPS	1% (intensity recorded)	NR	NR	2016-2017	Non-formalin	Biopsy, smears, FNA	30 pairs	Cell transfer cell block - 70%.	Lower	80%	Concordance: Histology vs cell block 80%, Histology vs smear - 94.4%
Noll et al 2018	NSCLC	82	22C3	TPS	1% (intensity recorded)	2	Averaged (<10%)/consensus	2014-2017	FFPE	Biopsy, EBUS FNA, CT FNA, smears	41 pairs	Cell block - 93% Smear - 90%	Higher	Histology: 73%; Cytology 76%; Smear only 76%	Concordance: Histology vs cell block 84%, Histology vs smear 97%
Jain et al 2018	NSCLC	52	SP263	TPS	≥25%	NR	NR	2016-2017	FFPE/Cytorich	Biopsy, FNA, smears	26 pairs	Retrospective - selected out	Lower	Histology: 46%; Cytology 75%; Smear only 60%	Concordance: 88.4%
Capizzi et al 2018	NSCLC	100	SP263 (+ CD68)	TPS	Any, + intensity	NR	NR	2014-2017	FFPE/MicroFix	Biopsy, smears, FNA	49 pairs	Retrospective - selected out (98% of included)	Lower	≥1%/≥50% TPS 34-40%/20-30%	Concordance: 90% (K 0.364)
Russell-Goldman et al 2018	Lung Cancer	102	E1L3N	TPS	1%, 50%	2	ICC (0.96)	2015-2016	FFPE	Resection, biopsy, FNA, PF, BW	46 pairs	NR	Lower	≥1%/≥50% TPS Histology: 63%/30%	Histology vs all samples: Spearman's - 0.69, Histology vs FNA cell block Spearman's 0.78
Hernandez et al 2019	NSCLC	52	22C3	TPS	1%, 50%	2	3rd decided	2016-2017	FFPE	Resection, biopsy, FNA,PF, BW	50 pairs	(100% ≥50 cells but 28% <100 cells)	Lower	≥1%/≥50% TPS Histology: 73%/44%; Cytology: 54%/35%	Concordance: All cases 67% (K0.51). Primary tissue only 83% (K 0.74)
Lozano et al 2019	NSCLC	113	22C3 SP263	TPS	1%, 50%	2	Consensus	2016-2018	FFPE	Resection, FNA, EUS FNA, smears	62 pairs	Retrospective - selected out	Lower	FFPE - 50%, Smears - 50%	Concordance: 97.3%
Grosu et al 2019	NSCLC	164	22C3	TPS	1%, 50%	2	Averaged (<10%)/consensus	2014-2018	FFPE	Resection, biopsy, PF	82 pairs	Cyto 71.3%	Higher	≥1%/≥50% TPS Histology: 56%/21%; Cytology: 60%/24%	Concordance: 78% (K 0.76)
Munari et al 2019	NSCLC	115	SP263	TPS	1%, 50%	2	NR	2003-2017	FFPE/Alcohol	Resection (TMAs), FNA	53 pairs	Retrospective - selected out	Lower	≥50% TPS 19-28%	Concordance: All samples Pearson's 0.9, ≥TPS 50% 90.6%, ≥1% TPS - 81.1%.
Sakata et al 2018	NSCLC	122	22C3	TPS	1%, 50%	NR	NR	2006-2016	NR	Resection, FNA	61 pairs	84% (cytology only)	Lower	48%	Concordance: ≥1% TPS 87%, ≥50% 82%; Lymph nodes only LN ≥1% TPS 86%, ≥50% 82 83%
Wang et al 2018	NSCLC	1419	22C3	TPS	1%, 50%	2	Consensus	2016-2017	FFPE/Alcohol	Resections, biopsies, EBUS FNA, EUS FNA, PF, BWs	27 pairs	93%	Higher	≥1% TPS (71%) ≥50% TPS (36%)	Pearson's r 0.925, (p<0.001)

Table 2.1.1 Papers with matched cytology and histology specimens comparing the expression of PD-L1.

EBUS, endobronchial ultrasound; EUS, oesophageal ultrasound; FNA, fine needle aspirate; PF, pleural fluid; BW, bronchial washings; CT, computer tomography; NR not recorded; FFPE, formalin fixed paraffin embedded; TPS, tumour proportion score; NSCLC, non-small cell lung cancer

As will be discussed in chapter 7, cytology specimens lose architectural features, and so many aspects of the TME cannot be fully examined. Perhaps most relevantly, whilst PD-L1 expression in NSCLC is limited to tumour cell scoring only (with the exception of the SP142 clone), many other tumour types involve immune-cell scoring as either a separate or combined component. Whilst the nature of a lymphocyte as a genuine TIL or not can be ascertained with biopsy samples from most primary tumour tissues, a fine-needle aspirate from a lymph node renders this distinction impossible, and therefore cannot be used to score immune-cell PD-L1 expression. Thus whilst cytology specimens are not a barrier to exploring PD-L1 expression *per se*, their use depends on the specific clinical context.

2.1.3 Pre-analytics – Other

Pre-analytical conditions outside of fixation and storage are important but under evidenced in regards to PD-L1 specifically, and as previously mentioned, IASLC guidelines thus refer to general good practice and knowledge from other immunochemical protocols to guide the parameters for these conditions: cold ischaemia time should be fewer than 30 mins and no more than 1 hour, the paraffin embedded sections should be 3µm to 5µm, de-calcification should be avoided but EDTA is best used if necessary, and so forth.[282, 323] There is, however, a limited number of articles that do discuss some of these other pre-analytical conditions and these will be discussed here.

Forest *et al.* 2019 is a small study that looked at decalcification in NSCLC samples (excluding negative and heterogeneously expressing samples) as well as placental tissue specimens.[276] Each sample was divided into multiple pieces, and after 24 hours of fixation in 10% NBF a piece from each sample was placed in either EDTA or DC3 Qpath decalcifier (a chlorhydric acid based de-calcifier) for 6 differing time periods ranging from 4 hours to 24 hours. A separate parallel study looked at placing the tissue in EDTA or DC3 for 1 day, 3 days or 5 days. PD-L1 expression on each piece of tissue was then assessed using the 22C3 and research only E1L3N clone. In the samples de-calcified for less than 24 hours, DC3 caused a significant reduction in PD-L1 expression by 22C3 by 4 hours, with only a non-significant trend for EDTA even by 24 hours. Staining intensity of 22C3 tissues was reduced at 8 hours for DC3 but took until 20 hours for the same decreased in the EDTA samples. Within the 24 hour period there was no significant loss of staining or staining intensity for the E1L3N samples. For the longer de-calcification periods, PD-L1 expression loss was seen in DC3 by day 1, but 48 hours was required before significant loss was seen using EDTA, though there was significant loss of PD-L1 expression intensity for both agents at all time points. For the E1L3N clone a significant loss in PD-L1 expression was seen by day 3 for DC3, but with no loss for EDTA even at day 5. Staining intensity was

significantly reduced in E1L3N stained tissue using DC3 by day 1, but only a non-significant loss in intensity for the EDTA samples by day 5.

At the time of the literature review this was the only significant paper to address the issue of decalcification, but two further papers that look at this issue are worth including here.

Strickland *et al.* 2020 performed a similarly comprehensive, albeit small study, on the use of various de-calcification agents and their effect on PD-L1 expression.[324] 10 placentas and 10 non-neoplastic lungs (selected for PD-L1 expressing macrophage populations) were fixed in 10% NBF and then divided into multiple pieces and placed into 4 different decalcifying agents: EDTA, FA/MC (formic-acid based) 12% hydrochloric acid (HCl) and DeCal STAT (23% HCl) for 4 different time periods each (1, 2, 6 and 24 hours) as well as a control piece for each sample that was not decalcified. TMAs were generated from the tissue and were stained for PD-L1 using the 22C3 clone and a H-score based on quantity and intensity of PD-L1 expression was assigned (as described earlier). For the EDTA and FA/MC samples there was no loss of PD-L1 expression or intensity in the placental specimens. For the 12% HCl and STAT samples there was significant loss of PD-L1 expression by 24 hours, with almost total loss for the STAT group in the placental tissues. In the lung tissues with macrophages expressing PD-L1, no loss was seen at any time point for the EDTA group, with a mild to moderate loss seen by 24 hours for the FA/MC group. Again the 12% HCl and STAT group fared poorly, with loss seen in both groups by 1 hour, and again almost total loss by 24 hours in the STAT group. Their conclusion was that EDTA, and to a lesser extent FA/MC were suitable decalcification agents for use with tissue destined for PD-L1 expression. These findings were echoed in a study by Pontarollo *et al.* 2020 whom looked at 84 NSCLC bone metastases specimens de-calcified using EDTA or formic acid over a wide range of non-specified times (4 hours + for biopsies, 4-15 days for surgical specimens) and stained for PD-L1 using the 22C3 clone.[325] The samples were grouped into TPS categories of 1-49% or ≥50%. There was no significant difference found in the rate of PD-L1 expression using these cut-offs between the de-calcified (n=39) versus non-decalcified specimens (n=45). Though there was no comparison between EDTA and formic acid, and no comment was made on intensity loss, these findings fit with the general consensus that EDTA is the de-calcification agent of choice.

There is little available data on other pre-analytical factors specific to PD-L1, though Rebelatto *et al.* 2016 do mention some relevant findings in addition to those mentioned above: using SP263 in NSCLC and HNSCC, they found no significant change in PD-L1 expression across tissue section thickness (from 3µm to 7µm) and there was no difference in expression depending on cold ischaemia time ranging from 0 to 24 hours or in slides stored at cold (2-8°C) or hot (30°C) conditions, though this study has consistently seen minimal change despite a wide variation of pre-analytical

factors that has seen difference in PD-L1 expression in other studies, and is likely due to the use of a single cut-off of 25% TPS to discern variation.[272]

Pre-analytical conditions therefore seem to have an effect on immunochemical expression in general, and this includes the specific use of PD-L1. The variation in findings between different studies simply highlights the lack of consistency between labs and if anything clearly illustrates the need for pre-analytical conditions to be robustly uniform and controlled between sites in order to minimise false change in PD-L1 expression. A consistent approach to these factors is important, but there remains several areas of limited evidence, and the question of how best to minimise pre-analytical variation, and how to overcome these effects is not fully answered at present.

2.2 Analytics

Variation in analytical factors include differing PD-L1 IHC antibodies, epitope retrieval methods and signal amplification, but there is little freely available data in regards to analytical factors affecting PD-L1 expression, as the nature of the specific PD-L1 antibodies, including epitope binding regions are largely propriety, with studies that have looked at this showing various data: 28-8, SP263, SP142 have both intracellular and extracellular binding regions, whilst 22C3 is predominantly extracellular and some of the antibodies may have overlapping but non-identical epitope binding regions.[326, 327] The clinically validated clones, 22C3, SP263, SP142 and 28-8 are all available as companion or complementary diagnostic tools that are used on specific platforms with validated protocols. Generally the advice is to use these where possible over LDTs,[266, 282] though where this is unavailable or impractical, reasonably good results can be attained with LDTs, as long as there is sufficient quality assurance steps.[328, 329]

2.2.0 Analytics – PD-L1 clones, LDTs and inter-pathologist concordance

Given the data pertaining to PD-L1 epitope binding regions, the studies that have shown PD-L1 antibody clones to have variable binding characteristics and distinct epitope mapping results and the visibly different staining pattern between the clones, it is unsurprising that many studies have looked to compare the similarities and differences of PD-L1 clone's staining pattern and quantity. Furthermore, as each series of clinical trials focusing on specific anti PD-1/PD-L1 therapies with different PD-L1 clones to guide therapy, and the scoring approach to each differed, the need to see which clones were equivalent became vitally important for guiding prediction of IM therapy. It is improbable that most clinical laboratories would have access to all validated platforms, and the logistics and communication required to ensure specific testing was applied to each sample would almost certainly lead to delay.[266] The most likely outcomes from this include: multiple testing of

different PD-L1 clones on samples (but this would waste tissue and be excessively expensive), testing would be ignored or only used in requested cases (but this would result in unsatisfactory numbers of patients being tested and unnecessary delays in acquiring results) or, as the simplest solution, a single clone on a single platform would be used in any given lab for NSCLC PD-L1 testing, and the TPS score would be used to guide therapy, depending on the particular cut-offs for any given IM. However, though this final approach is the most efficient and cost effective, this does require the PD-L1 diagnostic assays to be equivalent and for any clone used to give a universal TPS such that the relevant IM can be appropriately prescribed. This also requires scoring to be concordant between different pathologists regardless of the clone used, and the equivalence of using clones on non-approved platforms to create a LDT. As this review will now explore, this is not simple to achieve.

2.2.0.0 Early papers comparing PD-L1 IHC clones and LDTs

Scheel *et al.* 2016 is one of the first attempts to rationalise the different scoring criteria between each PD-L1 clone, and even go as far to generate an attempt at unifying the different approaches into a "Cologne Score".[330] Despite the lack of uptake of the Cologne Score (with the simpler percent based TPS becoming the *de facto* scoring approach), it is an important early paper that looked at multiple clones in NSCLC. A total of 30 NSCLC resections (18 ADC and 12 SCC) were scored in two rounds, the first using LDTs (E1L3N and SP142), and the second 4 different clones and their validated assays (SP263, SP142, 22C3, 28-8), with 9 pathologists included for scoring. Differences in staining type between the clones was observed and, even blinded, the pathologists were able to sort the images by antibody. Overall the findings were similar between the four clinically validated clones, but in four particular cases they noted that 28-8 and 22C3 were generally similar in intensity and total number of cells stained, but that SP263 was more intense and stained more cells, and SP142 was more intense than 28-8 and 22C3 but stained fewer cells, and these differences would have potentially changed clinical categories for specimens using cut-offs of 1% or 50%. They also found inter-observer concordance to be moderate: (Kappa 0.47-0.5 for the "Cologne Score" and Kappa 0.6-0.8 for the percent based cut-offs). They also note the concordance between assays for immune-cell scoring was very poor and inter-observer concordance for the same was very low (Kappa 0.12-0.25). Despite the relatively modest size of this paper, it does highlight some findings that are generally echoed in subsequent papers: SP142 stains fewer tumour cells but more immune cells, SP263 has a cleaner, more intense, staining pattern, inter-observer concordance is modest without being excellent, and immune-cell scoring with PD-L1 in NSCLC generally results in poor inter-pathologist concordance. In addition, attempts to rationalise scoring which have resulted in more complex categorisation prove both to be unpopular and less accurate for inter-observer concordance than simple scoring approaches.

A few months later Neuman *et al.* 2016 published work that looked at another key factor pertaining to PD-L1 clones by asking if a single clone can be used on different platforms to achieve comparable results.[331] Using the 22C3 clone they compared staining on the validated Dako autostainer Link48 to the Roche-Ventana BenchMark XT (using two types of Ventana detection kit on the BenchMark XT: OptiView and UltraView). A validation cohort of 41 NSCLC cases were selected and multiple pathologists scored these using the 22C3 on the Dako platform, and re-scored after an undefined washout period with cases assigned negative, weak or strongly positive TPSs (1%, 1-49%, ≥50%). Intra-observer concordance utilising these groups was perfect (Pearson's correlation = 1). Two pathologists then scored the case cohort that used the Ventana platform and detection kits. 66 protocols were used to optimise the process before the selected protocol was used in the comparison of PD-L1 scoring. The UltraView approach had agreement in 87.8% of cases, and the Optiview approach 85.3% of cases (in fact, a difference of one case). The Ventana protocols also showed high levels of intra-observer concordance (Pearson's 0.94). Most importantly, using the Ventana platform weak-negative cases were mis-classified as negative, and overall weaker staining or loss of staining within germinal centres was seen. This paper has several findings that are again echoed in later papers: high levels of concordant results are achieved when using the validated platforms, LDTs using differing platforms can be used to achieve comparable results but considerable time and effort is required to set these up, with a need for external validation very apparent, and even slight changes in protocol can change weak staining cells to negative in a manner than is clinically relevant.

It is worth at this point exploring some other data pertaining to the use of LDTs and the general finding that LDTs, if used improperly, can provide very poor and discordant results, but if optimised and utilised effectively, they can generate very comparable results to the validated platforms. A significant study in this regard is from Adam *et al.* 2018 which looked at 7 centres, each providing 6 cases of NSCLCs of varying known PD-L1 expression, and used 28-8, 22C3, SP263, SP142 and E1L3N on a wide variety of platforms and scored to generate a TPS.[328] Between centres, for the validated assays of 28-8, 22C3 and SP263, a generally good concordance was found between centres (0.79-0.94), but poorer when comparing the clones against each other (0.71-0.89). LDTs for each clone were compared to the validated assay, with E1L3N and SP142 compared to the validated SP263 assay. Using a weighted kappa concordance threshold of 0.75, they found only 51.8% of LDTs were concordant. Despite this, several other studies have reported LDTs can achieve good concordance if sufficiently optimised. Cogswell *et al.* 2017 used 28-8 and E1L3N as LDTs as well as using the validated 28-8 assay on a variety of tissues, including NSCLC tissues and cell lines to score a TPS.[332] They found the two clones were very similar when both using the same LDT protocol, but that the

validated protocol for 28-8 was generally more sensitive, with less non-specific staining. This study was limited, however, by small sample size and few matched cases. Villaruz *et al.* 2019 compared 22C3 and SP263 on the same platform (Ventana Benchmark Ultra).[305] The similarities of the clones are discussed later, but they also illustrate that very comparable results (0.89) between the clones can be attained using LDTs if sufficient effort is made to optimise them. Generally speaking LDTs may be suitable, but only if sufficient time and robust validation is implemented. It is also worth noting that although several studies use the correlation of average TPS scores to compare clones, this is not necessarily the best approach as the absolute change in TPS maybe small but still sufficient to re-categorise tumours into different clinical groups.

2.2.0.1 PD-L1 protein abundance and the use of TMAs to assess PD-L1 concordance

The considerable problem of PD-L1 expression heterogeneity, discussed in more depth in this thesis later, is likely to be a significant factor when using TMAs, but nonetheless several studies have used these in PD-L1 clone comparisons studies. Smith *et al.* 2016 compared the research clone E1L3N to SP263 using TMAs from 100 NSCLC specimens.[333] The authors describe a process of developing and optimising the analytical process including antigen retrieval and primary antibody concentrations before an agreed optimised process for both clones was used in the main study. Two pathologists were blinded and scored all the cases independently. They used an H-score incorporating the number of positive tumour cells as well as staining intensity, as described earlier to give a score of 0-300. Tumour associated immune cells were also scored and given as a percentage of positive immune cells of all immune cells at any staining intensity. This paper also included subjective feedback on how much the pathologists 'liked' each stain for each case, with the results strongly favouring the SP263 clone, with the greater staining intensity cited as the main reason. SP263 found more cases to be positive than E1L3N, and the H-score was also generally higher, with more cases with positive immune cells found as well. Finally, inter-pathologist concordance for SP263 for the tumour cells was higher (R^2 0.87 vs 0.82). Given the greater sensitivity, the equivalent specificity, higher pathologist preference and concordance, SP263 was deemed to be the superior clone. Whilst this highlights the effectiveness of SP263, E1L3N is a research clone, not validated for clinical use, so it was always highly improbable E1L3N would be considered the superior choice.

Hendry *et al.* 2018 looked at previously constructed TMAs from two samples each of 355 lung cancer patients from one cohort and one sample each from 68 cases from another cohort.[297] Tissue was stained for 22C3, 28-8, SP263 and SP142, with 22C3 used both on the approved Dako Link 48, and also on the Ventana Benchmark Ultra using the SP263 protocol. A pathologist scored each case to generate a TPS and IC score, excluding alveolar macrophages. Intensity of staining was recorded only

for the two 22C3 assays. Both absolute TPS/IC scores were captured, as well as groupings of the cases into cut-offs by the respective clones instructions (1% and 50% for 22C3, 1% for 28-8, 25% for SP263 and 50% TPS/ 10% IC for SP142). SP142 returned consistently lower TPSs than the other clones, and SP263 consistently higher, with 22C3 and 28-8 generally similar to each other. (Mean TPS 3.57, 12.05, 9.2 and 9.46 respectively). Curiously IC staining was highest for 28-8 and lowest for SP142, in contrast to most other studies. TPS overall concordance between the assays was 0.674, raising to 0.755 if SP142 was excluded, but highest between 22C3 and 28-8 (0.812), and lowest between SP263 and SP142 (0.525). Concordance for IC was generally poor for all clones (0.212) and for pairwise comparisons (0.027 – 0.403). Using clinical cut-offs for each clone as defined above compared to each other, the overall correlation was poor (0.43), with a total of 24.8% of cases being discordant between clones. Using specific cut-offs of ≥50% and ≥1% TPS showed variation between the clones, with SP142 consistently scoring fewer cases as positive, but a close agreement between the other 3 clones at ≥50% (9%, 101.1% and 11.8% classified as positive). The two 22C3 assays showed high degrees of concordance (overall 0.921 correlation, with kappa of 0.897 for ≥50% TPS cut-off) with 5.3% of cases discordant between them. The authors state the limitations of using small tissue samples as TMAs, and discuss the problems of heterogeneity (discussed in more detail later in this review), with only 88% of cases showing the same result in both cores. None the less, this paper illustrated an oft repeated finding that SP142 stains fewer tumour cells, and SP263 stains more, with corresponding changes in clinical groupings. Using a different platform to the approved one shows good, but imperfect correlation, though it is uncertain which is more 'correct'. This paper is also one of a small number to correlate ' PD-L1 positive', defined using each different clones' 'cut-off' (as opposed to using the same cut-off for each clone); with the findings that this approach does not improve concordance between the clones, once again demonstrating the difficulty in applying the clinical cut-offs consistently between them. Marchetti *et al.* 2017 looked at 4 different centres in which 100 NSCLC cases were used to generate TMAs and then stained for both 22C3 and SP263, and were assessed by 4 pathologists to generate TPSs using 1% and 50% TPS cut offs. Inter-pathologist concordance for each clone was good for 50% (0.95-0.97) but poor for 1% TPS (0.77 vs 0.82).[334] SP263 stained more intensely, but again good concordance between the clones was found at 50% TPS (0.99) but poorer when using 1% (0.80).

A number of studies utilised alternative detection techniques to compare PD-L1 protein abundances by multiple approaches. Brunnström *et al.* 2017 looked at 55 NSCLC samples, from which two 1mm cores were extracted to be stained with 28-8, 22C3, SP142 and SP263.[335] Seven pathologists scored the slides independently and were blinded to the clones, and each case assigned a score of 0-5 based on TPS cut-offs (from <1% to ≥50%) using any membranous staining. This study also collected

RNAseq data for 35 of the cases for PD-L1 mRNA. A consensus score for each case and clone was taken if the majority (≥4) of pathologists agreed. In 5% of cases, there was no majority agreement, and the consensus score was thus taken to be the median (in fact, equivalent to the mean in these same cases). Five or more pathologists agreed in only 86% of cases. Using these consensus scores, considerable variation between the clones were found with difference cut-offs. Using a ≥1% cut-off, 16-44% of case were deemed positive, and using a ≥50% cut-off, 5% to 24% of cases were deemed positive, in both instances SP142 scoring the fewest positive, and SP263 scoring the most positive. Correlation of the clones using pair-wise comparison was highest for 28-8 and 22C3 (0.88-0.91) and poorest for SP142 versus any other clone (0.45-0.63). Inter-observer concordance varied from 0.71-0.96, with SP142 showing the highest concordance, and SP263 the lowest. This study also found that using a ≥50% cut-off resulted in fewer cases changing clinical groups by different clones compared to a ≥1% cut-off ($p<0.01$), with as many as 20% of cases being re-classified using the ≥1% cut-off. Interestingly, the highest correlation of mRNA and PD-L1 TPS was for SP263 (rho 0.780) and lowest for SP142 (0.643). This suggests that SP263's trait of staining more tumour cells for PD-L1 is a more accurate observation of PD-L1 protein levels. The authors conclude that SP263, 22C3 and 28-8 are reasonably concordant, but that inter-pathologist concordance is a source of error for scoring PD-L1, particularly at the 1% cut-off threshold.

Soo *et al.* 2018, a study that looked at various samples for performing PD-L1 analysis, and included 20 specimens composed of histology and cytology specimens, stained PD-L1 by 22C3, 28-8, SP142, SP263 and E1L3N, and scored by three pathologists to generate TPS scores in 0%, 1% and 5% TPS increments thereafter.[336] This study also assessed PD-L1 mRNA levels by RNAscope. PD-L1 IHC were compared pairwise, and when considering TPS as a continuous variable, correlations were generally average: SP142 compared to other clones was generally very poor (R^2 0.31- 0.62), with the best correlation between 22C3 and E1L3N (0.72). Of note, correlation of SP263 and 22C3 was very poor (0.29) as was 28-8 with either 22C3 (0.42) or SP263 (0.41). Specimen type had no major impact of grouping of PD-L1, interestingly, given the inclusion of cytology specimens. Using each clones' respective cut-off according to clinical trials, the number of cases deemed 'positive' ranged from 33% (28-8, SP263) to 72% (22C3), with SP142 scoring 56% positive. Only 6 cases had positive PD-L1 mRNA present, and only the non-clinically validated clone E1L3N correlated significantly with RNA levels. Interestingly, if using the SP142 clone with the SP263 protocol, a higher level of correlation between the clones could be achieved (0.71). This paper had a small study size, but showed potential significant variation between clones, which interestingly did not correlate well with PD-L1 mRNA.

Gaule *et al.* 2017 looked at 6 different clones, 4 clinically validated (SP142, SP263, 22C3, 28-8) and 2 research clones (9A11, E1L3N).[337] TMAs were generated from cell lines, normal tissue and 30 NSCLC samples. PD-L1 was analysed using both IHC and quantitative immunofluorescence (QIF), and optimisation of each clone was performed (though minimal data is included in the manuscript on this process). No pathologist TPS scoring was performed, rather the positive pixel counting tool from Aperio was performed on the samples.[338] Each clone was compared pairwise, with generally high levels of concordance seen in the cell lines (0.83-0.96) and poorer concordance for tissue samples (0.42-0.68). Additional comparison using cell-line microarrays also found high levels of concordance between the clones, (0.76-0.99) with E1L3N being the biggest outlier. The stated aim of the authors was to show correlation of protein abundance and they cite heterogeneity as the primary reason for poorer correlation between tissue samples. However, positive pixel staining will also include non-specific, immune cell and cytoplasmic tumour cell staining and is therefore not directly linkable to TPS, and has been discussed, it is the subtle differences that change clinical grouping of tumours that is the biggest challenge.

Parra *et al.* 2017 compared a large number of clones, nine in total, in NSCLC and cell lines, and compared IHC results to Western blot results, to ascertain the suitability of each clone to detect PD-L1.[339] Of the original nine clones, only six passed validation by Western blot and IHC for subsequent use in the study: 22C3, SP263, SP142, 28-8 and the research clones E1L3N and E1J21. The Western blot analysis revealed a molecular mass band around 43kDA in most cell lines for all of these clones except 22C3. Two pathologists were blinded and scored the cell lines and tissue samples to generate a H-score based on quantity and intensity of staining (0-300) using Aperio Image Toolbox analysis software. Three cores of tissue from each tissue sample were averaged and included both tumour cell and immune cell positivity. The pathologists also used light microscopy to score the samples with a TPS and a third pathologist used in cases without an agreement to form a consensus score. Curiously, SP142 had the highest median H-score, but the in-depth data found generally good correlation between clones, but poor between SP263 and 28-8 (0.342), 22C3 (0.321) and E1L3N (0.345) and SP142 (0.373) with SP263 consistently returning higher H-scores, as well as consistently scoring the highest number of cases as 'positive' using cut-offs as ≥1%, ≥5% or ≥50%. The authors note that SP263 achieved this in both the tumour defined regions, but also the stromal regions, and suggest the image analysis software had a tendency to over score tissue, and thus oversight by pathologists was wholly necessary. This paper again highlights that absolute correlation of clones, either in terms of absolute protein abundance, or its specific location within the tissue, is only part of the issue between varying clones – correct interpretation to allow for precise clinical categorisation is arguably the harder, and more important task. Furthermore, when using scoring

methods that integrate staining intensity, SP142, that typically stains fewer tumour cells compared to other clones, may receive a disproportionately high H-score, possibly as a result of more intense staining secondary to antibody amplification, which other clones may not use.

2.2.0.2 Large studies comparing multiple PD-L1 clones

NSCLC was the primary focus of most early PD-L1 IHC studies, though it is interesting to note that similar limited studies in other tumour types were also being reported, for example in prostate (SP263 vs E1L3N)[340] and melanoma (SP142 vs E1L3N).[341] None-the-less, with multiple PD-L1 clones now developed for IHC in NSCLC, a series of larger studies comparing multiple different clones for concordance were reported, largely in 2017 and 2018. A general consensus of agreement was suggested, illustrated in the IASLC guidelines and largely based on the Blueprint study, although more recent data suggests that no two clones should be considered absolutely equivalent.

Rimm *et al.* 2017 was a major study looking at 28-8, 22C3, SP142 and E1L3N in 90 NSCLCs and scored by 13 pathologists.[342] The 3 clinically validated clones were compared to E1L3N as a LDT, although at the time, SP142 was not quite validated, so it was considered a LDT for the purposes of the paper. Again, an attempt at a unified scoring system was made (given the complex suggested scoring system provided by Roche-Ventana for SP142) in which a TPS and immune cell proportion score (ICPS) were given for each section, using a range of six categories for TPS (A-F) and three for ICPS (A-C). The line "the score of TPS or ICPS is based on membranous and cytoplasmic staining of any intensity" was not clarified in the supplementary tables, so it is uncertain if TPS was limited purely to membranous staining as is generally now accepted. The inter-pathologist concordance for TPS was generally good (0.83 to 0.88 for each antibody), but for ICPS very poor (0.17-0.23). Specific concordance for the TPS as the cut-offs of 1% and 50% were poorer, however, at 0.54 and 0.75 respectively. The average score for the TPS and ICPS for all pathologists was used to compare each clone pair-wise against each other, and found every pairing (except 28-8 versus E1L3N) were statistically different, though these ranged in small differences of TPS from only 0.246 to 1.2. They found the ICC to be 0.81, raising to 0.97 when excluding SP142 for TPS, and 0.27 raising to 0.8 for ICPS, reflecting the generally poor ability of SP142 to stain for tumour cells, but strong ability to stain for immune cells. The authors concluded that 28-8 and E1L3N were practically equivalent, and whilst 22C3 showed statistically less staining, the change was minimal, and was only found when using the pathologists' average score. This paper again demonstrates the issue that absolute variation between clones for TPS might be small, but the specific categorisation into clinical cut-off groups remains a significant hurdle. Inter-pathologist concordance is not perfect, with immune cells a

particular challenge, and a complicated scoring system does little to help this, and indeed has found little utility outside this study.

Ratcliffe *et al.* 2017 looked at 493 NSCLC samples using the SP263, 22C3 and 28-8 clones, and scored by a single pathologist using each manufacturer's instructions (cut-offs of 25%, 1%/50% and 1%/5%/10% TPS respectively) as well as TPSs given as 5% increments.[343] A second pathologist performed an independent review of 200 of the cases, with appropriate wash-out periods for both pathologists between scoring the same cases for different clones. Each clone was compared pair-wise at each 5% increment, with Spearman correlations for each clone at any cut-off over 0.9. Interestingly the lowest correlations (0.91) were seen at the ≥1% cut-off, with the highest (0.97) at the ≥50% cut-off. The overall percentage agreement was thus >0.9 for each clone compared to any other in the 493 samples, but this dropped when comparing the second pathologist's results in the 200 sample sub-cohort: again ≥50% cut-off achieved a high correlation (0.945 -0.975) but was worse at lower levels, with poorest concordance seen with the ≥1% cut-off (0.759-0.77). This paper highlights that, for most cases, particularly at higher TPS scores, the differing clones and differing pathologists return similar results, but the critical cut-offs at lower values, particularly around 1%, is a problematic issue in a significant minority of cases.

A major study in this area was Tsao *et al.* 2018, which was the final paper of the Blueprint study, an IASLC sponsored study in which 22C3, SP142, SP263, 28-8 and 73-10 (the first paper to include this latter clone) PD-L1 clones were compared by 24 experienced pulmonary pathologists, in 71 cases of NSCLC (and 10 case of small cell lung cancer (SCLC)) from various sampling methodologies on validated platforms and protocols.[304] A TPS was generated for all assays and results divided into seven categories (from <1% to 80-100%), as well as an IC scored generated based on the SP142 interpretation approach. The overall inter-pathologist concordance ranged from 0.80 to 0.93 (for both digital and glass slide, with glass slide slightly higher), with higher concordances achieved if excluding cytology and the SCLC specimens. (0.88 to 0.95). K statistics for various cut-offs were all >0.7, though they note this was weakest at the ≥1% and ≥80% cut-offs; the extremities of expression. The authors then used the mean TPS values across all pathologists for each assay and sample, and compared these both as a cohort and pairwise. 22C3, 28-8 and SP263 were the most similar, with SP142 consistently scoring lower TPSs, and 73-10 the highest. Pairwise comparison found the greatest similarity between 22C3 and 28-8, with SP263 scoring slighter higher TPSs on average. Inter-pathologist scoring of ICs was very poor (k 0.08 – 0.28), with the highest range achieved using the SP142 clone. The authors concluded that IC remains challenging, but for TPS, the 22C3, 28-8 and SP263 could be considered interchangeable, despite the higher sensitivity of the latter.

A number of other larger studies published at a similar time showed similar results. Batenchuk *et al.* 2018 looked at 1,930 samples of various cancer types, including 412 NSCLC, with paired 22C3 and 28-8 staining already performed.[344] NSCLC cases had a TPS scored using negative, weak and strong positive categories (<1%, 1-49% and ≥50% TPS) and 5% categories of TPS. In terms of absolute TPS difference, less than 10% difference between clones was noted in 94.6% of cases. Both clones scored higher or lower at points, with 22C3 slightly more often scoring higher TPSs in NSCLC (10.9% vs 6.6%) with overall concordance between the clones at various cut-offs generally being very high (0.90-0.95). Villaruz *et al.* 2019 looked at 302 NSCLC samples and compared SP263 and 22C3 on the same platform (Benchmark Ultra) as LDTs, and included cytology specimens as well as small biopsies and resections.[305] Two approaches were used to score TPS; one as negative, weak and strongly positive, and one using ≥1%, ≥5% or ≥10% cut-offs. They found SP263 to be on average slightly higher scoring than 22C3, with SP263 placing more cases in the ≥50% group (60 vs 40 cases) but overall good concordance (0.88). This group also had clinical outcome response for 44 patients treated with either pembrolizumab, nivolumab or atezolizumab. Using best-overall-response, they found a cut-off of ≥1%, ≥5% or ≥10% was associated with superior outcome if assessed by 22C3, but only for ≥10% by SP263, though the authors note the variation in drugs, line of treatment, and the fact not all approaches were, at that point, in keeping with standard of care. This group again highlights the high, but imperfect, correlation between 22C3 and SP263, and has the data to highlight these have potential clinical ramifications.

Velcheti *et al.* 2018 looked at a cohort of 6024 patients with NSCLC that had been tested with PD-L1 by either 22C3, 28-8, SP142, or LDTs that included the SP263 clone.[345] PD-L1 TPS were captured and categorised as negative, weak or strong positive (<1%, 1-49% and ≥50%). SP142 again consistently scored more cases as negative or a lower TPS, but there was no significant difference between 22C3 and 28-8. Using a ≥50% cut-off, 22C3 found 32.8%, 28-8 found 31.8%, SP263 20% and SP142 10.3% of cases positive. (LDTs as a single group found 22.6% of cases positive at ≥50%). These data would suggest further evidence of 22C3 and 28-8 as equivalent, and SP263 as not, but the SP263 was included as LDTs in this paper, and as previous work has shown, without sufficient optimisation LDTs can be quite unreliable. However, the question of the similarity between these 22C3 and SP263 is further examined by two papers from the same group, Munari *et al.* 2018 and Munari *et al.* 2019.[346, 347] The former constructed TMAs from 198 NSCLC cases and stained for 22C3 and SP263 and TPS scored and categorised with either a ≥1% or ≥50% TPS cut-off by two blinded pathologists, with a minimum of 30% tumour content and 5 cores for each case being available. 22C3 was used both on the validated platform and on the Ventana platform as a LDT. Using a ≥1% dichotomous cut-off, the average number of cases classified as 'positive' was 23.7%, 31% and 40.3% for the 22C3, 22C3 LDT

and SP263 respectively, and 7.6%, 9.8% and 15.6% using the ≥50% cut-off for the same, again highlighting the increased average TPS for SP263 vs 22C3. Overall concordance between 22C3 and SP263 was average (k 0.518, 0.390 for each pathologist), rising slightly when comparing SP263 to the 22C3 LDT (k 0.624 and 0.572), and similar when comparing the two 22C3 assays (k 0.595 and 0.583). Inter-observer agreement, overall, was also modest (k >0.6). Munari *et al.* 2019 looked at 165 resected NSCLC cases (some of which from the previous study) from which TMAs were constructed. PD-L1 was stained for by 22C3, SP263 and E1L3N and a TPS scored as before. Only cores with >30% tumour content were included, and only the highest scoring cores for any case were used, which due to heterogeneity is a potential flaw in this study. In addition, direct comparison was made between E1L3N and 22C3, and E1L3N and SP263, but not directly between the two clinically validated clones as per their previous study. Comparison of E1L3N and SP263 showed good correlation at ≥1% cut-off (0.95) and ≥50% (0.98) with 5.4% of cases differently categorised. E1L3N and 22C3 showed less strong correlation at ≥1% cut-off (0.77) and ≥50% (88.2) with 31% of cases differently categorised. SP263 was therefore considered closer to E1L3N with stronger intensity of staining than that seen by 22C3, and not concordant to 22C3, though no direct comparison was made. Inter-observer concordance at ≥1% varied for each clone (k 0.73-0.87) as well as for ≥50% TPS (k 0.76-082). The authors generate a potential flow chart for assessing PD-L1 that potentially involves staining a sample multiple times for PD-L1 by different clones which seems impractical in the routine clinical setting. Despite this, however, there is again the observation that SP263 is more intense, stains more cells, and is therefore not wholly interchangeable for 22C3, particularly at the ≥1% cut-off threshold.

Tseng *et al.* 2018 looked at 211 NSCLC specimens from both cytology and histology and stained matched specimens for both 22C3 and SP263.[348] PD-L1 was assessed by two pathologists and a TPS score generated. Using a ≥1% cut-off, 22C3 found 47.4% of patients to be positive, and SP263 27%, and for a ≥50% cut-off 12.8% for both clones. Concordance between the matched specimens increased with higher cut-offs, from 76.8% for ≥1%, 81.5% for ≥10%, 90.8% for ≥25% and 94.3% agreement for a ≥50% TPS cut-off. 34 patients also had data for patient outcome after treatment with nivolumab or pembrolizumab, with the only significant factor for improved overall response rate being strong expression by SP263, but not by 22C3. In addition PFS was significantly different using strong expression by SP263 (p=0.008) but again, not for 22C3 (p=0.061). This paper highlights important recurring findings: lower cut-offs, particularly ≥1% are difficult to achieve concordance, and differences between 22C3 and SP263 have potential clinical ramifications in regards their ability to determine positivity. Saito *et al.* 2018 which studied 420 resected NSCLC specimens, which were stained for both 22C3 and 28-8 and scored to give a TPS, found average concordance between them, though higher concordance at lower TPSs (≥1%, ≥25% and ≥50% cut-off correlation k 0.763, 0.677

and 0.643 respectively) with 22C3 consistently returning higher TPSs and more positive cases, indicates 28-8 and 22C3 are not equivalent.

2.2.0.3 Small studies comparing multiple PD-L1 clones

Kim *et al.* 2017 looked at 97 resected NSCLC cases split into two groups, a 'training set' utilising TMAs, and a 'validation set' utilising whole slides.[288] SP263, 22C3, SP142 and E1L3N were used to stain for PD-L1 and the analysis performed by two pathologists, whom gave TPS results in 5% increments, and also scored for immune cells for SP142. In their hands 22C3 had the highest TPS on average and the most intense membranous staining, followed by SP263 and E1L3N, with again SP142 having the lowest TPS scores. The clones were compared pair-wise, and for TPS cut-offs of 1-25% 22C3 and SP263 had reasonable concordance (k >0.7) but much poorer for ≥50% cut-offs (k 0.467), with EL13N showing very good concordance with SP263 (k = 0.905), with 8% of cases being re-classified depending on the clone used. Interestingly they note that substituting the validated cut-off of 'positive' for one clone for a different one for another actually reduces agreement, with the authors concluding each clone should indeed use its respective cut-off point. They also had 50 matched cases with both TMAs and whole slides, of which only 14 cases were positive for PD-L1, and of which seven had differences in TPS, one significantly so (90% vs 15% TPS).

Fujimoto *et al.* 2017 compared SP263, 28-8, 22C3 and SP142 in 40 NSCLC patient samples, and also captured clinical outcome data.[349] They categorised the samples as negative, weak or strongly positive (<1%, 1-49% and ≥50% TPS), and four pathologists were used to generate concordant scores, with re-staining and re-examination if necessary to reach a consensus, with each clone compared pairwise. They found reasonable concordance between 22C3, SP263 and 28-8 (k 0.64-0.71), albeit with slightly lower TPSs for 28-8, but significantly lower TPS scores and poor concordance when comparing SP142 to other clones (0.39-0.55). Using ≥50% as a cut-off and excluding SP142, the agreement rate was for 90% of cases, but much poorer using the ≥1% cut-off (65% of cases). Again this study highlights a reasonable but far from perfect concordance between clones, with particularly respect to the ≥1% cut-off. Unlike many of these studies, the authors also captured clinical outcome data, with some patients receiving nivolumab (it is not clear what percentage of patients did, and the cohort included patients with tumours harbouring *EGFR* and *ALK* mutants/translocations) and showed that generally speaking, higher PD-L1 expressing patients had superior outcomes to weak PD-L1 expressing patients, and better again than those with negative samples. PFS times for strong positive versus PD-L1 negative were significantly better if determined by SP263, 22C3 or 28-8, but non-significantly longer by the SP142 clone, with non-significant findings for all clones in the weak expressers. Between the clones, excluding SP142, there were four

discordant cases using the ≥50% cut-off, and for these the response rate was poorer than the concordance cases, but the authors admit these numbers are too small to generate any conclusions

Krawczyk *et al.* 2017 used 48 NSCLC resections to study PD-L1 with 22C3 and SP142 (as well as *ALK* and *EGFR* status) and scored a TPS and categorised as negative, weak or strong positive (<1%, 1-49% and ≥50% TPS).[350] 22C3 categorised more cases as positive than SP142 using a ≥1% cut-off (72.9% vs 60.4%), ≥5% cut-off (66.7% vs 39.6%) and ≥50% cut-off (45.8% vs 22.9%), the latter two cut-offs being statistically significant, with SP142 having weaker staining as well as fewer positive tumour cells. Xu *et al.* 2017 looked at 135 NSCLC samples, with each having matched samples stained for both 22C3 and SP142.[351] A TPS was generated for both clones, with 22C3 using negative, weak or strong positive (<1%, 1-49% and ≥50% TPS) and SP142 used TC1, TC2 or TC3, although various TPS cut-offs were used for direct comparison of clones. SP142 demonstrated generally weaker staining, with poor concordance of the clones using the 22C3 scoring algorithm (k 0.481) with SP142 generally scoring lower TPSs, and even worse concordance using the SP142 scoring algorithm for both (k 0.324). Pang *et al.* 2018 also demonstrate the lower sensitivity of SP142 by comparing 84 NSCLC cases of SP142 and SP262 that found a low concordance (k 0.53).[352] Despite using a ≥25% cut-off for SP263 versus ≥1% cut-off for SP142, SP263 still consistently returned more positive cases.

2.2.0.4 PD-L1 clone concordance in other studies

A number of other studies for which PD-L1 clone comparisons was not the primary focus of the paper nonetheless reported data on clone concordance.

Scheel *et al.* 2018 was a study primarily setup to compare interlaboratory concordance by comparing 22C3, SP263, 28-8 and SP142 on both validated platforms and by LDTs in 21 NSCLC cases used to create TMAs.[353] Their main finding was a high concordance between sites, and that with sufficient effort LDTs can achieve comparable quality. However, their findings for PD-L1 TPS was also comparable to some of the previous studies. Multiple pathologists scored the cases by TPS into one of six categories (from <1% to ≥50%) or one of three (1%, 1-49% and ≥50% TPS). As before a similarity in type of staining was seen for the first three clones (linear membranous staining) but a difference for SP142 (granular and/or linear deposits) a finding widely commented on in the studies mentioned here. They found 22C3 and 28-8 to be equivalent, SP263 to have a higher average TPS score, and SP142 to have a lower average TPS score.

Skov & Skov 2017 was primarily a study looking at comparing cytology and histology but also found a strong correlation between 22C3 and SP263 in both histology specimens at various TPS cut-offs (0.93-0.99) and cytology specimens (0.93-0.98).[307] Beck *et al.* 2019 looked at 80 samples from NSCLC

biopsies, stained for both 22C3 and SP263 and a TPS score was generated using cut-offs at ≥1%, ≥25% and ≥50%.[354] Overall correlation between the clones was good (0.892), though slightly poorer at each cut-off, respectively 0.878, 0.698 and 0.790. Even unconventional sampling techniques for PD-L1 appeared to show good concordance between clones: Lozano *et al.* 2019, primarily comparing smears to histology also compared 22C3 and SP263 in 55 matched cases and found good concordance (0.93) using a cut-off of ≥1% or ≥50% on cytology specimens, and ever higher concordance on the histology specimens (0.98).[317]

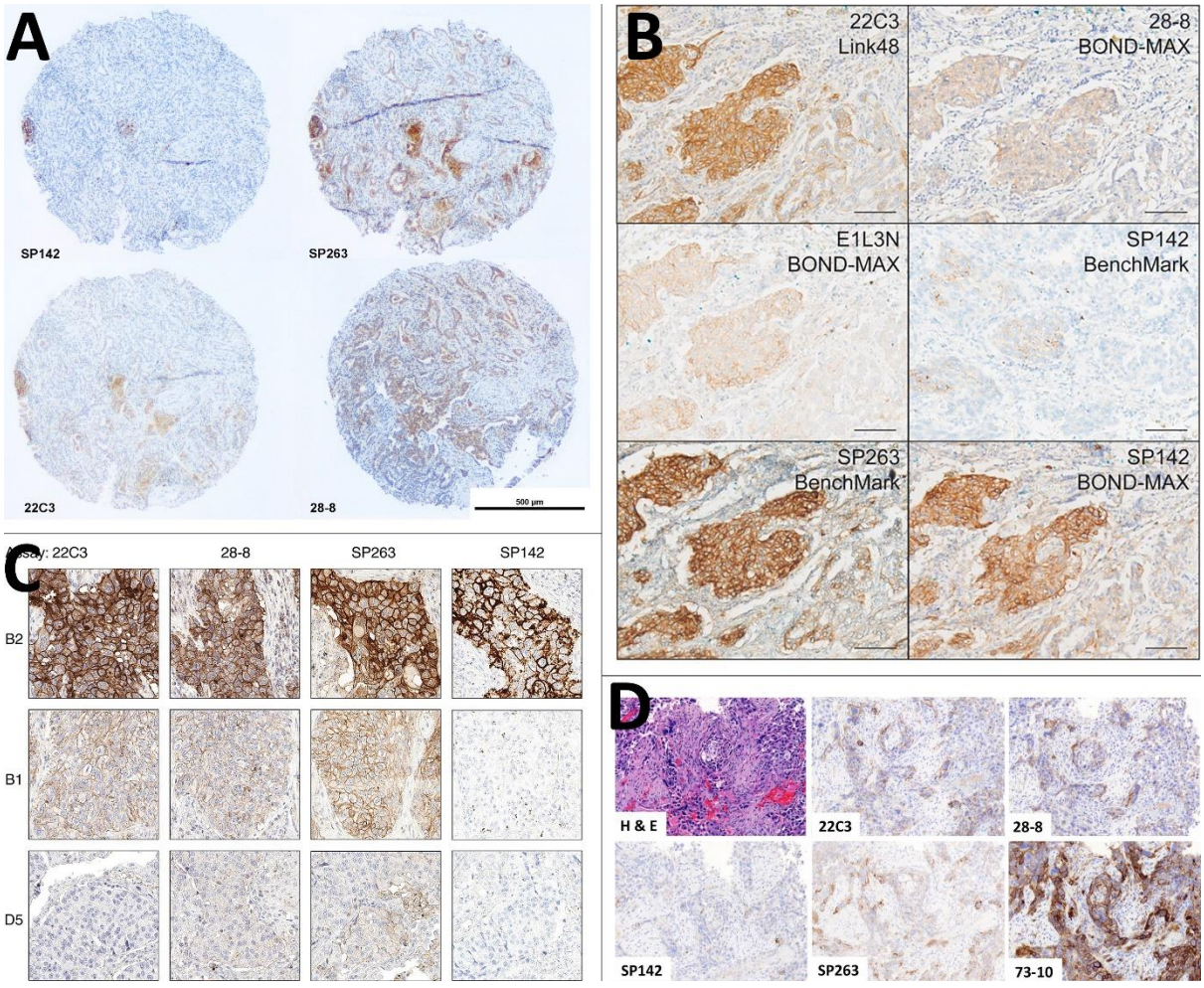


Fig 2.2.0 Composite image of four studies that compared different PD-L1 IHC clones in NSCLC. Despite different approaches and specimens used between the studies, they all illustrate similar findings: SP142 stains fewer tumour cells, the similarity of 22C3, SP263 and 28-8, with a slightly cleaner, more intense stain for SP263. **A** – Hendry *et al.* 2017 **B** – Soo *et al.* 2018 **C** – Scheel *et al.* 2018 **D** – Tsao *et al.* 2018.

Sheffield *et al.* 2016 primarily looked at heterogeneity of expression, but used four PD-L1 clones (SP142, 28-8, E1L3N and RBT-PDL1) and simply stated the agreement was high, but gave minimal data: a Kappa of 0.67.[355] Capizzi *et al.* 2018, already considered in this review for its cytological comparison also stained 50 biopsies with 22C3, 28-8 and SP263.[316] They compared each clone in turn to a 'gold standard', which was in fact whichever clone stained any given sample with the highest TPS, which is of dubious logic. By their approach they found concordance to the gold standard to vary (22C3 K 0.554, 28-8 K 0.698, SP263 0.908) which probably just reflects the fact SP263 stained with the greatest TPS in most cases, and, unfortunately no further direct analyses was made.

Despite the large number of studies comparing PD-L1 clones, the massive variation between studies in terms of specimens used and scoring approaches, and the apparently discordant results between many of them, several trends have emerged in regards to certain clones. The overall evidence suggests that SP142 consistently scores lowers TPSs, that 22C3, 28-8 and SP263 clones are generally similar, although SP263 trends towards higher TPSs than other clone (Fig 2.2.0). LDTs can achieve comparable results, but only if significant input into their optimisation is utilised. Inter-pathologist concordance is highly variable, with a reasonable level achievable for TPS, but generally very poor for IC scoring. Finally, despite several papers showing apparent concordance of clones, the majority show agreement of less than 90%, and most don't have clinical outcome data. For those that do, there is enough difference between the various clones to suggest that even the minor differences between 22C3, SP263 and 28-8 are likely to have clinical ramification in only a minority of cases. Perhaps the biggest limiting factor of these studies is the lack of consistency between them: some are large international studies with expert pathologists, some are smaller single centre studies, some use whole section slides and matched cases, and some compare cohorts using TMAs and so forth. This highlights in general the problem of attempting to rationalise the optimal approach to using PD-L1 IHC as a predictive biomarker. Furthermore, inherent limitations of immunochemistry limit its ability to provide consistent results, and may go some way to explaining why the predictive power of PD-L1 is not more reliable than it is.

2.2.1 Analytics- limitations of immunochemistry

There are generic limitations to immunochemistry, not least the fact it was designed as a diagnostic aid that allowed for the presence of specific proteins to be detected and their spatial profile to be determined and not as an inherently quantitative technology. Therefore attempts to use IHC to give a percentage score is potentially limited. As seen above, in additional to the widely used TPS, attempts have included various different cut-offs, as well as H-scores and other scoring methods that include intensity of staining and cell type. Ultimately the more complex approaches fail to be

used routinely, probably because of the time consuming nature that shows no better concordance. In addition, the PD-L1 protein itself can exist in states which may render it less susceptible to detection by IHC. Post-translational modification (PTM) of proteins is a well-recognised biological pathway for changing or optimising the function or structure of a protein, and PD-L1 can undergo various PTMs, though these have not been fully characterised. A comprehensive overview of PTM of PD-L1 is outside the scope of this literature review, but N-linked glycosylation, serine/threonine phosphorylation, ubiquitination and acetylation all play roles in the modification of PD-L1.[150, 356, 357] Glycosylation is a particularly important aspect: non-glycosylated PD-L1 has a relatively short half-life (4 hours) and will undergo phosphorylation before being degraded by the ubiquitination/proteasome pathway.[356, 358] N-linked glycosylation stabilises the half-life of PD-L1 and plays a role in PD-L1/PD-1 interaction.[150, 356, 359] Few studies have looked at the effect of PTMs on the detection of PD-L1 by IHC, but Morales-Bentazos *et al.* 2017 showed with the use of MS (mass-spectrometry) that heavily glycosylated PD-L1 cannot be detected by IHC.[360] The effect of all PTMs on detection by IHC is unknown, but may at least partially explain why the absence of PD-L1 in some specimens may still see clinical benefit from PD-1/PD-L1 IMs.

PD-L1 (and PD-1) are largely studied in the context of membrane-bound forms. However, soluble PD-L1 (sPD-L1) can also exist. It is far from fully characterised; it is uncertain if generation is via proteolytic cleave of the membrane bound form, or via alternative splicing as seen for TNF-α and soluble CTLA4 respectively.[361, 362] It has been shown that sPD-L1 can come from both tumour cells and immune cells,[363, 364] with a potentially distinct regulatory mechanism to that of PD-L1.[365] The role of sPD-L1 is uncertain, though there is evidence it can bind to PD-1 in the context of lung cancer[366] and generally speaking its presence is associated with poor OS in various cancers, including NSCLC.[367-369] sPD-L1 may therefore be an important component in achieving immune escape. Detection of sPD-L1 has largely been made in the context of plasma samples, rather than from tissue samples, but the consistent finding is that there is no correlation between sPD-L1 levels and PD-L1 expression by IHC in NSCLC,[370] RCC (renal cell carcinoma),[371] Gastric cancer,[372] ovarian cancer[373] or HCC (Hepatocellular carcinoma) .[374] To further muddy the waters, splice variants of PD-L1 resulting in secreted variants that appear to homodimerise but still infer an active immunosuppressive function have been described, and require potentially distinct IHC approaches to be appropriately detected.[174]

These variable forms of the PD-L1 protein may have greater or lesser roles in regulating the TME, and may be very susceptible, or, not at all susceptible, to PD-1/PD-L1 IM therapy, the prediction of which is ultimately the most important consideration for the use of PD-L1 IHC. However, it is important reminder than if reliable detection and quantification of a membrane bound variant of

PD-L1 is challenging enough, we must also remember the myriad other forms that may not be detectable by conventional means.

2.3 Heterogeneity of PD-L1 expression

So far in this review I have considered the various ways in which pre-analytics may affect PD-L1 detection (a false representation of what is genuinely there) and how different clones can both under or over represent PD-L1 expression (incorrect results from what protein is visible), and that the interpretation of the assays can also lead to poor concordance (incorrect results from interpretation error). The final major hurdle for achieving the greatest accuracy with PD-L1 IHC is that of tumour heterogeneity, and the heterogeneous expression of PD-L1 therein. PD-L1 expression is typically scored as a percentage, which by definition will result in certain tumours not uniformly expressing PD-L1 in all cells (Fig 2.3.0). The difficulty in studying heterogeneity of tumours is compounded by the multiple ways in which a tumour can be sampled: a core needle biopsy, an EBUS FNA, or a resected lung all provide vastly different quantities and qualities of tumour cell content, may use different fixatives and so forth. I have attempted to extricate the specific details of some of these aspects from papers considered within this review, but a considerable number of relevant studies use both cytology and histology to study heterogeneity, and thus bring a number of confounding factors with them.

I have split PD-L1 expression heterogeneity into 4 distinct types: intra-tumoural (variation within a single tumour site), inter-tumoural (variation between different tumour sites), temporal (change over time) and what I termed 'iatrogenic heterogeneity', in that treatments including XCT, XRT (radiotherapy) and IM can all variably effect PD-L1 expression. It is worth noting most of these heterogeneities can be confounding factors of each other; for example, metastases tend to occur over time, but also represent distinct background tissues for tumours to invade into, thus having elements of both inter-tumoural and temporal heterogeneity, or in studies that used matched samples but from different time points. Studying heterogeneity is therefore challenging, but I have chosen to split papers into the type of heterogeneity that they predominantly provide data on.

2.3.0 PD-L1 Heterogeneity – Intra-tumoural heterogeneity

Intra-tumoural heterogeneity, that is, expression of PD-L1 that varies across a single site of tumour, is a significant factor that has the potential to account for 'sampling error' in that a biopsy from a single site within a heterogeneous tumour may produce a falsely high or falsely low PD-L1 expression when considering the total PD-L1 expression within the tumour. (Fig 2.3.1). The underlying biology as to why this variation in PD-L1 expression occurs (or doesn't) and what impact it

has on clinical response is not fully characterised. Intra-tumoural heterogeneity may partially explain why apparently PD-L1 negative tumours respond to PD-1/PD-L1 IM therapy or why positive tumours respond poorly, as small biopsies and samples do not fully represent a tumour's PD-L1 expression levels.

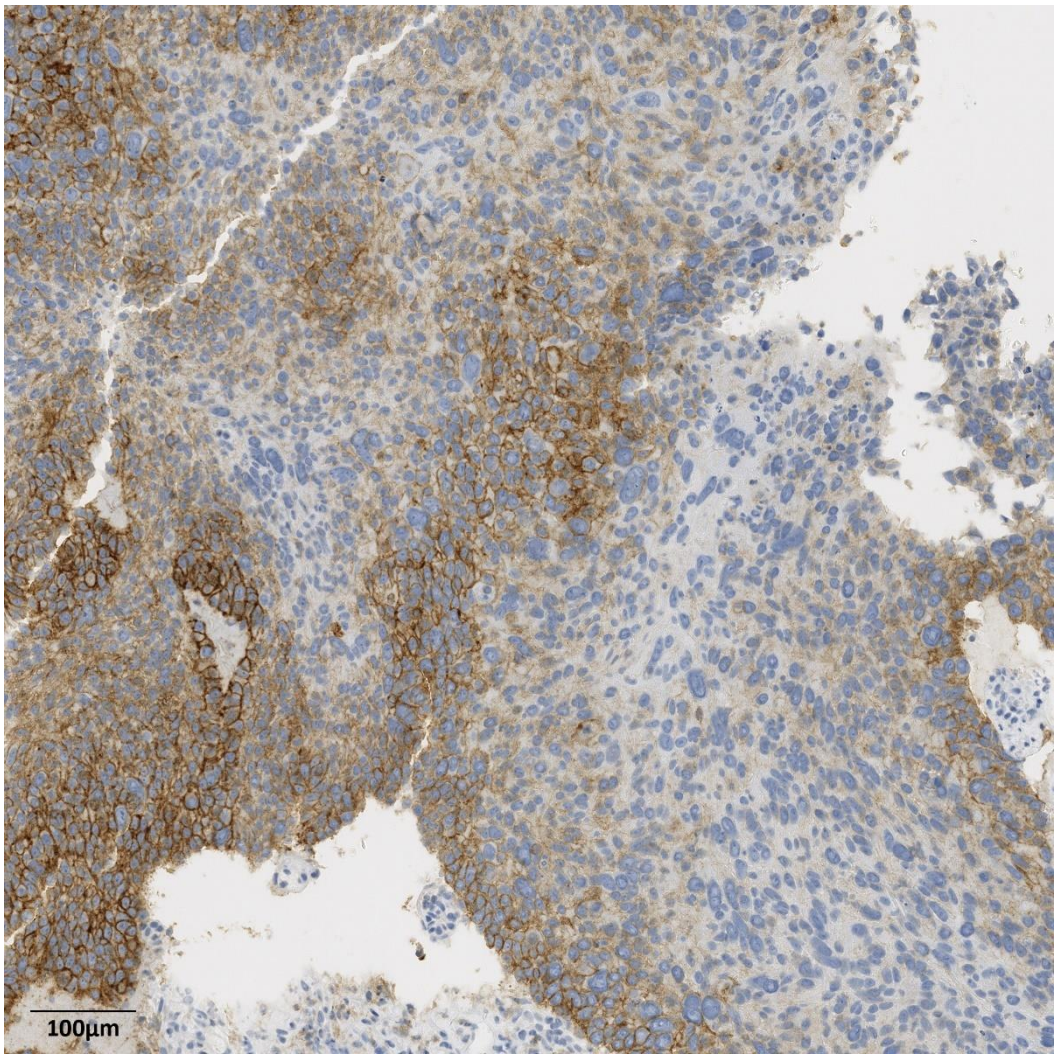


Fig 2.3.0 Heterogenous PD-L1 expression in NSCLC in which strength of expression varies across the field and is absent from some tumour cells. Thus even in this single field, tumour cells with no PD-L1 expression, and both weak and strong intensity of PD-L1 expression can be observed.

One of the earliest papers to consider heterogeneous PD-L1 expression within a single sample as a potentially relevant factor in tumour related mediation of immune responses in NSCLC is Konishi *et al.* 2004, whom studied 52 NSCLC sections for PD-L1 (MIH1 clone) and compared these to various clinical outcomes and the presence of immune markers.[375] They found focal expression of PD-L1 in all their cases, and in regions with high PD-L1, they found lower levels of PD-1 expressing TILs, and

proposed this contributed to the tumour induced immunosuppressive features of PD-L1 positive tumours. However, at this point in history the clinical impact of these findings was not apparent, and very little is in the literature until some 10 years later when the IM therapeutic options were in clinical trials. The potential impact heterogeneity had on sampling error was initially considered in regards to intra-tumoural heterogeneity, and these will be considered here.

2.3.0.0 Quantitative immunofluorescence to assess PD-L1 intra-tumoural heterogeneity

Velcheti *et al.* 2014 looked at two cohorts of NSCLC totalling 544 patients,[376] from which TMAs were constructed, with mRNA for PD-L1 RNA assessed by RNAScope, and IHC and immunofluorescence (IF) for PD-L1 protein using the 5H1 antibody, with QIF using an automated quantitative analysis (AQUA)[377] approach, which allows for measurements of a protein concentration within defined cellular compartments. This study largely focused on the QIF and mRNA findings, and showed that 25% and 36% of patients, from each cohort respectively, demonstrated PD-L1 expression above a defined cut-off of expression, with 53% and 51% of patients demonstrating the same for PD-L1 mRNA. In regards to heterogeneity, at least two TMA hotspots were evaluated from each specimen, and the results for each could be compared, and generated linear regression coefficients of 0.53 and 0.59 for each cohort, demonstrating intra-tumoural heterogeneity of PD-L1 expression even within small TMAs (0.6μm diameters) from the same specimens. The authors acknowledge that this is not something to be dismissed, though they say it may not present a major issue, but admit that TMAs may under-represent genuine heterogeneity, and these are not in keeping with routine clinical diagnostic specimens. McLaughlin *et al.* 2016 also used the AQUA method of QIF to assess PD-L1 as well as IHC by use of the E1L3N and SP142 clones.[378] 49 NSCLC cases were included and were composed of TMAs and matched whole-slide sections. As anticipated from other studies, there was poor concordance between E1L3N and SP142 in both assays, significantly so for QIF. Heterogeneity across the samples for PD-L1 expression was considerable and variable: co-efficient of variations (COV) between field-of-views (FOV) for individual cases ranged from 6.75% to 75.24% for E1L3N and 12.17% to 109.61% for SP142. Rehman *et al.* 2017 used the AQUA method of QIF using the SP142 clone, but in their case looked at 35 resected cases of otherwise untreated NSCLC on three blocks of primary tumour for each case.[379] By doing so intra-tumoural heterogeneity within each block and between the blocks could be studied by utilising whole-tissue sections. Five pathologists scored each whole-tissue section independently to generate a TPS, whilst a modified AQUA QIF approach was used to analyse the IF stained samples across the large tissue sections. No categories were used, so ICC only was captured, but these showed good inter-pathologist concordance (ICC 94%) but poorer inter-block concordance (ICC 75%), which was slightly improved using the QIF (ICC 95% and 88% for the same). Stroma scoring for PD-L1 was poorer in both respects (Inter-pathologist ICC 27%, Inter-

block 75%). QIF only was used to assess intra-block heterogeneity, and generally found that variation within a block was greater than between blocks (91% vs 9% field of view variance), suggesting that a single block is sufficient to provide an overall assessment of heterogeneity.

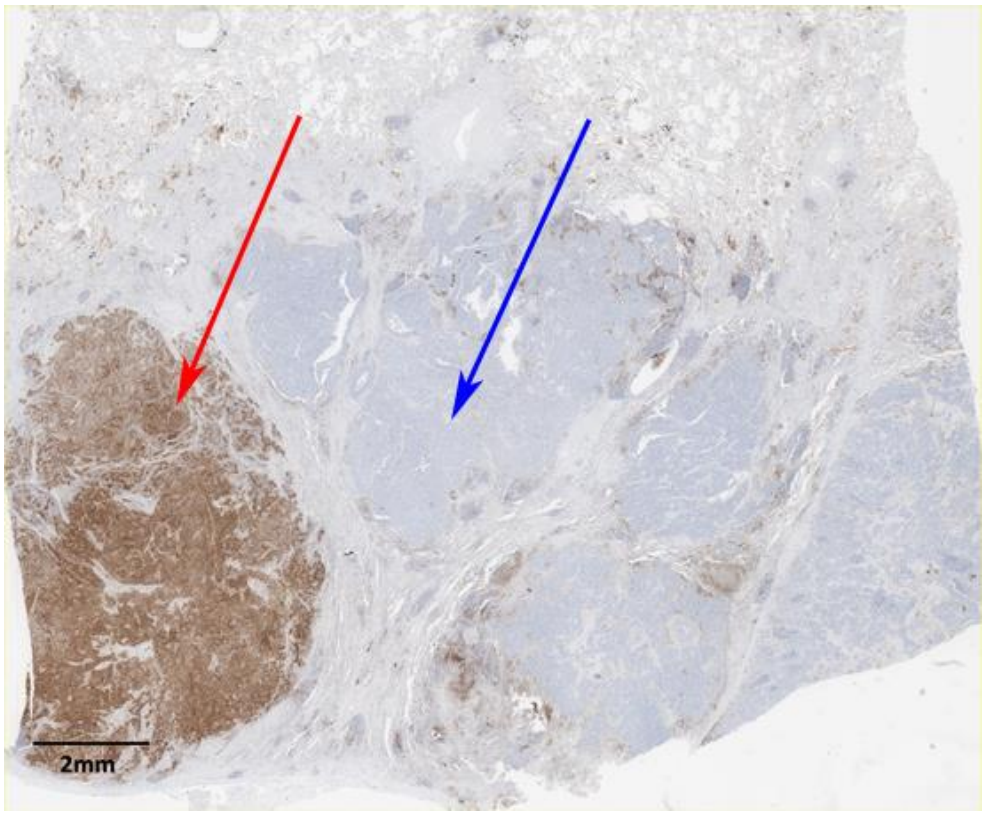


Fig 2.3.1 Intra-tumoural PD-L1 expression heterogeneity which illustrates the potential for sampling error. In the bottom left corner an area of strong uniform PD-L1 expression of tumour cells is seen (red arrow), but the majority of the rest of the tumour is negative for PD-L1 (blue arrow).

2.3.0.1 Using TMAs to assess PD-L1 intra-tumoural heterogeneity

Several other studies used TMAs as a measure of looking at intra-tumoural heterogeneity. Casadevall *et al.* 2017 constructed TMAs from 144 NSCLC resections by sampling two distinct regions from each primary tumour to generate 288 cores in the array.[380] SP142 was used to score for PD-L1, and the TC 1/2/3 and IC 1/2/3 scoring methodology utilised. ADC and SCC were considered separately due to both their distinct morphologies, and the differences in sampling the cores to maximise intra-tumoural heterogeneity (SCC used distance, ADC were selected based on variable morphology by H&E (haematoxylin and eosin)). Between the matched cores in the TMA, ADCs

demonstrated significant discordance of both TC (kappa 0.465) and IC (k 0.260), and this was even poorer for SCCs (TC = K 0.275; IC = k 0.124). When using a single cut-off of ≥5% TPS (TC2), these were improved, (ADC K 0.585, SCC k 0.543). These results overall demonstrate significant intra-tumoural heterogeneity, even if only using two TMA cores, although the tissues themselves were from 2004-2015, with some potentially 13 years old at the time of analysis. In addition, the use of SP142 means these are likely to generally underestimate PD-L1 expression. Li *et al.* 2017 constructed TMAs from 190 NSCLC patient specimens from 2008-2010, and utilised two cores (2mm) from each case to be stained with 22C3 and scored to give a TPS and classified into negative, weak or strong positive (0%, 1-49%, 50%). Two pathologists scored each case (inter-pathologist agreement K 0.94), with discrepant cases reviewed to achieve consensus.[381] Using a ≥1% cut-off only 36.9% of cases were positive, likely a result of using older tissue samples. Using the categories as above, and comparing the matched cores, the discordant rate was 18.9% (K 0.630), with a fairly equal split between cores over or underestimating PD-L1 expression. Using a single cut-off the discordant rates improved to 13.2% (K 0.711) and 6.8% (K 0.685) for ≥1% and ≥50% respectively. When comparing the TMAs to their matched whole tissue section, and looking for completely negative versus any positivity, the discordance rate was 2.6%, including, oddly, 3 cases positive in the TMAs, but entirely negative in the whole sections. Notwithstanding the use of older tissue, these results indicate again the relatively small but significant issue of intra-tumoural heterogeneity.

A number of papers, using similar approaches, found PD-L1 intra-tumoural heterogeneity to be rather more prevalent. Gniadek *et al.* 2017 constructed TMAs from 150 NSCLC cases, and sampled 4 cores (0.6mm diameter) from each specimen at random, with at least 3 in each case adequate for analysis. SP142 was used with a ≥1%, ≥10% or ≥50% TPS cut-off category utilised.[382] Based on a ≥1% cut-off, 39% of cases were positive, dropping to 11% for a ≥50% cut-off. Comparing the cores to each other showed that 13.3% of all cases had at least one core entirely negative and another positive for PD-L1, with heterogeneity of some description observed in many cases. The presentation of results in this study is not as comprehensive as it could be, but they do conclude that when using a ≥1% cut-off, a single TMA core will only be concordant in 87% and 90% of cases in ADC and SCC respectively, and 85% and 95% of cases when using a ≥50% cut-off in the same, and thus a single biopsy is likely, in a large number of cases, to not be fully representative of PD-L1 expression. Nakamura *et al.* 2017 generated Spiral Arrays, a technique similar to TMAs, but designed specifically to characterise tumour heterogeneity and optimised on NSCLC samples, and utilises a reeled layer of tissue.[383, 384] Prospective patient specimens of NSCLC were collected, as well as a small number of retrospectively collected SCLCs and large-cell carcinomas, totalling 138 cases. The spiral arrays were stained with 28-8 and a TPS scored and categorised using ≥1%, ≥5%, ≥10% and ≥50% cut-offs, and discordance

between pathologists discussed to reach a consensus score. Each array was split into eight segments to assess heterogeneity, with six selected to be compared to their matched whole tissue section. Positivity across the cohort using ≥1% was 30.4%, dropping to 15.2% for a ≥50% cut-off. Again data might have been presented more comprehensively, but the main finding was that between 50-76% of positive cases, depending on which cut-off was used, had half or more of their respective eight segments return a PD-L1 negative result. Their conclusion that intra-tumoural PD-L1 expression heterogeneity is a common finding is valid, but the translation of this in clinical terms when using Spiral Arrays is uncertain. Hendry *et al.* 2018, as already described, looked at a large number of previously constructed TMAs of NSCLC cases and looked at a variety of outcomes.[297] A short section in the results found that 515 cases (stained either with 22C3, SP263, 28-8 or SP142) showed the ICC between cores was quite variable, depending on clone, from 0.355 (22C3) to 0.600 (SP142). Using the cut-offs defined for each clone, overall they found a discordance rate of 12% between the cores as a measure of intra-tumoural heterogeneity. Gagne *et al.* 2018 also used TMAs to study 214 resected ADCs from between 2003 and 2012, and extracted up to five cores (1mm diameter, minimum of 3 cores) for every case to represent diverse areas of morphology. E1L3N was used and a TPS score generated, with ≥1%, ≥10% and ≥50% cut-offs used to define positivity. Using a ≥1% cut-off, 67.7% of patients were positive, reassuringly so given the older specimens used, dropping to 37.8% for a ≥50% cut-off. Variation between the cores using a ≥1% cut-off demonstrated heterogeneity in 39.8% of cases, and 32.4% and 22.4% of cases for ≥10% and ≥50% cut-offs respectively. When focusing on samples with at least one core negative for PD-L1, heterogeneity with respect to each cut-off was seen in 55.5%, 41.4% and 26.5%, demonstrating a considerable risk of false negative results. They also noted some variation in PD-L1 expression between morphological subtypes of ADC, such that mean TPS scores for solid and micropapillary were higher than acinar, papillary and lepidic patterns. This phenomenon is not well characterised, but has been noted in other studies, for example Koezuka *et al.* 2019, which looked at a variety of aspects relating to general ADC heterogeneity, noted that in their small number of PD-L1 positive tumours (6 cases, 22C3), expression was heterogeneous in all cases; specifically the solid and papillary components were positive, but the acinar components were not.[385] Evans *et al.* 2020 looked at a large number of NSCLC (10,005 cases, 22C3) and found higher PD-L1 TPS in solid and micropapillary, and low expression in lepidic, mucinous and papillary adenocarcinomas.[318]

Munari *et al.* 2017 was the first of two papers by the same group that constructed TMAs from resected NSCLC as surrogates of small biopsies in a bid to assess intra-tumoural heterogeneity, and ultimately provided data on the minimum required quantity of tissue to overcome this phenomena. In this first paper SP263 was used to create TMAs utilising multiple cores (up to five) from 239

surgically resected NSCLCs and scored to classify into one of five groups (0%, 1-4%, 2-5%, 10-49%, ≥50% TPS).[386] 40% of the cases were positive using a ≥1% cut-off, of which 52% were discordant between cores. When using a ≥50% cut-off, the discordance was higher (66%). When considering the entire cohort the discordance rate was lower (20%, K 0.53, for 1%, 7.9% K 0.48 for 50%). Based on these data the authors suggested at least four cores are required to attain an accurate result. The second paper was Munari *et al.* 2018 in which they constructed TMAs from NSCLC patients, consisting of 268 patients who had no treatment prior to surgical resection. They took five cores (1mm diameter) from each specimen to construct the TMAs, taken from 'diverse areas'. SP263 was used to stain for PD-L1 and scored independently by two blinded pathologists, with a third utilised in discordant cases, and a TPS score was generated and then grouped into negative, weak or strongly positive (<1%, 1-49%, >50% TPS).[387] Using a ≥1% TPS, 61% of cases were negative, and a ≥50% cut-off 90%, this high negative rate possibly accounted for by the fact that specimens were from 2003-2017 and therefore potentially 15 years old at the time of analysis. Inter-pathologist concordance was extremely good for ≥1% and ≥50% cut-offs (K 0.98 and 0.91 respectively). The majority of cases had five cores to assess, and based on cases that demonstrated heterogenous expression on the whole slide, (104 cases (38.8%)), concordance rates between the cores was 93% and 88% for a ≥1% and ≥50% cut-off respectively. Based on their findings, the authors also concluded that three cores were sufficient to provide enough tissue to adequately assess PD-L1 even in heterogenous cases. Whilst this appears to be true in this cohort, the fact that the inter-observer agreement was so high, and the fact their previous paper was from the same overall cohort of patients (in which they suggested four cores as a minimum) might suggest these specimens were particularly favourable in regards to this analysis, and the authors admit the limitations in using TMAs for studying heterogeneity.

2.3.0.2 Small biopsies to assess PD-L1 intra-tumoural heterogeneity

Several studies looked at PD-L1 intra-tumoural heterogeneity by comparing small biopsies of NSCLC with the corresponding matched resection block, on the basis that a block of NSCLC will be sufficient to provide an accurate PD-L1 expression assessment. Kitazono *et al.* 2015 looked at 79 patients with NSCLC who had a small biopsy prior to surgical resection, (including 12 (15.2%) who had needle aspiration of the primary tumour).[388] No data was given on the time between initial sampling and surgical resection, and so temporal heterogeneity is potentially a factor as well. PD-L1 was assessed using a research clone (Prosci-Inc 4059) and assessed by two pathologists to give a H-Score (0-300), presumably together to generate an agreed result. Using the H-score, samples were grouped as 0, 1-50, 51-100 or 101-300. Using a ≥1 score cut-off as positive, concordance between the biopsies and resections was good (92.4%, K 0.8366), but dropped if a score of ≥51 was used as a cut-off (83.5% and k 0.3969), with the authors concluding this was overall a reasonable level of agreement.

However, Ilie *et al.* 2016 found poor concordance in PD-L1 expression between biopsies and matched resected tissue.[389] They looked at 160 resected NSCLC specimens with matched biopsies of the primary tumour (although 12 (7%) were biopsies of metastatic deposits in regional lymph nodes) stained for SP142 and scored for both TC and IC. Using the SP142 scoring system (TC 1/2/3, IC 1/2/3) they found a discordance rate of 48% (k 0.218) for any positivity, slightly improving for high positive tumours (TC3/IC3) with a discordance rate of 20% (k 0.528), and similar if only using a TC score (discordance 20% k 0.396), with the interesting finding that in all cases the biopsies underestimated the PD-L1 expression levels. It might be anticipated that heterogeneity would see a number overestimated as well, but again this study does not mention the time between primary biopsy sampling and surgical resection, and as we have seen age of tissue is a known factor of under representing PD-L1 expression. In addition, this study used SP142 which has been shown to be problematic in NSCLC, but nonetheless provides data that intra-tumoural heterogeneity is a potentially significant factor. Gradecki *et al.* 2018 looked at 51 patients with pre-treatment biopsies and subsequent resections of their NSCLC from 2011 to 2014.[390] They used the SP142 clone to stain for PD-L1, though curiously in their hands, based on previous (presumably unpublished) work this was found to be concordant with 22C3 in ≥95% of cases. They gave virtually no information as to what this previous work entailed to arrive at this concordant rate, but as this review has noted, this is a much higher concordant rate than might be expected. They scored PD-L1 for a TPS and used <1%, 1-49% and ≥50% groupings. Using a ≥1% cut-off they found 58.8% of resections to be positive, which is perhaps much higher than might be expected with the SP142 clone, and 15.7% of cases were positive using a ≥50% cut-off, which is perhaps still a little on the high side for SP142. This point notwithstanding, they found four cases (7.8%) to be discordant between core biopsy and resection with the ≥50% cut-off, and twice this number discordant with a ≥1% cut-off, with an overall concordance rate recorded as 92.2% (K 0.7). Though the authors make the reasonable conclusion that based on their data a core biopsy may well be representative, this is yet another study using a clone generally not used in NSCLC (SP142) or, in the case of other studies, non-clinically validated clones, which does bring into question the applicability of these conclusions to routine diagnostics. Kim *et al.* 2019 also looked a pre-resection small biopsies and compared them to the matched resected tissue, but looked at these with three PD-L1 clones (SP263, SP142 and 22C3).[306] They looked at 46 NSCLC patients who met this criteria. Unfortunately, 18% of these included cytology specimens (PCNA (percutaneous needle aspiration) and EBUS-TBNA) but were not separated from the small biopsies in the analysis, and also included a potentially small but unspecified number of samples from regional lymph node metastases. These discrepancies notwithstanding, the paper focuses on the use of small biopsies for PD-L1 staining by three different clones, each scored for a

TPS and categorised with various cut-offs at ≥1%, ≥5%, ≥10% and ≥50%. 26 cases were stained for 22C3, and the concordance between the biopsies and resections ranged from 96% (K 0.649) to 73% (K 0.455) depending on cut-off. However the concordance was worse for SP142 (20 cases) with concordances ranging between 65% (K 0.146) and 80% (K 0.231). SP263 was the only clone used on all 46 cases, and had a concordance rate of between 72% (K 0.143) and 91% (K 0.292). The highest Kappa achieved was 0.649, with agreements generally poor. This may well reflect the mixed nature of the 'small biopsies' and relatively small number of samples. Bizarrely, the authors conclude that small biopsies are "reliable compared with… surgical specimens in patients with NSCLC", but confounding factors such as inter-tumoural heterogeneity and variation in sampling techniques could not be separated in this paper. Despite evidence that discordance is a significant issue, it is challenging to ascertain how much of this is due to intra-tumoural heterogeneity, but these papers do nicely illustrate the challenge of comparing various studies that all take such different approaches. For example, where most of these studies used retrospective tissue, often sampled many years prior to analysis, Tsunoda *et al.* 2019 took a prospective approach to compare small biopsies to their matched resections in 30 NSCLC cases.[391] In total they looked at 153 cases sampled by various means, which included a small number of cytology specimens (23 EBUS-TBNAs), and they made no comment as to whether any of the EBUS sampled specimens might have been from regional lymph node metastases, and therefore include an element of inter-tumoural heterogeneity. This notwithstanding, specimens were stained with 22C3 and each scored for a TPS and categorised into <1%, 1-49% and ≥50% categories. As might be expected, these categories were composed of roughly a third each of all specimens (34.6%, 31.4% and 31.4% respectively). Comparing the 30 cases of small biopsies to their resections, they found a concordance rate of 86.7%.

Elfving *et al.* 2019 combined previous approaches by using pre-operative biopsies and surgical resections, but also used the latter to construct TMAs to assess intra-tumoural heterogeneity.[392] 58 NSCLC cases from between 2006-2010 were selected, with two cores (1mm diameter) from each resection used to construct the TMAs. SP263 was used to stain the tissues and scored by two pathologists independently to generate a TPS, with a consensus reached in cases of disagreement, and positive cut-offs of ≥1% and ≥50% used for analysis. Using a ≥1% cut-off 48% of resections were positive, and 19% using the ≥50% cut-off, which is maybe slightly low for SP263, but possibly accounted for by the use of older tissue. The proportion of positive samples by each method was not significantly different, with the agreement between biopsies and resections found to be reasonable (K 0.69 and 0.83 for ≥1% and ≥50% respectively) with the inverse finding when comparing TMAs to the resections (K 0.83 and 0.69 for ≥1% and ≥50% respectively). Taking the resection score to be correct, biopsies were incorrect for 16% and 5% for each cut-off, and 9% for both cut-offs using the

TMAs. The authors conclude, quite reasonably, that small biopsies might therefore misclassify patients by PD-L1 score. Their approach suggests that possibly two smaller biopsies from separate regions of the tumour might be more accurate than 1 larger biopsy from a single site.

2.3.0.3 Whole sections from resected NSCLCs to assess PD-L1 intra-tumoural heterogeneity

Scorer *et al.* 2018 looked at comparing whole sections of tumour acquired from commercial sources.[393] They looked at one small cohort of 15 patients, (5 NSCLC, 5 NHSCC and 5 UC (urothelial carcinoma)) as well as a larger cohort composed of 200 NSCLC patients. For the first cohort they selected two blocks of tumour from each case, serially sectioned them 51 times and stained sections 2, 25 and 50 for PD-L1 with the SP263 clone (1 and 50 were H&E) and scored for a TPS at <1%, 1-4%, 5-9%, 10% and 5% increments thereafter, and categorised subsequently using ≥1%, ≥10%, ≥25% and ≥50% cut-offs for positivity. The second cohort was from Ratcliffe *et al.* 2017, and had two sections taken from each block, at least 70µm apart, and prepared with at least seven months between them (though sectioned fresh at each time point) to allow for sufficient wash-out periods between scoring.[343] These were also stained for SP263 and scored with the same approach. IC scoring was also captured. For TPS scores of any cut-off, concordance was perfect within blocks in any of the first cohort cases, though it was as low as 70% in some cases for IC scoring. In the second cohort, the OPA (overall percentage agreement) ranged from 91% (≥1% cut-off) to 98.5% (≥25% cut-off), with negative and positive percentage agreements generally favourable (81.4% to 100%), though again this was poorer for IC scoring. Between block analysis in the first cohort again found 100% concordance for OPA, except for ≥50% cut-off in NSCLC (OPA 80%), with again wider ranges for IC scoring (60-100% OPA). This paper demonstrates that tissue quantities in the range of a whole section from a tumour is typically representative of the entire tumour's PD-L1 expression, and is therefore sufficient quantity of tissue to overcome intra-tumoural heterogeneity. Rehman *et al* 2017, described above, used QIF on whole tissue sections and found that variation within a block was generally greater than between blocks (91% vs 9% field of view variance), suggesting that a single block is sufficient to provide an overall assessment of intra-tumoural heterogeneity.[379]

Broadly speaking intra-tumoural heterogeneity is a significant issue, with the perhaps predictable conclusion that the more tissue sampled, the less likely sampling error will result in inaccurate PD-L1 expression analysis. TMAs are more problematic than small biopsies, which are more problematic than whole sections of resected tumour. Clearly where these are available, larger pieces of tumour are preferable, but the reality of NSCLC sampling is that small biopsies are the typical diagnostic specimens utilised for scoring PD-L1, and thus PD-L1 intra-tumoural heterogeneity remains a challenge.

2.3.1 PD-L1 Heterogeneity– Inter-tumoural and temporal heterogeneity

If intra-tumoural heterogeneity can be overcome with sufficient sampling of the primary lesion, it does not help address the fact that inter-tumoural heterogeneity between samples may result in discrepancies between tumour sites. Many diagnostic specimens in NSCLC are from regional lymph node metastases, or distant metastases, and the question of how representative they are of the primary lesion is an important one. Perhaps more importantly, there is the question of what differences between lesions may mean for patient response to treatment, and what biological processes underpin these changes. To begin to understand this, we must first appreciate the scope of inter-tumoural heterogeneity by a review of the studies that have attempted to quantify this issue. As metastases occur over time, and several studies have not sampled primary and metastatic tissue concurrently, the issue of temporal heterogeneity as a confounding factor must also be considered for these papers.

Pinato *et al.* 2016 took an unusual approach to studying heterogeneity by using post mortem FFPE blocks from patients diagnosed with various types of lung cancer but whom had received no treatment for it, totalling 98 patients from between 1970 and 2005.[394] They looked at inter-tumoural heterogeneity by constructing TMAs using multiple samples from both primary and metastatic tumour, with PD-L1 stained for with E1L3N and deemed 'positive' if ≥5% TPS with at least moderate staining by two pathologists. Though these were scored independently the authors merely state the results were 'consistent'. 76% of the patients had metastases suitable for this analysis in addition to the primary tumour. 66% of the cohort were NSCLC, with SCLCs also included. Only eight (12%) of the NSCLC patients were PD-L1 positive in the primary tumour, of which 16 of the 19 (84%) matched metastatic samples were negative. Conversely, of the remaining negative primary NSCLC cases, only one had a PD-L1 positive matched metastatic deposit. All SCLC cases were negative. The low rate of PD-L1 positive expression can probably be put down to both the requirement of 'moderate' staining of ≥5% or greater, and, possibly more importantly, the age of the specimens. As previously seen, old blocks tend to lose PD-L1 expression, and even the most contemporary in this study was some 10 years old at the time of the analysis. Despite multiple sampling of primary sites, unfortunately no comment was given to intra-tumoural heterogeneity (despite the title of the paper), but it does provide evidence of inter-tumoural heterogeneity.

2.3.1.0 PD-L1 Inter-tumoural heterogeneity between primary lung and regional lymph node metastases

Comparison of primary lung tumour and regional lymph node metastases, given the high prevalence of the latter as diagnostic specimens,[299] is an important comparison which is the focus of several

studies. Sheffield *et al.* 2016 constructed TMAs (0.6mm diameter) of matched non-squamous NSCLC primaries and nodal metastases from 78 patients with a core from each site. PD-L1 was stained for by SP142, 28-8, E1L3N and RBT-PDL1 (a rabbit monoclonal research clone) and scored by 3 pathologists independently (with consensus scoring for discrepant cases) by an H-score (0-300). PD-L1 mRNA was also assessed with RNAscope. 28 (35%) of the primary tumours were positive based on at least two of the four antibodies with an H-score of ≥1, with 28 (36%) of the lymph nodes also positive using the same method. Comparing the primary and nodal metastases found 17 (22%) of cases were discordant between them. They also compared mRNA for PD-L1 and found varying levels of concordance for each clone, with the lowest agreement for SP142 (R2 0.53) and highest for 28-8 (R2 0.76), suggesting SP142 does indeed under represent PD-L1 expression, though the authors do not comment on this. Kim *et al.* 2015 studied PD-L1 and PD-L2 as well as PD-1+ve and CD8+ve TILs in 77 resected lung SCC with matched nodal disease. E1L3N was used for PD-L1, and scored into a 4 category bin using both TPS and staining intensity (with 2+ (weak/moderate TPS ≥10% and 3+ (strong staining with TPS ≥10%) taken to be positive) with discordance between primary and matched nodal metastases seen in 29.7% of cases, with both negative and positive 'conversion' seen for both PD-L1 and PD-L2. Uruga *et al.* 2016 looked at 109 resected NSCLCs with matched nodal metastases (split into N1 and N2 as per TNM staging)[395] using representative sections stained with E1L3N and independently scored by three pathologists (with consensus scoring for discrepant cases) and split into TPS categories by ≥1%, ≥5% or ≥50% cut-offs. A total of 66 cases had sufficient tissue in the nodal metastases, with overall correlation of scores to matched primary tumour reasonable for N1 ($r = 0.7173$, $P < 0.00001$) and less so for N2 ($r = 0.4990$, $p = 0.0036$). When using the TPS categories, the differences were significantly discordant, ranging from 15% to 18% of N1, and 9.4% and 38% for N2. The authors also compare N1 to N2 but it is not certain if these are matched or just overall analysis. They do, however, state that differences are maybe due to divergent histology between the primary and matched nodal metastases.

A few studies looked at both regional lymph node metastases and intra-tumoural heterogeneity. A unique example is Liu *et al.* 2018 whom focused purely on adenosquamous (ADSC) carcinomas of the lung. 72 patients with ADSC underwent surgical resection of which 21 (52.5%) had matched regional lymph nodes. E1L3N was used and scoring via TPS and using ≥1% and ≥5% cut-offs, and re-scored including TILs to generate a CPS. 48.6% of primary tumours were positive in any component for TPS. Discordance between adenomatous and squamous components was moderate when using a TPS (≥1% cut-off: 42.2%, K 0.477, ≥5% cut-off 51.9% K0.527) but also seen for a CPS (38.6% K 0.338, 51.6% K 0.493 for ≥1% and ≥5% respectively). When comparing matched regional lymph node samples, they found similar levels of concordance for TPS (57.8% K0.477, 48.1% K 0.527 for ≥1% and

≥5%) respectively and CPS (61.4%, K 0.338, 48.4% K 0.493 for ≥1% and ≥5% respectively). They also looked at morphology of the metastatic sites, and comparing like for like (e.g. squamous component of metastasis to squamous component of primary) found much better concordance for squamous components (90% K 0.792 80% K0.596 for ≥1% and ≥5%), and moderately improved for adenomatous components (77.8%. 74.1%, K not given), and found that 38.9% of patients had different histological patterns between the primary tumour and metastatic site. This highlights a significant role of inter-tumoural variation in regional lymph nodes, but also attempts to explain this may be coupled with morphological changes between sites, as variation is also seen intra-tumourally due to this reason. This is, however, easier to ascertain in ADSCs, given the clear morphological difference in regions, but these remain a small percentage of NSCLC and it is not certain how easily translatable these are to other types of NSCLC. Keller *et al.* 2018 studied 378 lung SCCs by constructing TMAs using eight cores (0.6mm diameter) from a section of primary tumour, and whole tissue section in a cohort of 41 N2 patients to compare the TMAs to the primary tumour and to the lymph node metastases. E1L3N and SP142 were both used and scored by a single pathologist to generate TPS and categorise using <1, 1-49 and ≥50% categories. A final score TPS as a mean of the eight cores was generated for analysis. Although not the primary purpose of the study, they do mention differences between SP142 and E1L3N: generally SP142 scored more cases as negative (62% vs 44%), and fewer as strong positive (13% vs 20%). Comparing primary to N2 nodal metastases TPS found a significant correlation of the clones (p value only given no R or correlation amount) and category agreement in 85% of cases using three categories, and 98% using a ≥50% cut-off. They also noted that correlation of TMAs to whole section TPS was good (R 0.781, p <0.001) with a concordance rate between cores of 29 (of 41 cases) for both E1L3N and SP142 using the three categories. Comparing primary to N2 nodal metastases in 40 cases found a significant correlation (again only a p value only given) with agreement into the three categories in 60% and 70% for E1L3N and SP142 respectively, rising to 90% and 95% for a ≥50% cut-off. This study backs up the notion that sufficient primary sampling will overcome intra-tumoural heterogeneity, and that eight TMA cores may be sufficient, but there is a difference again between both PD-L1 clones and between primary and regional lymph nodes.

A number of other studies looked at various sites of metastases, but for which regional lymph nodes were the predominant site, though they do not distinguish between the sites during analysis. Kim *et al.* 2017 looked at 146 patients with lung ADC with matched metastases, totalling 161 paired samples, (15 patients had multiple metastases included) of which 83.2% were regional lymph nodes and the majority (127) were synchronous metastatic samples.[396] Samples were stained with E1L3N and scored using a ≥1%, ≥5%, ≥10% or ≥50% cut-off for TPS, scored by two pathologists and a

consensus score reach in cases of discrepancy. Primary tumours were positive for PD-L1 in 41 cases (28.1%) to 19 cases (13%) using ≥1% and ≥50% cut-offs respectively. Concordance between matched samples was 75.2% (K 0.433), with the highest concordant rates between negative (<1%) and strong (≥50%) expressers (87.2% and 84.7% respectively), but only 20.8% concordance for other expressers. Dichotomising the groups by ≥1% and ≥50% TPS improved overall concordance to 80.1% (k 0.492) and 90.7% (k 0.598) respectively. They also noted that *EGFR* mutant patients had a slightly worse concordance than *EGFR* wild-type (71.9% vs 74.4% respectively), as well as for smokers vs non-smokers (80.6% vs 71.2% respectively). They also looked at a small number of metachronously metastatic or distantly metastatic disease as separate findings; the former had comparable concordance rates for ≥1% and ≥50% (76.5% and 88.2% respectively) but poorer for distant metastases overall (59.3%). The authors conclude that the different samples may be helpful for analysing a patient's overall PD-L1 status, but the concordance rates would urge caution. They cannot explain the reason for poorer concordance for PD-L1 expressers not at the extremes of expression, though it might perhaps be explained by the fact that negative tumours are less likely to induce PD-L1 at all, and very strong will include constitutive expressers, but weak expressers may represent the most active and dynamic tumours in regards to their immune escape profile. Munari *et al.* 2018 looked at 84 cases of NSCLC of which the majority (67) were from regional lymph nodes, all of which were synchronous, but also included a smaller number of metachronous distant metastases and nine local recurrences (median 40 months range 8-91 months for local recurrence).[346] SP263 was used (using a <1, 1-49, ≥50% TPS groupings). They specifically mention all discordant cases were for regional lymph nodes, but only occurred in nine cases at ≥1% cut-off (12% K0.75) and seven cases at ≥50% cut-off (9.3% k 0.61). Local recurrences was only a small cohort, but discordance was found in three (33%) and one (11%) of cases using a ≥1% and ≥50% cut-off respectively.

2.3.1.1 PD-L1 Inter-tumoural heterogeneity between primary lung and distant metastases

Various other studies focused on a range of distant metastases. Kim *et al.* 2017 studied 37 NSCLC patients who underwent surgical resection between 2005 and 2012, all of which had paired metastatic samples available, including contralateral lung, brain and pleura, in what are presumably largely or entirely metachronous incidences (the authors don't specify but do mention median and range of recurrence time (17.8, 2.5-52.5 months)).[397] Staining was with SP142 and used TPS ranges of <5%, 5-49% and ≥50% alongside intensity to generate an H-score. Two pathologists scored the specimens and agreed a consensus in cases of discrepancy. Overall concordance between the matched samples was 78.4% (K 0.374) , with an equal split of cases expressing more or less PD-L1 in the metastases. The authors also used RISH (propriety RNA ISH)[398] to study PD-L1 RNA levels in

matched specimens, and found a poorer concordance (62.% (K 0.186)), with slightly more cases expressing less PD-L1 in the metastases, which may reflect a more variable level of PD-L1 mRNA than protein, and again suggests these two metrics are not equivalent. Mansfield *et al.* 2016 focused on multifocal lung cancers, and used morphology and NGS to determine if these were independent primaries or related intra-pulmonary metastases.[399] 67 specimens from 32 patients were included (including one large cell neuroendocrine and one sarcomatoid carcinoma), with 23 patient sets found to be independent primaries and nine intrapulmonary metastases. E1L3N was used and scored TPS in 5% increments, with ≥5% considered positive, with IC scored in the same fashion. Only 25% of cases were positive for PD-L1 by TPS. Of the 32 patients, 12 (37.5%) were discordant between specimens, of which 11 were in cases of independent primaries for both TPS (K -0.31) and IC (k 0.02), and only one discordant case in the intrapulmonary metastases for TPS (K 0.73) and IC (K 0.34). The authors also performed agreement analysis and noted that the degree of positivity within one specimen does not predict the positivity of a paired lesion. Evans *et al.* 2018 was a study primarily considering the molecular heterogeneity of NSCLC specimens by looking at 10,005 NSCLC cases scored for PD-L1 as well as for their *EGFR*, *ALK* or *KRAS* status. 22C3 was used to score a TPS and provide categories (<1, 1-49 and ≥50%) by 2-3 pathologists with discordant cases discussed to reach consensus scores. Mutant genes had variable effects on PD-L1 expression, with rare *EGFR* variants and *ALK* translocations associated with higher TPSs. Although not the primary aim of the paper, they do note that pleural metastases and regional lymph node metastases have generally higher rates of PD-L1, both of which remained apparent regardless of sampling technique, including cytology, in the region of around 10-20% increase in TPS versus lung samples. However, these samples were not matched and were simply cohort comparison, so the value of this is lessened, although it does represent a considerable size cohort.

Several studies have focused on brain metastases specifically. Mansfield *et al.* 2016b looked at 146 paired primary NSCLC specimens and matched brain metastases from 76 patients (including one SCLC), the majority of which were whole-tissue sections, and included a range of time intervals from sampled primary and metastases (synchronous in eight cases, median interval 11 months).[400] E1L3N was used, with scoring the same as Mansfield *et al.* 2016a[399], with the exception that only 2 or 3+ intensity of PD-L1 staining was included. There was discordance in ten cases for TPS (14%, k 0.71) and 19 cases for IC (26%, K 0.38). The authors also give some information on temporal heterogeneity, by comparing cases attained within six months to those attained after, with the former faring slightly better (discordant TPS in two (3%) versus eight (11%) cases respectively, IC eight (11%) and eleven (15%) respectively)), though this was not statistically significant. Zhou *et al.* 2018 looked at 25 NSCLC patients with matched brain metastases from cases from 2006 to 2014.[401]

PD-L1 was stained with 6E8 (research antibody), which the authors note has been used before in a clinical trial, though this was only a Phase I novel anti-PD1 monoclonal antibody in oesophageal SCC.[402] Its similarity to other clones is therefore unknown, but they used a TPS ≥5% as positive, ≥50% as strong positive, and IC ≥10% was also recorded. Seven cases (28%) were discordant between matched specimens for TPS, and worse for IC (10 cases, 40%). In their study *EGFR* status had no effect on concordance. Though a small study with a novel clone, it adds further data to suggest a significant inter-tumoural heterogeneity between primary and brain metastases, and this is higher for ICs than for TPS, though as we have seen scoring of IC is consistently more challenging. Teglasi *et al.* 2019 studied 61 matched lung ADC and brain metastases for PD-L1 with the SP142 clone to create TMAs using 3 cores (2 mm diameter) taken from each sample.[403] TPS and IC were both scored (TPS ≥1%, ≥5%, ≥50% IC 0-3) by two pathologists with discordant cases discussed to generate a consensus score. Overall TPS generated a Pearson R of 0.464, which was poorer if using any given TPS cut-off (0.390, 09.409, 0.393 for 1%, ≥5%, ≥50% respectively). IC generally had no significant correlations, with only IC3 stated to be so (R 0.322). This paper is perhaps limited by the use of SP142, and as with many factors explored in this review, uses a slightly different approach to other similar papers, but nonetheless describes significant inter-tumoural heterogeneity.

In contrast to these findings, Wang *et al.* 2019 studied 580 consecutive cases of NSCLC with metastatic samples from various sites from between 2016 and 2018.[404] Sites included brain, bone, non-regional lymph nodes and others. They also included 101 specimens with regional lymph nodes, and included a variety of samples including cytology specimens, though these were mostly fixed with 10% NBF. The 22C3 clone was used and scored by three pathologists, with consensus scoring for discordant cases, and assigned a TPS and split into negative, weak and strong positive categories (<1%, 1-49%, ≥50% TPS) and also, separately, given an intensity score (0 - 3+). 547 cases were adequate, with overall expression of PD-L1 in metastatic tissues distributed as one might expect: 33.7%, 28.7% and 37.6% for <1, 1-49 and ≥50% respectively, and were not significantly different between sites, including regional lymph nodes. 35 patients had multiple metastatic sites, and of these six were discordant between them, though overall there was no significantly different TPS scores either by continuous TPS or by categories. For such a large prospective study, this is a rather disappointing set of findings. Although overall distribution was as expected, (roughly a third in each category) they had no primary tumour for comparison. It does add strength to the notion that a tumour site or sampling methodology does not change PD-L1 expression *per se*, but they cannot comment on the likelihood of any given tumour changing expression as it metastasises to a new site. Their bold conclusion that any site will yield appropriate clinical information is, therefore, not in keeping with the strength of their study. Russell-Goldman *et al.* 2018, mentioned previously for

cytology comparison, also had a small number of cases from different anatomical sites (15) and compared these to those samples acquired from the same site and found a much poorer correlation for PD-L1 expression by SP263 (Pearsons 0.84 vs 0.08).[311]

Cho *et al.* 2017 also looked at local recurrence as well as metastatic samples in NSCLC patients.[405] They selected 91 patients with samples at two time points, the first mostly primary NSCLC (92.3%) with the rest diagnosed via metastatic deposits, and the second time point being a mixture of local recurrence (60.4%) and distant metastases (39.6%). The median time point between collection was 20.2 months (range 0.1-94) with 91.2% of specimens collected >3 months apart. This study also makes note of treatment during the interval period, of which just under half (47.3%) received either XCT, XRT, additional surgery or TKIS. PD-L1 expression was looked for via 22C3 and scored for TPS and categorised into negative, weak and strong positive categories (<1%, 1-49%, ≥50% TPS). Overall concordance between matched samples across the entire cohort when using a ≥1% cut-off was 67%, including negative to positive and positive to negative cases. When using categorical variables to compare samples the concordance is poorer at 57%. Despite this, when looking at TPS as a continuous variable, they found no significant difference between matched specimens based on time of sampling, treatment type or collection method (surgical resection vs. biopsy), though the authors note the study was not powered for these sub-group analyses. Despite their overall concordance of 67%, this highlights that temporal heterogeneity has a role to play, though it is disappointing there was not more comment on local recurrence versus metastatic disease differences. It also highlights the point that TPS as a continuous variable may provide falsely reassuring concordances, but when treated as clinically relevant categorical data the differences may be more pronounced.

Overall the data suggests inter-tumoural heterogeneity, both in synchronous or metachronous samples, represents a significant challenge for PD-L1 expression, in that PD-L1 often varies between sites and it is impossible to know how representative any single sample is of a patients overall tumour PD-L1 expression. Perhaps more importantly the question of what effect this heterogeneity has clinically remains unknown. What is apparent from this review is the difficulty in comparing all the various studies that look at the same issue, but often from many different perspectives. In many articles a single issue or a small number of questions are asked, but based on other data from other papers, the reliability of these results may be bought into question. For example, many studies looking at heterogeneity used tissue older than three years of age, but we know from other studies this will likely result in high false negative rates. Many studies used multiple pathologists to score PD-L1 but make no or minimal comment on inter-pathologist concordance agreement, and it is

interesting to note that in studies where inter-observer variation is not a primary outcome, agreement is generally much more favourable than papers which do address this specifically.

The clones used vary between papers, and the studies, of course, have confounding factors, many of which are unavoidable, but they are all slightly different, using different approaches, with different assay conditions and asking subtly different questions. An attempt to make overall conclusions is therefore very challenging! However the general consensus seems to be heterogeneity, intra-, inter-, or temporal are a significant issue. Adding further data to this are similar studies that have looked at PD-L1 expression heterogeneity in other tumour types, with intra-tumoural, inter-tumoural or spatial heterogeneity seen in breast cancer,[406, 407] malignant mesothelioma,[408] head and neck cancers,[409] colorectal cancers,[410] lymphomas[411] and various tumours that have metastasised to the lung.[412, 413] The inescapable conclusion is that heterogeneity of PD-L1 expression is a significant factor in assessing patients for IM therapy, but the full impact of this on clinical decision making is not, at present, fully understood.

2.3.2 Heterogeneity – iatrogenic heterogeneity

A difficulty in dealing with inter-tumoural variation is the issue of separating local recurrence, independent lung primaries, regional lymph node metastases or distant metastases from each other for analysis, as well as the confounding factors of temporal heterogeneity. As an additional challenge, several studies include patients whom underwent various treatments, (or no treatment at all), in the intervening time period between sampling, potentially impacting PD-L1 expression. The question of specific treatments and what effect they may have on PD-L1 expression in NSCLC will be discussed here.

Some attempts at *in vitro* effects of treatment on PD-L1 expression have been previously made, with the limitations this entails. Shen *et al.* 2017 used NSCLC cell lines exposed to ionising radiation (6 Gy) and stained for PD-L1 (MAB1086 clone) pre and post radiation, and found generally that this resulted in an increase in PD-L1 and proposed mechanisms by which this may occur, and suggested that IM therapy for patients receiving radiotherapy maybe more effective.[414] Several studies have looked at the effects of XCT in various cancer cell lines, with both increases[415] or decreases[416] of PD-L1 expression seen, probably as a result of the specific XCT regime used and/or cell line looked at.[417]

Generally, specific treatments are associated with a trend across a cohort, but in almost all studies there are instances of both increases and decreases of PD-L1 expression following treatment. Zhang *et al.* 2016 looked at a sub-population of their NSCLC cohort (n=30) treated with neoadjuvant XCT and with matched pre-surgical biopsies and surgical resection stained for PD-L1 (Abcam,

unspecified).[418] Across this cohort they see a general trend of fewer positive cases (56.6% to 43.3%). Similarly, Sheng *et al.* 2016 looked at 32 NSCLCs with pre-treatment biopsies, treated with neoadjuvant XCT and matched surgical resection.[419] PD-L1 by E1L3N was positive in 75% of cases, which as a cohort overall decreased to 37.5% by TPS, and also via an H-Score (95 to 75). This was a significant change, and interestingly they noted the inverse trend in IC (43.8% to 56.2%) expression. However, though the cohort's overall trend was a decreased PD-L1 expression, in their figures they still show a minority of cases that do increase post XCT.

Song *et al.* 2016 recruited 76 primary SCC of the lung and used a non-clinically validated clone (Proteintech Unspecified) with H-score of ≥5 deemed to be positive.[420] Patients had pre-surgical biopsies taken, followed by platinum based XCT, and then a surgical resection of the primary lung tumour. The paired specimens were compared to each other, with the initial biopsies positive for PD-L1 in 52.6% of cases, and increasing to 61.8% in the surgical specimens, and they also noted that in addition to the nine cases that went from negative to positive, two cases also went from positive to negative. They also noted *EGFR* had no effect on PD-L1 expression. Omori *et al.* 2017 looked at the effect of various adjuvant therapies on 76 NSCLC patients on PD-L1 expression by E1L3N and found an overall decrease in PD-L1 expression (50% positive to 36% positive) with 24% of cases changing category from positive to negative or strong positive to weak positive, but also found 14% of cases increased PD-L1 across the same categories.[421] Treatment with *EGFR* TKIs had a significant change towards more positive expression, however, and trended towards more positive change in receiving XCT. It should be noted that they used older samples, and this may account for the low initial PD-L1 expression rate, and they also, for reasons not fully explained (though presumably related to their use as internal controls), excluded samples if internal macrophages were negative. Rojko *et al.* 2018 also looked at neoadjuvant XCT in 41 lung cancer patients, including four SCLC cases, using SP142 and scored for both TC and IC.[422] They found overall 22% and 7.3% of cases decreased or increased in TC after XCT respectively, however they found the inverse for IC scoring (24.4% increased and 19.5% decreased). They conclude TC PD-L1 positivity is significantly decreased after XCT, but in fact demonstrate that most patients have no change, and for those that do the change can be both more or less positive. Fujimoto *et al.* 2017 looked at 35 pairs of NSCLC samples, with pre-treatment biopsy, neoadjuvant concurrent XCT/XRT and then surgical resection scored with 28-8.[349] Overall PD-L1 levels decreased in the second sample, (16 cases) but four cases demonstrated an increased TPS. Of the 15 unchanged cases, 11 were negative pre and post treatment.

Table 2.3.0 Papers studying the effect of PD-L1 expression from samples pre- and post-treatment.

A variety of treatments and cohorts were examined, but there is no definite trend for PD-L1 expression to decrease or increase following the administration of any given treatment type.

1ry, Primary; TC, tumour cell; IC, Immune cell; TPS, tumour proportion score; NSCLC, non-small cell lung cancer; FFPE, formalin fixed paraffin embedded; RT, radiotherapy; IM, immunomodulatory therapy

				PD-L1			Patholo		Age of					Pre-treatment	Post-treatment PD-L1 expression
Paper	Cancer	Stage	Specimens (n)	Clone	Scoring	TPS for 'positive'	gists	Difference	specimens	Fixation	Type of specimens	Treatment	Setting	PD-L1 rate	changes
Song et al 2016	SCC Lung	NR	76 x2	Protein tech	H-Score	H score ≥5	Two	Average	2010-2014	FFPE	Pre - 1ry Biopsy; Post - resection	Chemotherapy	Neoadjuvant	53%	9 switched from positive to negative, 2 from negative to positive
Haratake et al 2017	NSCLC	NR	4 x2	22C3	TPS	Not specified	NR	Not recorded	2008-2016	NR	Pre - 1ry Biopsy; Post - resection	Chemotherapy/ RT/IM	Neoadjuvant	25%	2 of 4 went from neg to pos.
Fujimoto etl al 2017	NSCLC	II/III	35 x2	28-8	1%, 5% increments	<1%, 1-49%, ≥50%	Two	Consensus	2004-2013	FFPE	Pre - 1ry Biopsy; Post - resection	Chemotherapy/RT	Neoadjuvant	62%	Overall decrease to 60%
Omori et al 2018	NSCLC	I-IV	76 x2	E1L3N	TPS	<1%, 1-49%, ≥50%	Two	Consensus	2002-2014	FFPE	Pre - 1ry Biopsy/resection; Post - Biopsy/PF/autopsy	Chemotherapy/ RT/TKI	Adjuvant	50%	Overall decrease to 36%
Rojko et al 2018	Lung Cancer	I-III	41 x 2	SP142	TC / IC	Various	NR	Not recorded	NR	FFPE	Pre - 1ry Biopsy; Post - TMAs from resection	Chemotherapy	Neoadjuvant	NR	TC: 22% reduced, 7% increased. IC: 20% reduced, 24% increased
Sheng et al 2016	NSCLC	III	32x2	E1L3N	H-Score, TC IC	Various	Two	Consensus TPS - H score average	2010-2014	FFPE	Pre - 1ry Biopsy; Post - resection	Chemotherapy	Neoadjuvant	TC 75% IC 44%	TC decreased to 37.5% (p=0.003). IC increased to 56% (not sig)
Zhang et al 2016	NSCLC	NR	30 x2	AbCam	Modified H-Score	H Score ≥1	Two	Average	NR	FFPE	Pre - 1ry Biopsy; Post - resection	Chemotherapy	Neoadjuvant	57%	Decreased to 43%
Sakai et al 2019	NSCLC	I-IV	17 x2	22C3	TPS	>1%	Two	Consensus	2014-2016	FFPE	Pre - 1ry Biopsy; Post - 1ry Biopsy + LN + Mets	Chemotherapy	Adjuvant	55%	Increased to 65%
Shin et al 2019	NSCLC	I-IV	86x2	E1L3N	TPS	Various	Two	Consensus	2003-2014	NR	Pre - 1ry Biopsy; Post - resection	Chemotherapy (<10% RT)	Neoadjuvant	52%	Increased to 76%
Guo et al 2019	Lung Cancer	I-III	63x2	22C3	TPS	<1%, 1-49%, ≥50%	Two	Consensus	2011-2018	FFPE	Pre - 1ry Biopsy; Post - resection	Chemotherapy	Neoadjuvant	18%	Increased to 40%

The majority of these papers have shown an overall trend for PD-L1 TPS to decrease following treatment, largely XCT, but a number of papers also show the inverse. Haratake *et al.* 2017 is a case series of four NSCLC patients who had pre-treatment biopsy followed by neo-adjuvant therapy (XCT/XRT/IM) prior to surgical resection.[423] 22C3 was used to score a TPS. Three patients had 0% TPS on initial biopsy, with two of them becoming positive post treatment. They conclude treatment therefore results in positive conversion, but the very small sample size and mixed treatment options mean these results should be interpreted with caution. Sakai *et al.* 2019 reviewed 17 cases treated with adjuvant XCT, with matched pre and post treatment biopsies, (23.5% of the post-treatment biopsies were from regional lymph nodes and distant metastases) stained with 22C3 with ≥1% TPS as 'positive', and found specimens to have an overall trend towards increasing expression of PD-L1 post treatment (43.8% to 64.7% of cases PD-L1 positive), with a similar finding in metastatic cases only were included (33.3% to 69.2% of cases).[424] They specifically mention seven cases becoming 'more positive', though they do not mention if any cases become 'more negative' or not. Shin *et al.* 2019 looked at neoadjuvant XCT in 86 patients with matched pre-treatment biopsies and surgical resections, stained with E1L3N and used various cut-offs (<1%, 1%-5%, 5%-50% and ≥50% TPS)) and found a general increase in PD-L1 expression after treatment (52.3% vs 75.6% of cases), and noted 30% of patients went from negative to positive, but only 7% went from positive to negative, and also noted that a generally increased PD-L1 TPS was typically only seen in patients who did not respond to XCT.[425] Guo *et al.* 2019 looked at 63 lung cancer cases (including nine neuroendocrine) treated with neoadjuvant XCT with matched pre-treatment biopsies and surgical resections stained with 22C3.[426] Using a cut-off of ≥1% TPS only 17.5% of pre-treatment cases were positive, though this raised to 39.7% post-treatment, though there were also 2 patients who had PD-L1 reduced. The same results were echoed when scored for IC (19% and 71.4%). Interestingly, in contrast to Shin *et al.* 2019, they note in their cohort that changes in PD-L1 were more associated patients with poorer response to XCT.

Thus treatment itself seems to be associated with both general increase and decrease in PD-L1 TPS, though often this is across a cohort, and some papers make no mention of the more clinically relevant change across cut-offs. It is also important to note that in virtually every study, there were specific patients for whom both increases or decreases in PD-L1 were seen. The effect of treatment has been largely limited to XCT, but it is fair to assume the future will see further studies considering the effects of IM itself. The study that did include IM consisted of only 4 cases, so it is difficult to extract too much meaningful data. However, a more consistent finding was that PD-L1 expression on immune cells almost always increased across a cohort after treatment. These results are summarised in Table 2.3.0. Anti-cancer treatment, particularly XCT and XRT, induce significant

immune-related changes within the TME, partially as a result of the apoptotic and necrotic impact of these treatments and the subsequent cytokine activity they induce.[427] Indeed pre-IM XRT has been suggested by some groups as a means of evoking a profound immune response prior to immune checkpoint blockade, including in NSCLC. [428, 429] It is outside the scope of this thesis to provide a comprehensive overview of this approach, but suffice to say that whilst this is an exciting area which much promising early data, the widespread and dramatic impact that XRT has on the TME is far from predictable, and thus the use of IMs in combination with XRT remains a challenging area. Importantly, anti-cancer treatment may effect PD-L1 expression, (and indeed many other factors within the TME) but due to the inconsistent nature of these changes, this 'iatrogenic heterogeneity' is difficult to separate from concurrent temporal heterogeneity. Therefore patient/sample matching to minimise the impact of these changes is therefore of utmost importance. Finally, most of the studies looking at treatment effects on PD-L1 expression were retrospective with specimens potentially many years old, and many utilising small biopsies, both of which may also impact PD-L1 expression due to heterogeneity or pre-analytical variation.

2.4 Literature review conclusion

The number of studies involving PD-L1 has increased steadily since the use of PD-1/PD-L1 IM therapy. However, certain areas that have seen intense scrutiny in previous years now receive far less attention as an generally accepted consensus is achieved for certain aspects of PD-L1 IHC. For example, it is generally accepted that the SP142 clone scores fewer tumour cells but more immune cells than other clones, and SP263 and 22C3 are very similar, though not identical, and as such there are fewer papers comparing these clones in 2019 than there were in 2018. Some consensus has been reach for other aspects, such as the fact that approved IVD assays are generally more reliable than LDTs, though the latter can be used with sufficient optimisation, and inter-observer interpretation of PD-L1 varies considerably between studies, and remains a problematic issue. Other areas remain under evidenced; with most data on pre-analytical conditions having little or no PD-L1 specific evidence, for example. The issue of tumoural heterogeneity is a widely accepted phenomena but is perhaps under characterised and no easy solution to reliably overcome the impact of this on assessing PD-L1 expression.

This literature review has been concerned with the use of PD-L1 IHC and what impact pre-analytical, analytical, interpretation and heterogeneity factors will have on its ability to reliably assess PD-L1 expression so as to guide PD-1/PD-L1 IM therapy. Before additional biomarkers or supplementary features can be looked for, it is key that PD-L1 IHC is as optimised and accurate as it can be, and so

the first set of objectives for this thesis must be concerned with contributing to this body of knowledge by addressing gaps relating to PD-L1 IHC, as determined by this literature review.

Chapter 3 - Materials and methods

The specific materials and methods used within each set of experiments is detailed within the relevant chapter. However, some techniques are used multiple times throughout several projects, so for ease of reference and to minimise repetition, these will be shown here. In addition, the main materials and equipment, as well as details pertaining to the different patient/tissue cohorts alongside a brief overview of the statistical approaches utilised throughout this thesis will also be given in this chapter.

3.0 Materials

3.0.0 General Materials and Equipment

- Ventana BenchMark ULTRA (Ventana Medical Systems, Inc)
- Leica ST5020 Multistainer (Leica Biosystems)
- Leica CV5030 Fully Automated Glass Coverslipper (Leica Biosystems)
- Dako CR10030 Fully automated coverslipper (Agilent Technologies)
- Leica RM2135 rotary microtome (Leica Biosystems)
- Leica DM4000 B Microscope (Leica Biosytems)
- Aperio CS2 Scanscope slide scanner
- Roche Ventana DP200 slide scanner
- SuperFrost Plus Adhesion slides (Thermo Scientific)
- X-tra Adhesive slides (Leica Biosytems)

3.0.1 Reagents and antibodies

- Ventana PD-L1 (SP263) Rabbit Monoclonal Primary Antibody (~1.61µg/ml)
- Ventana Bluing Reagent (0.1M Li_2CO_3, 0.5M Na_2CO_3)
- Ventana Hematoxylin II counterstain reagent (≤60%)
- OptiView DAB IHC Detection Kit (OptiView) composed of:
    - Copper (5.0g/l)
    - H_2O_2 (0.04%)
    - DAB (0.2%)
    - HRP Multimer (<40µg/ml)
    - HQ Universal Linker (<50µg/ml)
    - Peroxidase inhibitor (3.05)
- Leica Gill II Hematoxylin (Ethylene Glycol <30% w/w, Acetic Acid <2%w/w, Aluminum sulfate <2% w/w)

- Leica Aqueous Eosin 1%
- Ventana EZ Prep 10x Concentrate (diluted 2l in 20l deionised H_2O) (Cola Terge LFD-C 30-50% w/w)
- Ventana ULTRA LCS (Predilute) (White mineral oil (petroleum) 50-70%w/w)
- Ventana Reaction Buffer Concentrate 10x (diluted 2l in 20l deionised H_2O) (Acetic acid 3-5% w/w, Poly(oxy-1,2-ethanediyl), .alpha.-dodecyl.omega.-hydroxy- 1-3%w/w)
- Ventana ULTRA Cell Conditioning (ULTRA CC1)
- Ventana ULTRA Cell Conditioning (ULTRA CC2) – (1,2-Ethanediol 5-10%w/w, Sulfuric acid monododecyl ester sodium salt (1:1) 1-5%w/w, Citric acid monohydrate 1-5%w/w, Disulfurous acid, sodium salt (1:2) 1-3%w/w)
- Ventana SSC 10x (diluted 2l in 20l deionised H_2O) Sodium Chloride Sodium Citrate buffer

3.0.2 Software

- Aperio ImageScope - Pathology Slide Viewing Software (Leica Biosystems)[430]
- QuPath (https://qupath.github.io – opensource digital slide viewer)[431]
- uPath Roche Ventana slide image software and MDT resource[432]
- IBM SPSS version 25
- Microsoft Office
- Graphpad Prism version 6.01
- Photoshop CC 2019
- R 3.6.1 via RStudio 1.2.1335

3.1 Patient and specimen cohorts

Tumours were classified according to the then current World Health Organisation 2015 criteria or the current 2018 criteria and staged according to the then current seventh edition of the Union for International Cancer Control (UICC) TNM staging system or the current 8th edition depending on the age of the sample.[3, 395, 433]

3.1.0 Liverpool Lung Project (LLP Cohort)

The LLP is an ongoing prospective study that collects lung cancer and case-control tissues from patients receiving treatment for their lung cancer in the Merseyside region, alongside clinical outcome and demographic data and is core funded by the Roy Castle Lung Cancer Foundation. The

tissue is stored as part of a biobank for patients with informed consent. The LLP cohort included in this thesis is a subset of patients from the LLP biobank selected as 139 consecutive patients with the following criteria: NSCLC diagnosis, N1 or N2 metastatic status, surgical resection of the NSCLC undertaken and surgical tissue being available. Demographic and pathological details of the cohort are included in Chapter 6 (Table 6.1.0). In brief, cases came from patients with surgical resections undertaken in 2009-2015, included a range of morphological NSCLC diagnoses, ranged in staging from T1-T4, M0-M1 with all having N1/N2 disease, represented across all lung lobes and a variety of other pathological features including necrosis and pleural invasion, with demographics including smokers and non-smokers and a mixture of patients receiving adjuvant XCT, radiotherapy, chemoradiotherapy or no adjuvant treatment. (No patients received neo-adjuvant therapy or IM therapy). In essence this represents a broad but realistic depiction of the NSCLC population with nodal disease with all tissue taken prior to systemic therapy. FFPE blocks and all glass slides (H&E in all cases, diagnostic IHC and ALK IHC where performed) were retrieved from the RLUH (Royal Liverpool University Hospital) pathology department archives and included all stored surgically resected tissue alongside any pre- or post- surgical samples for each patient. All blocks and slides were manually reviewed to confirm morphological diagnosis (2 cases were re-classified from adenosquamous to squamous cell carcinoma under the approval of Prof. Gosney) and the quantity and quality of the remaining tissue in the FFPE blocks. Cases were excluded if there was insufficient tissue remaining, a non-NSCLC diagnosis made, if significant neuroendocrine components were present, or if blocks and slides were missing to the extent confirmation of the inclusion criteria could not be made. 16 cases were thus excluded to leave 120 cases to be used, for which the term 'LLP cohort' will refer to throughout this thesis. Ethical approval for the use of these specimens was given by the Liverpool Research Ethics Committee (Reference number 97/141).

3.1.1 Clatterbridge Cancer Centre (CCC Cohort)

The CCC is the regional oncology tertiary referral centre and covers a population of 2.3million across Merseyside, Cheshire and the Isle of Man. As part of a collaborative project with CCC, the IMPULSE study (Immune-checkpoint inhibitors in NSCLC: using molecular characterisation of the TME from pre and post IM therapy patient samples to improve efficacy of the anti PD-1/PD-L1 class of drugs) was formed. This study cohort is composed of patients with advanced/metastatic NSCLC whom have been treated with anti-PD-1/PD-L1 IM therapy at CCC between September 2016 and September 2018. The major part of this cohort consists of patients who were not alive at the time of the study and included 309 patients with NSCLC with a range of TNM stages, morphological diagnoses, pathological and demographic features. They also represent a range of treatments, but all patients received either pembrolizumab, atezolizumab or nivolumab, alone or in combination with

chemoradiotherapy or ipilimumab. FFPE blocks and glass sections, including H&E, diagnostic IHC and previously performed PD-L1 IHC with the 22C3 clone as part of clinical workup was retrieved from RLUH and Whiston Hospital pathology departments. All tissue blocks and slides were reviewed to confirm diagnosis and to check remaining tissue quantity and quality. Cases were excluded if FFPE blocks were not available, if there was insufficient tissue remaining, a non-NSCLC diagnosis made or if significant neuroendocrine components were present. A total of 253 cases were ultimately included for which the term 'CCC cohort' will refer to throughout this thesis. The demographic data for these are provided in Chapter 8 (Table 8.2.). Ethical approval 20/EM/0091.

3.1.2 Other cohorts

The LLP cohort was used throughout the thesis in multiple projects, and the CCC cohort was used on a small number of projects in the thesis, both with future work planned. In addition, a number of small cohorts were collected for specific projects, details of which are given here.

The Merseyside-Salford (MS cohort) consisted of 60 pairs of cytology specimens collected via prospective acquisition of EBUS-FNAs during bronchoscopic examination of patients for the explicit use of comparing fixatives on PD-L1 expression. Bronchoscopy is an expert procedure requiring many years of experience to master and outside my skillset, and thus involved collaboration with respiratory physicians, for whom this technical procedure is a routine process, over a time period of March 2018 to February 2019 (Dr Andrew Wight, Dr Hock Tan, Dr Victoria Tippett and Dr Seamus Grundy are respiratory consultant physicians based at Arrow Park Hospital, Aintree University Hospital and Salford Royal). The sampling site for each patient was regional intra-thoracic lymph nodes with metastatic deposits of NSCLC, with patients undergoing the procedure as part of routine clinical follow-up for diagnosis and/or staging of known or suspected lung cancer under clinically validated conditions. Each patient's tumour deposit was assessed during the bronchoscopy, and if adequate material was present, and patient was tolerating the procedure well, a second pass of the lymph node would be performed, to provide a pair of samples for each patient. Further details are provided in Chapter 4.

The Royal Liverpool University Hospital (RLUH cohort) consisted of 122 prospectively collected NSCLC specimens of various types, including cytology EBUS FNAs, small biopsies, pleural fluids and so forth. These came from three pathologists reported as part of clinical predictive profiling at the RLUH on new diagnoses of NSCLC, (myself, Prof Gosney, Dr Piya Parashar) and were all reported for PD-L1 (SP263 clone) from August 2019 to February 2020. The H&E and PD-L1 sections were collected and used to create anonymised digital slide images for explicit use with the PD-L1 interpretative

machine-learning based algorithm. Cases were excluded if insufficient tissue was present for PD-L1 interpretation. Further details are provided in Chapter 5.

The Eli Lilly Cohort consisted of 65 commercially acquired samples of NSCLC, gastric cancer, tonsil and placenta from Asterand Bioscience (Detroit, MI, USA), US Biomax (Rockville, MD, USA), Tristar Technology Group (Washington, DC, USA), and Indiana University Health Methodist Hospital biobank (Indianapolis, IN, USA) in either tissue microarray (TMA) or whole section format. Tissue blocks were obtained between 2012 and 2018. Some 1206 tissue sections were evaluated from 35 gastric carcinomas [tubular adenocarcinoma], 10 NSCLCs [2 primary squamous cell carcinoma (SCC) and 8 primary adenocarcinoma (ADC)], 6 tonsil, and 6 placenta samples. This work was performed during my secondment at Eli Lilly & Co in Indianapolis (IN, USA) and so the FFPE tissue retrieval, storage and preparation was performed by Eli Lilly & Co staff (Michael Soper, Dimple Das). These were used within the context of this thesis for the pre-analytical and analytics studies. Further details are provided in Chapter 4 and Chapter 5.

The vast majority of the work performed within this thesis has been on tissue sections cut from FFPE tissue. Most projects required at least H&E and PD-L1 IHC to be prepared, with others requiring the placement of tissue onto glass slides prior to subsequent analysis by various assays. This preparation is given here.

3.2 Methods

3.2.0 Microtomy

Microtomy was used to prepare tissue from FFPE blocks onto microscope slides for further use including H&E staining, PD-L1 IHC, multiplex IHC, multiplex immunofluorescence (mIF) and MS.

FFPE blocks were placed on an ice tray for 10 minutes tissue side down. Each block is then removed and 4µm sections were cut using the microtome (Leica RM21350). These were placed onto a water bath pre-heated to 45°C and then removed onto appropriate slides (Generally speaking, X-tra adhesive for H&E and MS, Superfrost plus for IHC). Slides are dried in an oven at various temperatures and for various times depending on specific protocols.

3.2.1 Haematoxylin and eosin staining

Tissue is prepared onto slides as per microtomy. Slides are dried for 2 hours in an oven at 50°C. They are then stained for H&E using the Leica ST5020 Multistainer: The slides were washed in Xylene to deparaffinise them (5 minutes) then washed through absolute alcohol (x2, 20 seconds) and washed again in deionised water (20 seconds). The slides were placed in Gill's (II) Hematox (5mins), washed

in deionised water (20 seconds), 0.2% acid water (20 seconds), deionised water (20 seconds), Scott's tap water (150 seconds), deionised water (20 seconds), absolute alcohol (20 seconds) and then placed in Eosin (20 seconds). The slides were then washed in deionised water (45 seconds), absolute alcohol (x2, 30 seconds) and finally in xylene (45 seconds), after which they are coverslipped (Leica CV5030 Fully Automated Glass Coverslipper).

3.2.2 PD-L1 SP263 immunohistochemistry

Tissue is prepared onto slides as per microtomy with an additional step: a FFPE block of anonymised control tissue (placenta for the LLP cohort, multiblock of tonsil, placenta and tumour for other cohorts) acquired from LCL (Liverpool Clinical Laboratories) is sectioned following the same method in microtomy and a section placed on each slide to act as a positive control for the SP263 antibody, in accordance with the manufacturers recommendations. IHC staining for PD-L1 with the SP263 antibody is performed according to the manufacturers guidelines.[251] The slides are placed in an oven for one hour at 60°C. They are then stored in a fridge at 2°C for a maximum of six weeks (typically <48 hours). Slides are then placed in the Ventana BenchMark ULTRA and stained using the manufacturer's protocol for SP263 IHC.[251] The slides are then removed and placed in EZ prep and washed with tap water for 5 minutes. The slides are then coverslipped (Dako CR10030 Fully automated coverslipper).

3.2.3 Digital slide scanning and viewing

All glass slides were reviewed post cover slipping to assess tissue and staining quality adequacy and then scanned to create a digital image of the slide using the Aperio CS2 Scanscope slide scanner and Aperio Scancope console software at 20x maginifcation or the Roche Ventana DP200 at 20x magnification (true magnification 200x). These images were pseudo-anonymised by removing specimen label details and assigning them arbitrary numbers for the purposes of interpretation. Images were viewed on either Aperio ImageScope software, the opensource QuPath software package or the Roche uPath system.[430-432]

3.2.4 Scoring PD-L1 expression

Interpretation of PD-L1 expression is more complex than other IHC stains and requires a careful analysis of the correct cellular component, as well as a quantitative analysis of these features. PD-L1 expression by IHC varies between clones and tissue types, but all tissue scored in this thesis is NSCLC (with a small number of exceptions) and generally speaking the SP263 clone or 22C3 clone have been used (again, with a small number of exceptions). For SP263 and 22C3, a tumour proportion score (TPS) is generated from each specimen and is calculated as:

TPS = (Total number of PD-L1 positive tumour cells/total number of tumour cells)*100

I have received specialist training in both SP263 and 22C3 interpretation by authorised trainers at the start of this thesis, and I now also sit as an expert assessor on a panel for UKNEQAS PD-L1 assessment, as well as having reporting several hundred NSCLC PD-L1 cases clinically per year, and have given numerous talks on the pitfalls and challenges of PD-L1 IHC, in addition to the work performed during this thesis.

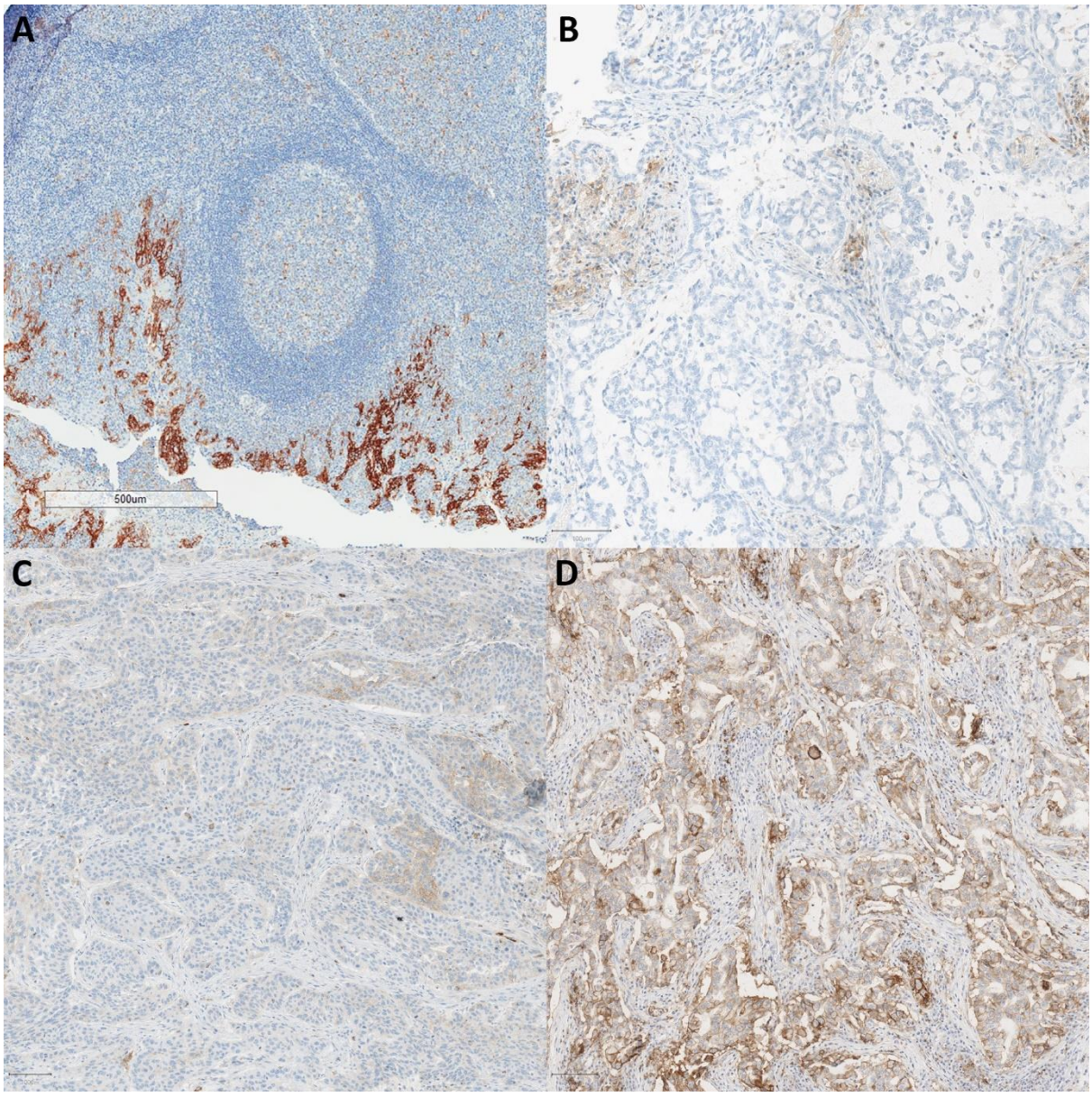


Fig 3.2.0 Assessment of PD-L1 expression **A** - PD-L1 expression in tonsil showing strong epithelial staining and weaker germinal centre staining. This is an ideal control tissue to use on slides with tissue stained for PD-L1 due to these properties. **B-D** PD-L1 expression in NSCLC showing different levels of TPS by conventional clinical groupings of negative, weak positive and strong positive: **B -** <1% TPS **C** – 1-49% TPS **D -** ≥50% TPS.

Detailed guides in the interpretation of PD-L1 are given by each clone's manufacturer[250-253] as well as guidelines from the IASLC,[282] but a brief summary of the salient points of assessing PD-L1 IHC in NSCLC will be given here.

An initial assessment of the H&E section is required to validate the quantity and quality of tissue present. A minimum of 100 viable tumour cells are required. A negative control slide should also be provided to confirm the absence of aberrant PD-L1 staining. On the PD-L1 IHC slide, positive controls are first assessed to ensure adequacy and accuracy of the assay, with the preferred positive control being tissue tonsil, as well as known positive and negative tumours. Tonsil tissue is particularly convenient as the epithelial components have strong membranous staining for PD-L1, with the germinal centres expressing weaker staining, both of which must be present (Fig 3.2.0). Partial or complete membranous staining of a tumour cell means that a cell can be scored as PD-L1 positive, including weak membranous (visible at 20x) or convincing membranous granular staining; cytoplasmic or nuclear staining only do not qualify. The total number of tumour cells in the specimen are calculated, as are the number of positive tumour cells to give a percentage TPS, example shown in Fig 3.2.0. Any other staining is not included, and common challenges include: macrophage staining, lymphocyte staining, strong cytoplasmic staining, crush artefact, diathermy artefact, necrosis, edge artefact, pre-analytical conditions including poor fixation and non-specific staining, all of which may falsely change the TPS, typically by false positives. In my experience the most challenging areas are admixed macrophages that can mimic tumour cells, the close proximity of PD-L1 positive TILs to PD-L1 negative tumour cells and the aberrant expression of PD-L1 by necrotic or apoptotic cells. Examples of these are shown in detail in Chapter 5. With experience and careful review of the H&E, these errors can be minimised.

The same basic approach is used when scoring a TPS for SP263, 22C3, 28-8 or E1L3N clones. For the SP142 clone the scoring is slightly different. The current approach to SP142 in NSCLC is that tumours scoring a TPS of <50% TPS (or TC of 2 or less) are then assessed for immune cell or 'IC' score, in which the area of the tumour covered by PD-L1 positive immune cells is calculated, with IC scores given as 0-3 (IC0 1<1%, IC1 ≥1-5%, IC2 ≥5-10%, IC3 ≥10%).[434] These are NSCLC specific, and other tumour types may require the generation of a CPS (combined positive score) in which both tumour cells and immune cells are scored at the same time, but this approach has not been validated in NSCLC. A final note is that the TPS (or IC) required to define a tumour as 'PD-L1 positive' varies between tumours, indications and clinical trials, and is discussed in more detail later in this thesis, but for NSCLC the typical clinical management of patients for PD-1/PD-L1 IM therapy is by the division of patients into weak positive (1-49% TPS) or strong positive (≥50% TPS), although exceptions do exist.

3.3 Statistics

Throughout this thesis a number of different statistical tests have been used, selected as being most appropriate for each dataset and for each question to be addressed. Every chapter will contain the most important and relevant findings, positive or negative, but to avoid repetition, a brief overview of the assumptions used to determine which statistical tests to use and the interpretation of the statistical tests used throughout this work is given here.

Assumptions for parametric tests include normality of data, homogeneity of variance and independence of observations. Additional assumptions, depending on the test, include linearity, skewness/kurtosis, and equality of group sizes and if data is continuous or categorical and measured on a nominal, ordinal, interval or ratio scale. To determine the validity of these assumptions the data collected throughout this thesis are assessed in the following way. Checking for distribution of data is first performed visually through the use of histograms, density plots and Q-Q plots, with subsequent formal testing by Shapiro-Wilk if deemed necessary ($p<0.05$, null hypothesis rejected and data is not normally distributed). Homogeneity of variance is checked for by Levene's test ($p<0.05$, null hypothesis rejected and data groups have unequal variance). Independence of samples will determine if the use of a repeated measures (or equivalent) test is used, and analysed on a case by case basis, as is the type of variable (e.g. use of continuous data on a ratio scale (TPS from 0 to 100) or categorical data on an interval scale (TPS is negative <1%, weak 1-49% or strong ≥50%)). Data descriptives include skewness (parametric testing suitable if ±2) and kurtosis (parametric testing suitable if ±7) and group sizes (very unequal group sizes can potentially impact parametric tests). Linearity of data is checked for visually by scatterplots and numerically by Pearson's correlation coefficient if assumptions are met (e.g. if variables are measured on intervals, or if data is normally distributed etc.) and Spearman's rank-order correlation if not.

Data are thus described to ascertain the appropriate test, with statistical tests for significance chosen as detailed in Table 3.3.0 As much of the data in this thesis involves the use of TPSs, which generally have a heavy negative skew and non-normal distribution, non-parametric testing was typically recommended. However, medians of TPS datasets tend to be very low relative to the data range (1-2) but means are typically higher (around 50), and parametric testing can still be utilised in larger datasets even if not normally distributed. Thus, for completeness, means are typically included within this thesis, and results by non-parametric testing are confirmed by parametric testing, with any significant differences highlighted within the text. By analysing data in this manner, it is hoped to improve the robustness of the included results.

Statistical analysis was performed using IBM SPSS statistics software, version 25 (IBM Corp, Armonk, NY, USA), GraphPad Prism, version 6.01 (GraphPad Software Inc, CA, USA) or R 3.6.1 via RStudio 1.2.1335 (RStudio Inc, MA, USA). All significances taken as $p<0.05$.

Further statistical testing particular to each project is described in more detail within the relevant chapter.

Data	**Statistical test of significance**
2 groups of data	
Paired, parametric	Paired T-test
Paired, non-parametric	Wilcoxon matched-pairs signed-rank
Unpaired, parametric	Independent samples t-test (± Welch's correction for variance)
Unpaired, non-parametric	Mann-Whitney U/Wilcoxon signed-rank
≥ 3 groups of data	
Paired, parametric	Repeated measures one-way ANOVA
Paired, non-parametric	Freidman Test
Unpaired, parametric	Ordinary one-way ANOVA
Unpaired, non-parametric	Kruskal-Wallis
Post-hoc tests	
One-way ANOVA - equal variances (all data)	Tukey's
One-way ANOVA - equal variances (subset of data)	Dunn's
One-way ANOVA - unequal variances (n<50)	Dunnett's T3
One-way ANOVA - unequal variances (n≥50)	Games-Howell
Kruskal-Wallis	Dunn-Bonferroni
Interrater reliability analysis	
Continuous data	Intraclass correlation coefficient (Two-Way Random-Effects Model)
Categorical data	Cohen's Kappa
Others	
Cross-tabulated proportions (2x2)	Fisher's exact test
Cross-tabulated proportions (> 2x2)	Chi Squared Goodness of Fit

Table 3.3.0 Types of data this thesis has used and the associated statistical test to look for significant differences between groups.

Chapter 4 – The effect of pre-analytical conditions on PD-L1 expression

4.0 Introduction

Pre-analytical conditions include any part of specimen handling up to the point of analysis. Pathology specimens selected for PD-L1 IHC include biopsies, surgical specimens and FNAs, and require a well characterised pathway of processing to be transformed from freshly sampled tissue to a substrate suitable for clinical assays. Each step has the potential to impact tissue quality and is therefore a potential risk factor for changing PD-L1 expression levels such that falsely raised or falsely lowered PD-L1 levels maybe ultimately observed. A brief overview of a typical tissue pathway from sampling to sectioning, that includes pre-analytical factors of potential impact is given here.

4.0.1 Sampling methods

Different sampling methods have their own intrinsic benefits and downsides, and in many instances the various sampling techniques are used together to optimise the treatment pathway; the use of both histology specimens (core biopsies, surgical specimens etc.) and cytology specimens (FNAs, bronchial washings, etc.) in lung cancer for diagnosis, including diagnostic IHC stains is well established.[435-437] Differences between the sampling techniques can affect the quantity and quality of tissue, though sampling via FNA is not a barrier to predictive profiling even by NGS *per se, (*as results are typically concordant with biopsy or surgical specimens), rather it is the issue of tissue insufficiency that raises the biggest concern.[308, 438-440] One particular aspect of PD-L1 expression as a predictive biomarker is the quantitative nature of interpretation, and so even subtle variations in expression may be of clinical significance. The impact of different sampling techniques on PD-L1 expression is difficult to separate from other factors, such as tumour heterogeneity, but as many as a third of NSCLC patients have only EBUS samples for diagnostic and predictive profiling, including for PD-L1,[299] and therefore any potential impact on PD-L1 expression scoring as a result of sampling technique should be considered.

Once tissue has been removed from the body it is either chilled (e.g. with dry ice) for transport and potential analysis (for example, a frozen section for intra-operative diagnosis) before being placed in a fixative, or, alternatively, the tissue specimen is immediately placed in a fixative. Cytology specimens, such as FNAs, will be placed immediately into fixative for further processing, though a direct smear of cells onto slides for assessment is often also performed. Both processing of the cellular fluids to create FFPE 'cell blocks' from which conventional sections can be made, or the use of the direct smears can be used for IHC. In all instances, fixation is a critical step: the type of fixative, the effect of delay in fixation and the total fixation time may adversely affect IHC intensity

and overall PD-L1 expression.[272, 274, 275, 323] Manufacturers' manuals for the clinically validated clones 22C3, 28-8, SP263 and SP142 all explicitly state they are for use in FFPE tissue, and typically make no mention of either fresh-frozen tissue or other fixative methods.[250-253] Most histology specimens are indeed formalin fixed, but the use of other fixatives, particularly in regards to cytology specimens is an important question.

4.0.2 Tissue fixation

A complete review of the biochemistry of formalin fixation falls outside the scope of this thesis, but broadly speaking the objective of any fixative is to prevent the degradation of tissue by decomposition or autolysis, and to preserve the cellular architecture and composition of tissue, including proteins and carbohydrates, such that the tissue is able to undergo processing to form thin sections of tissue (typically 3-5µm) for subsequent analysis. Fixation requires penetration through the tissue by the agent of choice, hence this may require 'opening' of larger specimens: a series of incisions to ensure internal aspects are equally accessible to the fixative. Broadly, the 4 main groups of fixatives are aldehydes, (including formaldehyde which is the active component of formalin), alcohol (including methanol and ethanol), oxidising agents and metallic group fixatives. Each group fixes tissue by distinct mechanisms, and each has benefits and downsides, with the agent of choice usually selected based on anticipated processing needs. Formalin is the most widely used fixative due its cost, ease of use and wide range of applications, and typically used as 10% NBF (~4% formaldehyde in water with 1% methanol).[441] As it penetrates tissue, formaldehyde binds to amino and imino groups of amino acids and DNA to form Schiff bases, which acts as an intermediary between other amino groups and subsequently forms a crosslink between cellular components, including 'protein to protein crosslinks' and 'protein to DNA crosslinks'.[441-444] The formation of these crosslinks occurs within minutes, but typically requires 24-48 hours for a whole specimen to be completely fixed in this manner. Crosslinks formed at this stage are reversible, which is crucial for immunogenic assays including IHC as crosslinking can alter the 3D structure of proteins, and is non-specific, and thus wholly unrelated proteins maybe linked to each other. In both instances, the immunogenicity of a specimen will be adversely effected by the presence of crosslinks, as the epitope binding region of an antigen maybe disrupted (if discontinuous) or 'masked' (if linear). Antigen retrieval (AR) methods which can reverse crosslinking are therefore deployed to overcome this, but it should be noted that prolonged fixation in formalin can result in permanent covalent binding between proteins and DNA, and thus reduce the immunogenicity of tissue in a manner in which AR is ineffective.[441, 443]

Alcohol based fixatives have been considered as alternatives to formalin as they do not cause crosslinking, and would thus overcome the need for AR. Alcohol fixation essentially works by dehydrating tissue which results in coagulation and denaturing of proteins, mostly cytosolic and extracellular proteins, though this can harden the tissue, negatively affecting morphology, including the damage of cell membrane structures, microtubules and nuclear contents, and thus is normally reserved for smaller specimens such as those used for cytology.[445-447] When combined with other chemicals, however, alcohol-based fixatives can result in good preservation of tissue morphology, and hence is a major component of a number of commercial fixatives (UMFIX, CytoLyt etc), which have been demonstrated to be equivalent, if not superior to 10% NBF for IHC, and assays requiring extraction of DNA or RNA, including for NGS.[448-454] Despite this, due to increased costs, lack of robust evidence in all areas, and a system largely dependent on formalin, alcohol based fixatives are the exception in clinical histology laboratories.

Once suitably fixed, the tissue must be further processed, either by inclusion of all available tissue if small biopsies or cytology specimens, or the macroscopic dissection of larger specimens for informative areas, such that the tissue can be embedded in paraffin wax, to produce FFPE tissue blocks, from which sections of tissue can be sliced and floated onto glass slides ready for analysis.

4.0.3 Specimen age and storage

Another important pre-analytical factor in the suitability of specimens for PD-L1 expression analysis is the age of the specimens. This is important for two reasons: the clinical and biological relevance of an older specimen in regards to how representative it is of the patient's current disease status, and the possibility of older tissue falsely representing the status of PD-L1 by under or over staining. The former is dealt with in Chapter 6 (Heterogeneity of PD-L1), and the latter considered more comprehensively in the literature review. In brief, several studies have found that FFPE blocks stored for periods over a year should be avoided as there is a high risk of under expression of PD-L1,[283-285, 287, 291] though use of tissue stored for less time, such as 6 months, seems to be less susceptible to this loss,[236, 286, 289] with IASLC guidelines recommending the use of blocks no older than 3 years of age.[282]

One of the major issues in old tissue is antigen degradation that results in a loss of immunogenicity and under staining of IHC stains. This phenomena is not fully understood, and there is variation of the severity of this between tissues, protein targets and specific antibody clones. The reasons for loss of immunogenicity is complex: a review by Engel *et al*. 2011 looked at 29 pre-analytical factors and found that 15 of these affected immunogenicity of FFPE stored tissue such that IHC staining was negatively impacted; including fixation – (including delay to fixation, type of fixation and duration of fixation), dehydration, paraffin impregnation, slide drying and slide storage.[455] Environmental factors

of storing FFPE tissue is a major area in which attempts to find the underlying cause of immunogenic loss have been made, with a wide range of factors implicated, including pre-storage pre-analytical factors, storage time, ultraviolet A (UVA) exposure, oxidation, humidity, and temperature.[292, 293, 295, 456-459]

For PD-L1 IHC specifically, manufacturers of the PD-L1 clones recommend the immediate staining of sections for PD-L1 once prepared,[250-253] and indeed it has been demonstrated that unstained sections rapidly lose their immunogenicity for PD-L1,[296] with IASLC guidelines recommending storage of tissue in a cool temperature away from light, heat and humidity.[282] The use of stored FFPE sections for IHC is not typical practice in the routine clinical laboratory, with FFPE blocks typically sectioned as required for reflex testing or on an 'as requested' basis. However, the underlying mechanism of loss of antigenicity in FFPE tissue is not fully characterised and the use of sections to study this phenomena is beneficial in that it saves tissue, likely reflects the same process that occurs in stored FFPE blocks and is time effective. In addition, there are scenarios in which FFPE sections are stored and transported over longer periods of times, and therefore the ideal storage conditions of tissue for any given biomarker is an important consideration. For example, in the development of companion diagnostics during international clinical trials, the tissue samples may come from multiple countries, but the biomarker in development maybe only tested at a single site, or some included countries may not allow for the collection of FFPE blocks for clinical trials. Indeed there is considerable variation between countries in regards to ethical, legal and regulatory practices in regards to FFPE tissue.[460] Anticipating the impact of storage conditions on the biomarker would be hugely valuable in their development. To that end, understanding the reasons for antigenicity loss, and having an environment in which loss can be minimised would be useful for both routine storage of clinical tissues and in the development of novel assays.

The work that follows in this chapter therefore sets out to address three major questions pertaining to the pre-analytical factors' impact on PD-L1 staining. Firstly, the impact of sample type on PD-L1 expression by a retrospective review of over 2,000 clinically reported cases. Secondly, the use of differing fixatives, by considering cytology specimens that have been fixed in either alcohol or 10% NBF. Thirdly, the role of environmental factors on stored unstained FFPE sections by using a controlled environment to alter temperature, humidity and oxygen levels. In so doing, the specific impact of these factors on PD-L1 expression can be better characterised, and practical advice of clinical relevance can be generated.

4.1 Methods

4.1.0 Review of retrospective cases

A search of the LCL pathology database was performed using Telepath (Reflection for UNIX 1985-2008) with the following parameters: a two year period between January 1st 2018 and December 31st 2019; SNOMED (Systematised Nomenclature of Medicine Clinical Terms) procedure code (P) "P6000" (PD-L1 analysis) to generate a list of consecutive NSCLCs tested for PD-L1. All returned cases were reviewed for suitability as having a NSCLC diagnosis (including ADC, SCC, NSCLC NOS) and excluded if there was incomplete data, a non-NSCLC diagnosis, or no attempt at PD-L1 IHC was performed. The remit of LCL over this time period included regional and extra-regional testing of PD-L1 for a number of centres, and so included specimens from across Merseyside, Greater Manchester and Cumbria, and all cases were stained for PD-L1 with SP263 clone as per main methods section.

Manual review of listed cases was then performed to extract the following data: diagnosis, specimen type and PD-L1 TPS to generate a raw data spreadsheet of information. This data was extracted and collated by my colleague Dr Felicity Elwin. Additional review of data to ensure completeness was then performed, and the TPS scores, including the return of 'insufficient tissue' (fewer than 100 viable tumour cells) were stratified and grouped by specimen type to account for both sampling methodology (bronchial washings and brushings, aspirates, tissue biopsies, resections) and anatomical site of sampling (primary lung tissue, regional lymph node, distant metastases, pleural/pericardial metastases). PD-L1 TPS scores were categorised as negative (<1%), weak (1-49%) or strong (≥50%) positive for PD-L1.

4.1.1 Use of alcohol or formaldehyde based fixatives

This used the Mersey-Salford cohort as described in the main methods section. As previously detailed, 60 paired specimens of EBUS-FNAs were collected during bronchoscopy, with two passes of regional lymph node metastatic deposits taken to provide each pair. The tissue collected from the first pass of each lymph node was placed into the alcohol-based fixative of choice, depending on the centre (CytoRich Red or CytoLyt), with tissue from the second pass of each lymph node placed into 10% NBF. In so doing this minimises the confounding factors of tumour heterogeneity, both spatially and temporally, ensures no delay to fixation, and minimises inter-procedural variation. After routine diagnostics were performed by regional pathology teams to determine diagnosis/staging and predictive profiling, specimens were then collated anonymously for this study, and labelled as "Pair 1", "Pair 2" etc. All specimens were prepared into FFPE cell blocks with the aid of cytology BMS (Claire Chadwick): all specimens were centrifuged for 10 minutes at 2500 rpm, and the supernatant

was decanted. Specimens fixed in CytoRich Red had an additional intermediate fixation step of 30 minutes to 2 hours in 10% NBF before processing. Resulting pellets from the CytoLyt specimens were placed into an alcoholic preservative solution (PreservCyt; Hologic UK, Ltd). After decantation of the supernatant from all specimens after a second spin, agar was placed onto the pellet to produce a solid cell block. Pellets were then bio-wrapped, placed into a labelled histology cassettes, and processed overnight (12 hours) in a Leica Peloris III HistoCore processor. All specimens, CytoRich Red–fixed or CytoLyt-fixed, had a post-fixation step in 10% NBF for a minimum of 45 minutes while they were in the processor. The specimens were placed through different gradients of alcohol, xylene, and paraffin wax. The FFPE blocks produced at this point were then sectioned and stained for H&E and PD-L1 (SP263) as previously described. Due to the assistance of Claire, the specimens were thus prepared as "Pair 1", "Pair 2" etc. in a blinded fashion in that I was unable to determine which of any given pair were fixed in alcohol or in formalin. Each section was scored by myself to give a TPS, and then re-scored independently by Prof. John Gosney, whom was also blinded to the nature of the fixatives. Discrepancies between scorers was resolved by generating a consensus score with both pathologists reviewing via a multi-headed microscope. Thus each pair was given a TPS score for both specimens (or rated as inadequate if fewer than 100 viable tumour cells were present) and the nature of the fixative revealed post-analysis. TPS scores for each fixative could thus be compared as a cohort.

4.1.2 Environmental factors

This study used the Eli Lilly cohort as described in Chapter 3. A mixture of gastric carcinoma, NSCLC, placenta and tonsil tissues were used. This work was performed during my secondment at Eli Lily & Co, where the acceleration chamber construction, processing of tissue and scanning of slides was performed by Eli Lilly & Co staff. Comparison of storage conditions involved the microtomy of FFPE tissues to create sections of tissue (performed immediately prior to the commencement of each experiment) in either normal storage conditions or within the acceleration chamber with variable environmental conditions. Normal storage conditions referred to a monitored and controlled laboratory environment with a relative humidity range of 14.4–80.5% (average 46.8%) and a temperature range of 20.1–31.0 °C (average 21.6 °C) where tissue sections were not exposed to direct light. The acceleration chamber was contained within an incubator (Panasonic MIR-154-PA, Seacaucus, NJ, USA) without direct light exposure where humidity, oxygen concentration, and temperature could be regulated and measured. Oxygen concentration was maintained and monitored using an oxygen meter [Apogee Instruments Oxygen Meter (MO-200); Logan, UT, USA] and humidity levels were measured using a Lockdown Hygrometer (Lockdown Vault Accessories; Columbia, MO, USA). A series of preliminary experiments were performed to ascertain the

conditions required to suitably accelerate loss of antigenicity of FFPE sections, such that loss of pan-cytokeratin (pan-CK) and PD-L1 by IHC could be observed in a step-wise fashion to significant levels of loss over an approximately 4 week period. The baseline parameters to achieve this desirable loss of accelerated antigenicity were found to be environmental conditions of 37°C, 100% oxygen, and humidity of ~80% (range 75–85%). Each variable could be changed to create different temperatures (20 °C vs 37 °C vs 60 °C), oxygen levels (100% vs 20% oxygen) and humidity (~80% vs ~45%), with experiments aimed to be run for 4 weeks, but at our discretion the experiments could be terminated early (if extremely rapid loss was seen) or extended further (if no discernible loss was seen). The use of desiccant to protect against chamber conditions involved the use of sections placed within the chamber in a closed box, sealed in a protective bag (Minigrip Commercial LLC UV Protection Bag; Alpharetta, GA, USA) with desiccant (Fisherbrand Humidity Sponge Desiccant; Lenexa, KS, USA) and a humidity indicator card (WiseSorbent Technology (Marlton, NJ, USA)).

IHC was performed in a similar manner to that detailed in the methods sections; briefly, serial sections were cut at 4-µm thickness and allowed to dry at room temperature overnight. IHC staining for PD-L1 was performed using four different anti-PD-L1 clones: 22C3 and 28-8 as per manufacturer's guidelines;[252, 253] E1L3N and SP142 as LDTs using EnVisionTM FLEX detection (High pH) on the Autostainer Link 48 (Agilent). Immunostaining for pan-CK was assessed as a control using an antihuman CK, clone AE1/AE3. A representative section from each block was stained for H&E on day 0 of each experiment. All stained tissue sections were scanned on an Aperio ScanScope AT Slide Imager (Leica Biosystems; Buffalo Grove, IL, USA) at ×40 magnification, and images viewed on Aperio ImageScope (v12.3.2).[430]

For each series of experiments the glass slides and digital images were reviewed to score antigenicity of the IHC stains. For NSCLC a TPS score was generated as previously described for the 22C3, 28-8 and E1L3N clones. SP142 was scored for a TPS, but this clone also requires individual scoring of the tumour cell component and the immune cell component to generate a TC/IC score, and gastric carcinomas are clinically assigned a combined positive score (CPS) that counts both positive tumour and relevant immune cells,[461] and thus a CPS score was used in these instances. For the placenta and tonsil tissues, as well as for pan-CK stains in all tissues, a TPS/CPS is not possible to be generated. Instead the digital images of all sections (including the NSCLC and gastric cancer PD-L1 sections) were reviewed in Aperio ImageScope and assessed using the Aperio ImageScope integrated image analysis 'Positive Pixel Count v9'[338] by annotating regions of interest and allowing the 'positive pixel count' to be generated. The number of positive pixels taken as a proportion of the total number of tissue pixels was used to define the positivity score and given as either absolute values (positivity index) or as a percentage change relative to the corresponding sample at day 0 (positivity %). For the

NSCLC and gastric cancer sections scored by PD-L1, the TPS/CPS thus provides a measure of the quantity of relevant cells being stained in a clinically relevant manner, and the positivity score provides an objective measure of the intensity of staining, and quantification of IHC staining intensity change between sections.

Mass-spectrometry (MS) was used in this project as an assessment of PD-L1 protein quantities, discussed more fully in Chapter 7 (The Tumour Microenvironment), and to quantify potential oxidation within the FFPE tissues. MS was performed by Protypia with fuller methods detailed in previously published work.[360, 462] In brief, two PD-L1 peptides found in the extracellular domain (LQDAGVYR and AEVIWTSSDHQVLSGK, referred henceforth to as LQD and AEV respectively) were analysed. These were selected as LQD contains no easily oxidisable residues, and LQD levels are comparable to IHC levels of PD-L1 by E1L3N[360] thus LQD was used as a measure of PD-L1 abundance. Conversely, AEV contains an oxidisable tryptophan (ω) and lies within the recognition sequence for the 22C3 antibody. Thus AEV and ω-oxidation products of AEV could be detected as a measure of oxidation, with potentially relevant findings for the impact on 22C3 expression levels. In addition, global MS analysis was performed that used 274 methionine (M)-containing peptides. The biochemistry of M-oxidation is outside the scope of this thesis, but briefly the oxidation of methionine residues is one of the common post-translational modifications of proteins with ultimately can result in the loss of conformational stability, decrease the biological activity of proteins and play a role in protein ageing.[463] MS has been used extensively to characterize M-oxidation in proteins, and tandem MS (MS/MS) (that is, the selection and fragmentation of already fragmented ions based on m/z by a second MS) can identify the presence and location of this modification.[464] Thus the detection of M-Oxidation products by MS in a global analysis of these 274 peptides is a measure of oxidation within the tissue sample as a whole.

4.1.3 Statistics

Descriptive exploration of data for testing of assumptions prior to formal statistical testing was performed as described in the methods chapter. For the retrospective case reviews data was unpaired and did not fit assumptions for parametric testing, therefore the Mann-Whitney U test was used to compare two groups, and Kruskal-Wallis with Dunn-Bonferroni post-hoc analysis was used for comparison of multiple groups, with Chi-Square testing to compare clinical categories. For the different fixatives in cytology specimens segment, data did not meet assumptions for parametric testing but was paired, so Wilcoxon matched-pairs signed-rank was used to compare means, and Chi-Square testing to compare clinical categories. For the acceleration chamber experiments, comparisons of multiple groups over time were performed using repeated measures ANOVA with

Bonferroni correction and Tukey's post-hoc analysis. Comparisons of two groups at a single point were performed using independent samples or paired Student's t test as appropriate. Relationships between variables were assessed using Pearson's correlation and, if appropriate, linear regression. Analysis of the effect of incubation conditions on methionine oxidized peptides compared the baseline condition and day 28 accelerated degradation condition and used peptide count data for peptides with at least ten spectral counts using Fisher's exact test (two-sided). The Wilcoxon Signed Rank test was used to determine significance in the number of oxidized peptides between the baseline and day 28 conditions. All significances were taken as $p<0.05$.

4.2 Results

4.2.0 Retrospective case review of PD-L1 expression by sampling method

A total of 2,016 consecutive NSCLC specimens scored by SP263 were included for analysis. Overall distribution of PD-L1 TPS into negative, weak or strong were roughly equal (32% vs 30% vs 31% respectively) with an overall 7% inadequate rate. (Table 4.2.0). Separating data by NSCLC morphology found no significant difference of distribution of TPS into clinical groups (p=0.081) or of TPS means (p=0.404) with similar inadequate rates noted (Table 4.2.0)

Division of samples by sampling technique found that biopsies, resections and aspirates generally followed a similar distribution, and although bronchial washings/brushings had fewer strong positive rates there was no significant difference between the numbers in each clinical group (p=0.079) with a non-significant change in mean TPS (p=0.503), Table 4.2.0. Representative images of strong PD-L1 expression in different tissue types is shown in Fig 4.2.0

Division of samples by cytology versus histology also found no significant changes in clinical groups (p=0.075) or mean TPS (p=0.992) and similar inadequate rates (7% vs 8%). (Table 4.2.0)

Division of samples by anatomical site found similar distribution of cases into clinical groups for tissue from primary lung, regional lymph nodes and distant metastases, but a significantly increased mean TPS in pleural/pericardial specimens versus primary tissue (40 vs 30 p=0.008) and regional lymph nodes (40 vs 30 p=0.009) and a significantly increased number of pleural/pericardial cases in strong positive clinical group than non-pleural/pericardial samples. (44% vs 32% p=0.003). (Table 4.2.1)

	N	%	PD-L1 n(%):	<1%	1-49%	≥50%	Inadequate
Total	2016	100		653 (32)	595 (30)	629 (31)	139 (7)
Specimen Type							
Biopsies	1240	62		382 (31)	395 (32)	382 (31)	81 (7)
Aspirates	695	34		237 (34)	181 (26)	218 (32)	59 (8)
Resections	61	3		26 (39)	14 (23)	21 (34)	0
BW/BBs	20	1		8 (40)	6 (30)	3 (15)	3 (15)
Morphology							
ADC	1132	56		374 (33)	315 (28)	358 (31)	89 (8)
SCC	695	34		220 (31)	238 (34)	204 (29)	33 (5)
ADSC	13	<1		3 (23)	3 (23)	5 (38)	2 (15)
NSCLC-NOS	161	8		49 (30)	38 (24)	58 (36)	16 (10)
Other	15	<1		7 (47)	3 (20)	4 (27)	1 (7)
Histo/Cyto							
Histology	1302	65		409 (31)	409 (31)	403 (31)	81 (7)
Cytology	714	35		241 (34)	186 (26)	227 (32)	60 (8)

Table 4.2.0 PD-L1 expression by SP263 in different NSCLC specimens by sampling methodology and by morphology. BW, bronchial washings; BB, bronchial brushings; ADC, adenocarcinoma; SCC, squamous cell carcinoma; ADSC, adenosquamous carcinoma; NSCLC NOS, non-small cell lung carcinoma not otherwise specified

Typically only adenocarcinoma NSCLCs will metastasise to body cavities, thus a further subdivision of specimens concerning only ADCs revealed pleural fluids to have significantly greater mean TPS than primary lung samples (40 vs 27 p=0.005) and distant regional lymph nodes, (40 vs 32 p=0.034) and trended towards greater versus distant metastases (40 vs 26 p=0.084). Although all pericardial fluids had a TPS of 100%, there were only 4 in the cohort and these were not included in a separate analysis. Pleural biopsies trended towards higher average TPS versus non pleural specimens, but no significance was reached. These results are summarised in Fig 4.2.1. Representative images of strong PD-L1 expression in pleural fluid, pleural biopsies and pericardial fluid are shown in Fig 4.2.2.

Specimen	Number	PD-L1 TPS <1% (%)	PD-L1 TPS 1-49% (%)	PD-L1 TPS ≥50% (%)	Average TPS
Primary	603	35	31	34	27
Regional lymph nodes	252	39	26	35	32
Distant Metastases	51	33	33	34	26
Pleural Biopsy	28	29	25	46	38
Pleural fluid	110	21	34	45	40
Pericardial fluid	4	0	0	100	94

Table 4.2.1 PD-L1 expression by SP263 in different NSCLC specimens by anatomical location

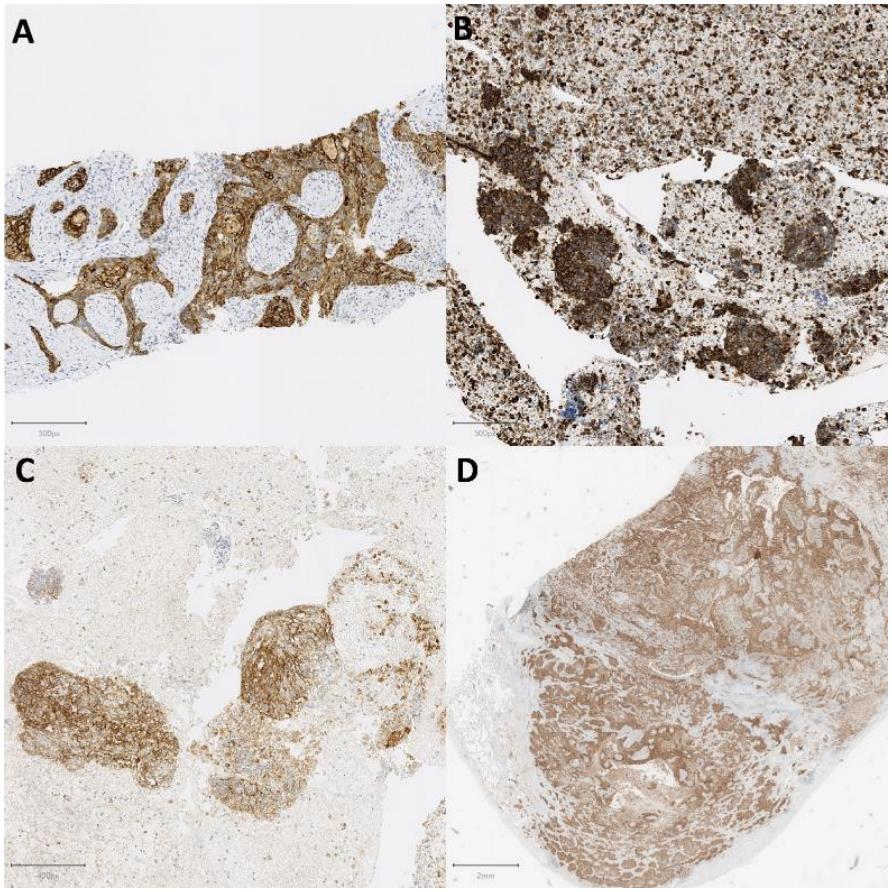


Fig 4.2.0 PD-L1 expression assessed by SP263 IHC in NSCLC for high (≥50%) TPS in four different specimen types: **A** – Core biopsy, **B** – Pleural fluid aspirate, **C** - Bronchial washings, **D** – Resection of primary lung tumour

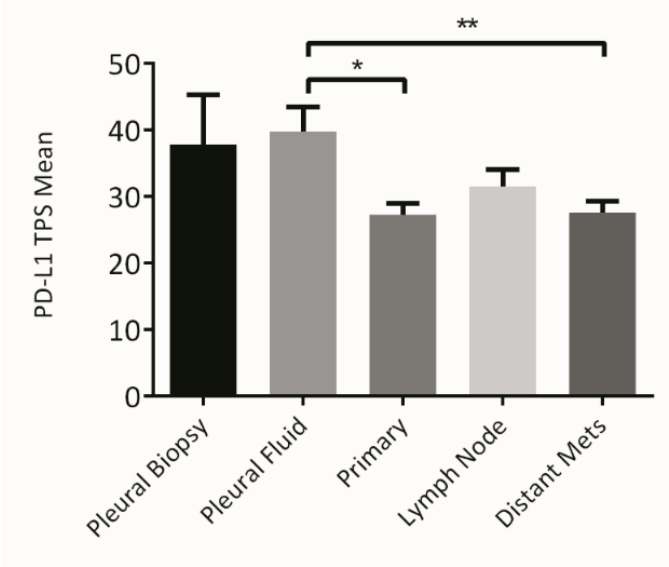


Fig 4.2.1 PD-L1 TPS average for lung adenocarcinomas in different specimens. Pleural fluid is significantly higher than primary tissue and regional lymph nodes. Bars represent mean ± SEM, *=p<0.05, **=0<0.01.

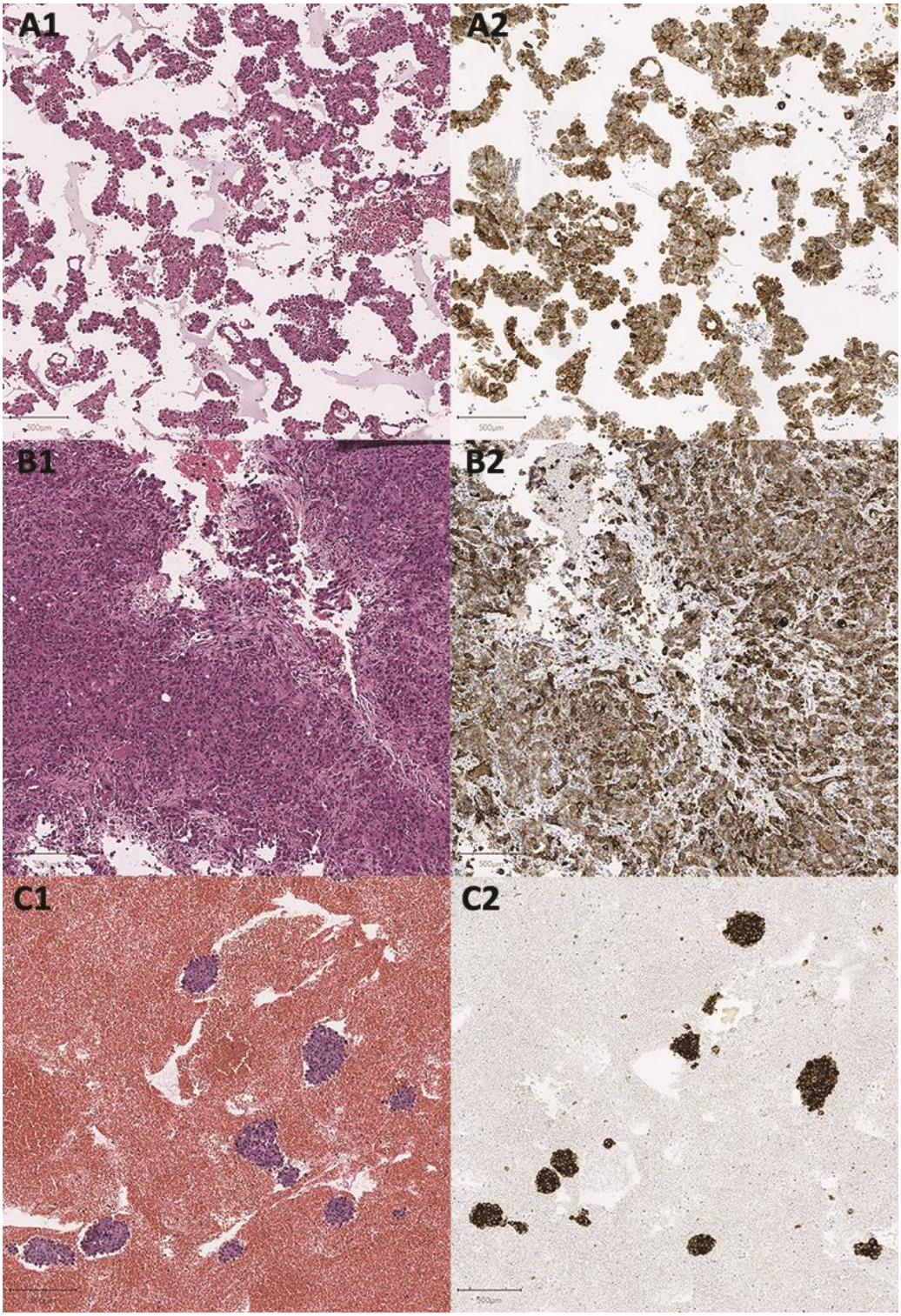


Fig 4.2.2 H&E and PD-L1 expression (SP263) for: **A** – Pleural fluid aspirate; **B** – Pleural Biopsy; **C** – Pericardial Aspirate. Overwhelming positivity of tumour cells and strikingly strong expression is observed in all 3 examples.

4.2.1 Impact of different fixatives on PD-L1 expression in paired cytology specimens

A total of 60 pairs of EBUS FNA specimens were included to provide 120 samples: 60 fixed in 10% NBF, and 60 fixed in alcohol. Of the alcohol fixed cases, 12 (20%) were fixed in Cytolyt, the rest were fixed in CytoRich Red. The fixatives, diagnosis, and TPS scores for both pathologists, including a consensus score is show in Table 4.2.2.

	Diagnosis	**Alcohol**			**Formalin**			**Alcohol fixative**
		Path 1	**Path 2**	**Cons.**	**Path 1**	**Path 2**	**Cons.**	
1	ADC	INS	INS	INS	INS	INS	INS	CytoLyt
2	ADC	0	0	0	0	0	0	CytoLyt
3	SCC	0	0	0	0	0	0	CytoLyt
4	NSCLC NOS	INS	INS	INS	20	20	20	CytoLyt
5	ADC	INS	INS	INS	INS	INS	INS	CytoLyt
6	ADC	10	20	15	10	20	15	CytoLyt
7	ADC	100	100	100	100	100	100	CytoLyt
8	NSCLC NOS	0	0	0	0	0	0	CytoLyt
9	SCC	50	50	50	50	50	50	CytoLyt
10	ADC	<1	<1	<1	INS	INS	INS	CytoRich Red
11	ADC	0	0	0	0	0	0	CytoRich Red
12	ADSC	80	80	80	80	80	80	CytoRich Red
13	ADC	30	30	30	30	30	30	CytoRich Red
14	SCC	<1	<1	<1	<1	<1	<1	CytoLyt
15	SCC	70	70	70	70	70	70	CytoRich Red
16	ADC	60	70	60	60	70	70	CytoLyt
17	SCC	<1	<1	<1	<1	<1	<1	CytoRich Red
18	SCC	0	0	0	0	0	0	CytoRich Red
19	ADC	0	0	0	0	0	0	CytoRich Red
20	ADC	60	60	60	60	60	60	CytoRich Red
21	SCC	90	90	90	90	80	90	CytoRich Red
22	ADC	INS	INS	INS	0	0	0	CytoRich Red
23	SCC	INS	INS	INS	50	50	50	CytoRich Red
24	ADC	80	70	80	80	90	90	CytoRich Red
25	NSCLC NOS	100	100	100	100	100	100	CytoRich Red
26	ADC	95	95	95	95	90	95	CytoRich Red
27	ADC	0	0	0	0	0	0	CytoRich Red
28	ADC	<1	<1	<1	<1	<1	<1	CytoRich Red
29	SCC	90	80	85	INS	INS	INS	CytoRich Red
30	SCC	0	0	0	0	0	0	CytoRich Red
31	ADC	90	80	85	INS	INS	INS	CytoRich Red
32	SCC	90	90	90	90	90	90	CytoRich Red
33	ADC	1	1	1	1	1	1	CytoRich Red
34	SCC	70	70	70	70	70	70	CytoRich Red
35	ADC	1	1	1	1	1	1	CytoRich Red
36	ADC	20	25	20	20	20	20	CytoRich Red
37	ADC	0	0	0	0	0	0	CytoRich Red
38	ADC	90	90	90	90	90	90	CytoRich Red
39	ADC	0	0	0	0	0	0	CytoRich Red

40	SCC	<1	<1	<1	INS	INS	INS	CytoRich Red
41	SCC	0	0	0	0	0	0	CytoRich Red
42	SCC	INS	INS	INS	90	90	90	CytoRich Red
43	ADC	0	0	0	0	0	0	CytoRich Red
44	ADC	0	0	0	0	0	0	CytoRich Red
45	ADC	0	0	0	0	0	0	CytoRich Red
46	ADC	0	0	0	0	0	0	CytoRich Red
47	Basaloid SCC	<1	<1	<1	<1	<1	<1	CytoRich Red
48	ADC	90	90	90	90	90	90	CytoRich Red
49	ADC	100	100	100	100	100	100	CytoRich Red
50	ADC	<1	<1	<1	<1	<1	<1	CytoRich Red
51	ADC	0	0	0	0	0	0	CytoRich Red
52	ADC	<1	<1	<1	<1	<1	<1	CytoRich Red
53	ADC	<1	<1	<1	INS	INS	INS	CytoRich Red
54	SCC	2	2	2	2	2	2	CytoRich Red
55	ADC	30	30	30	30	30	30	CytoRich Red
56	ADC	95	95	95	95	95	95	CytoRich Red
57	ADC	100	100	100	100	100	100	CytoRich Red
58	ADC	95	95	95	95	95	95	CytoRich Red
59	ADC	99	100	99	99	95	99	CytoRich Red
60	ADC	5	5	5	5	5	5	CytoRich Red

Table 4.2.2 PD-L1 expression by SP263 in 60 paired specimens of EBUS FNAs scored by two pathologists, with concordance scores and fixative details.
Cons., consensus score; INS, insufficient; ADC, adenocarcinoma; SCC, squamous cell carcinoma; NSCLC, non-small cell lung carcinoma not otherwise specified.

Of the 120 samples, a total of 13 (10.8%) were insufficient, and there was no significant difference between the fixatives for insufficient rate; 2 cases being insufficient for both fixatives, 4 insufficient for alcohol only, and 5 insufficient for 10% NBF only (p=0.974). There was no significant difference comparing TPS means for 10% NBF vs alcohol (35 vs 35, p=0.194), nor any difference in clinical groups, as all 60 pairs remained in the same clinical category (p=0.974). Separating the groups into alcohol fixatives also found no significant difference in mean TPS comparing CytoLyt to 10% NBF (28 vs 29 p=0.351), CytoRich Red to 10% NBF (36 vs 37 p=0.323), or CytoLyt to CytoRich Red (28 vs 37, p=0.112). Representative images of PD-L1 expression in paired cytology specimens are shown in Fig 4.2.3.

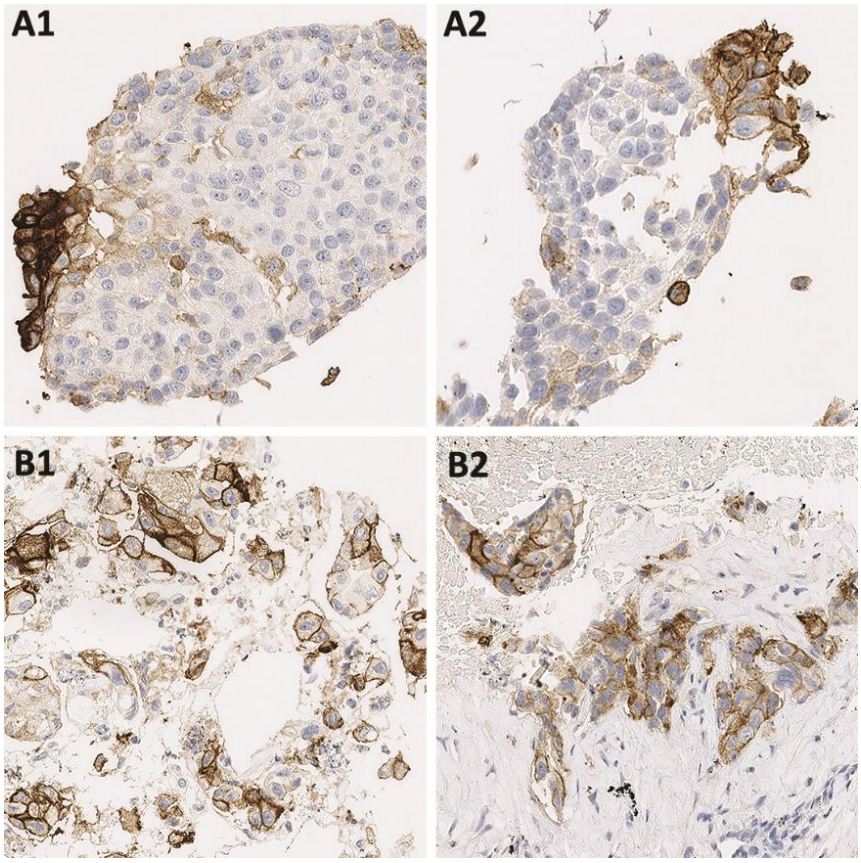


Fig 4.2.3 Expression of PD-L1 (SP263) in two pairs of EBUS FNAs, with one (**A**) containing squamous carcinoma (TPS, 2%) and the other (**B**) containing adenocarcinoma (TPS, 90%). One aspirate of each pair was fixed in an alcohol-based fixative, and (**A2, B2**) the other was fixed in formalin (**A1, B1**). There was no discernible qualitative or quantitative difference according to fixation in any of the 50 pairs studied.

4.2.2 PD-L1 loss in normal conditions

PD-L1 loss in stored tissue sections has been previously demonstrated, and to confirm we could replicate this and attempt to quantify the amount of PD-L1 expression loss in FFPE tissue sections, TMA sections containing 35 gastric carcinomas were stored for up to 24 months in normal storage conditions. Sections were stained for E1L3N and SP142 at day 0, 4.5 months and 24 months. Positive pixel count scoring (positivity) demonstrated that both clones showed significant loss of PD-L1 expression scores over time as expected: average positivity score for E1L3N was 0.197, 0.107 and 0.074 at 0, 4.5 and 24 months respectively, (0 vs 4.5 $p=0.05$, 0 vs 24 $p<0.001$) and for SP142 was 0.128, 0.075 and 0.074 at 0, 4.5 and 24 months respectively (0 vs 4.5 $p<0.001$, 0 vs 24 $p<0.001$). However, only E1L3N has any appreciable loss between 4.5 and 24 months, but this was not significant (0.107 vs. 0.074 $p=0.285$) with SP142 having almost no change (0.075 vs 0.074 $p=0.890$).

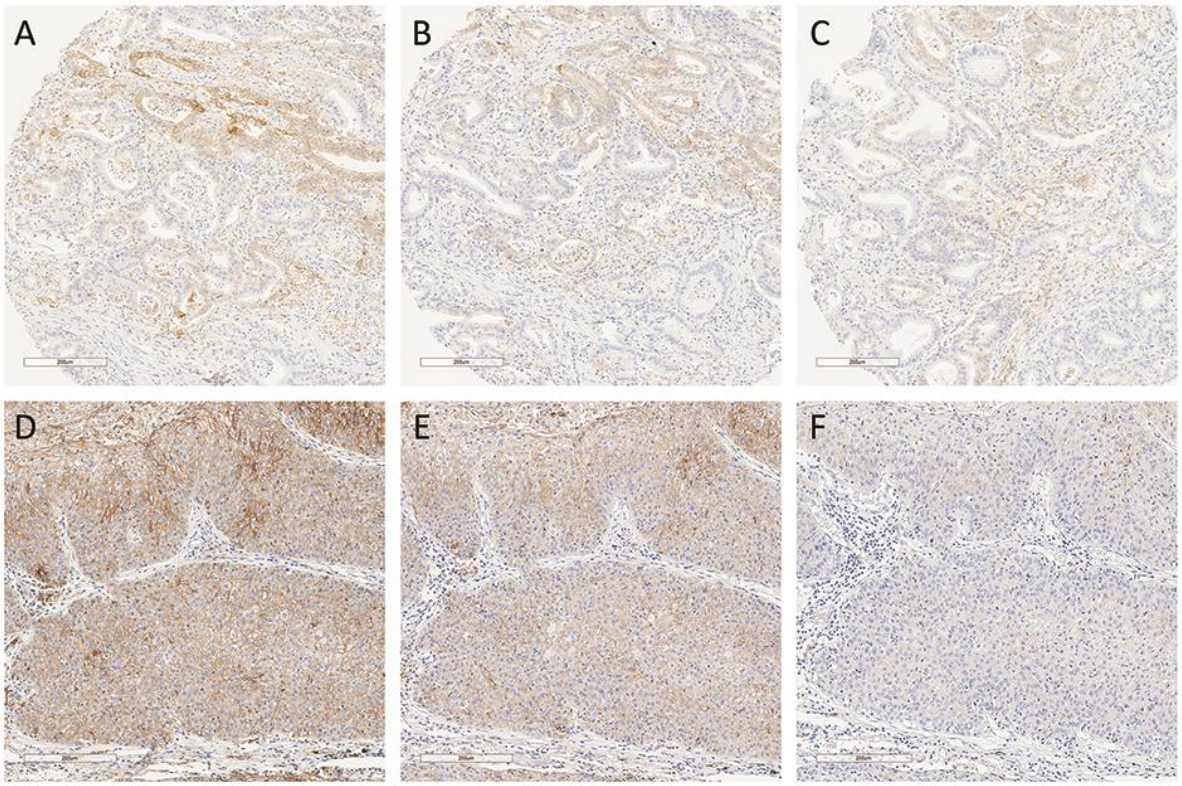


Fig 4.2.4 PD-L1 expression in aged tissue and tissue under accelerated conditions. Representative PD-L1 expression assessed by E1L3N IHC in gastric carcinoma under normal atmospheric conditions (**A-C**) and in NSCLC under acceleration conditions (**D-F**). **A** Day 0, **B** 4.5 months, **C** 24 months; **D** Day 0, **E** Day 9, **F** Day 28.

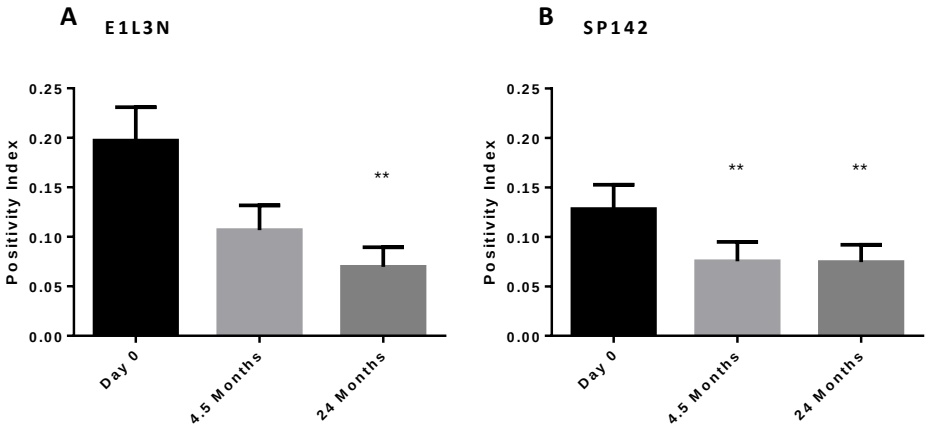


Fig 4.2.5 PD-L1 expression in gastric cancer TMAs over 24 months for E1L3N (**A**) and SP142 (**B**). E1L3N has a stronger average signal than SP142 with a significant loss in PD-L1 expression only seen by 24 months. SP142 has a significant reduction in PD-L1 expression by 4.5 months, but with no further discernible loss at 24 months. (Bars represent mean ± SEM, **=p<0.01)

CPS assessments for the gastric cancer TMAs were higher on average for E1L3N than for SP142 (CPS 40 vs 30; $p < 0.05$). Clinically relevant loss of CPS (from ≥1% to <1%) was seen by 4.5 months for both clones (E1L3N 13% of cases, SP142 20% of cases) with further loss by 24 months (E1L3N 33% of cases, SP142 37% of cases) (Fig 4.2.6).

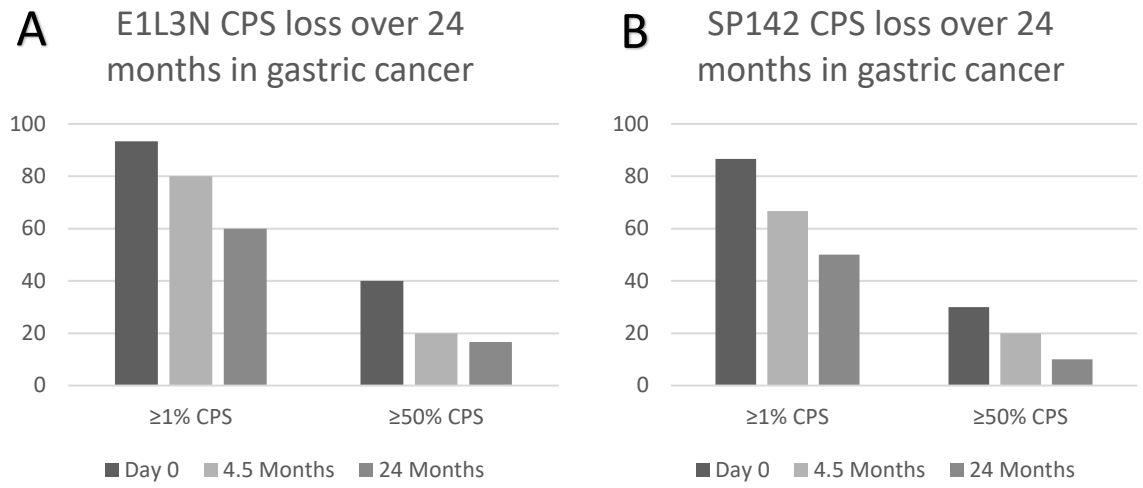


Fig 4.2.6 PD-L1 expression in gastric cancer TMAs over 24 months for E1L3N (**A**) and SP142 (**B**) by CPS for all positive cases (≥1%) and strongly positive cases (≥50%) with a reduction of both categories in both clones over time.

4.2.3 PD-L1 loss in accelerated conditions

To determine whether the natural loss of immunoreactivity could be reproduced in an accelerated fashion, tissues were subjected to controlled environmental stress. Storing unstained sections of the NSCLCs in the acceleration chamber at 100% oxygen, 37 °C, and 80% humidity resulted in repeatable stepwise loss of PD-L1 expression over 28 days comparable with loss, in effect, seen over 24 months in ambient conditions (Fig. 4.2.4 D-F).The NSCLC tissues were stained and assessed for PD-L1 expression using multiple antibody clones. Image analysis of these demonstrated day 0 PD-L1 expression was broadly equivalent for 22C3 and 28-8, with a slightly increased average expression for E1L3N, and markedly lower average expression for SP142, in keeping with what might be expected. For the 22C3, 28-8, and E1L3N clones a stepwise loss of PD-L1 expression is seen over 28 days within the acceleration chamber. The low immunostaining by SP142 at day 0 resulted in minimal further detection of expression decrease across NSCLC specimens thereafter. These results are shown in Fig 4.2.7.

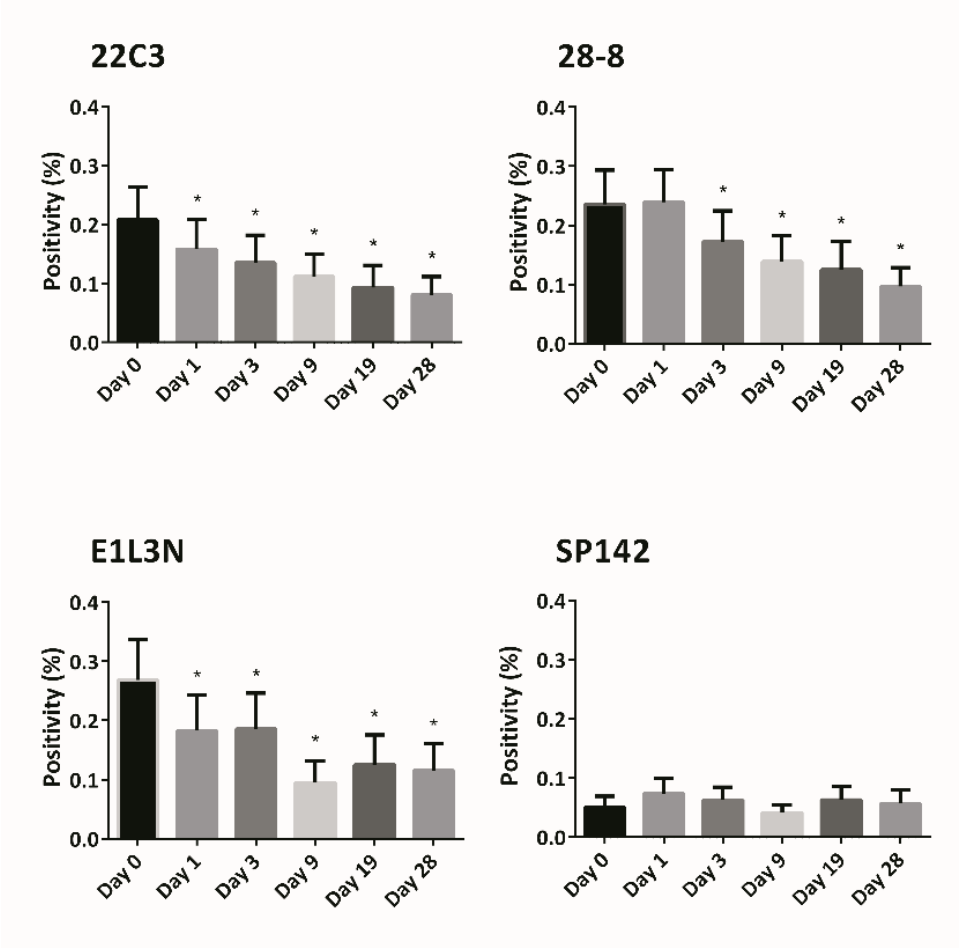


Fig 4.2.7 PD-L1 expression by positive pixel count in NSCLC sections over time in the acceleration chamber with conditions of 100% oxygen, 80% humidity, and 37°C for 22C3, 28-8, E1L3N, and SP142 PD-L1 clones. Significant reduction in PD-L1 expression is measured against day 0 values. Bars represent mean ± SEM. *p<0.05.

PD-L1 expression by TPS found similar numbers of positive (TPS ≥1%) and strongly positive (TPS ≥50%) cases when assessed with the 22C3 (seven and four cases, respectively), 28-8 (nine and four cases, respectively), and E1L3N (9 and 5 cases, respectively) clones, but fewer positive and strongly positive cases for SP142 (three and one, respectively). Loss of PD-L1 TPS in NSCLC sections to levels below these prescribing guideline cut-offs occurred for all clones at varying time points, with over half of cases changing from diagnostically positive to diagnostically negative by day 19 for 22C3 (TPS ≥1% and ≥50%), day 9 for 28-8 (TPS ≥1% and ≥50%), and day 9 for positive and day 19 for strongly positive for E1L3N (Fig. 4.2.8)

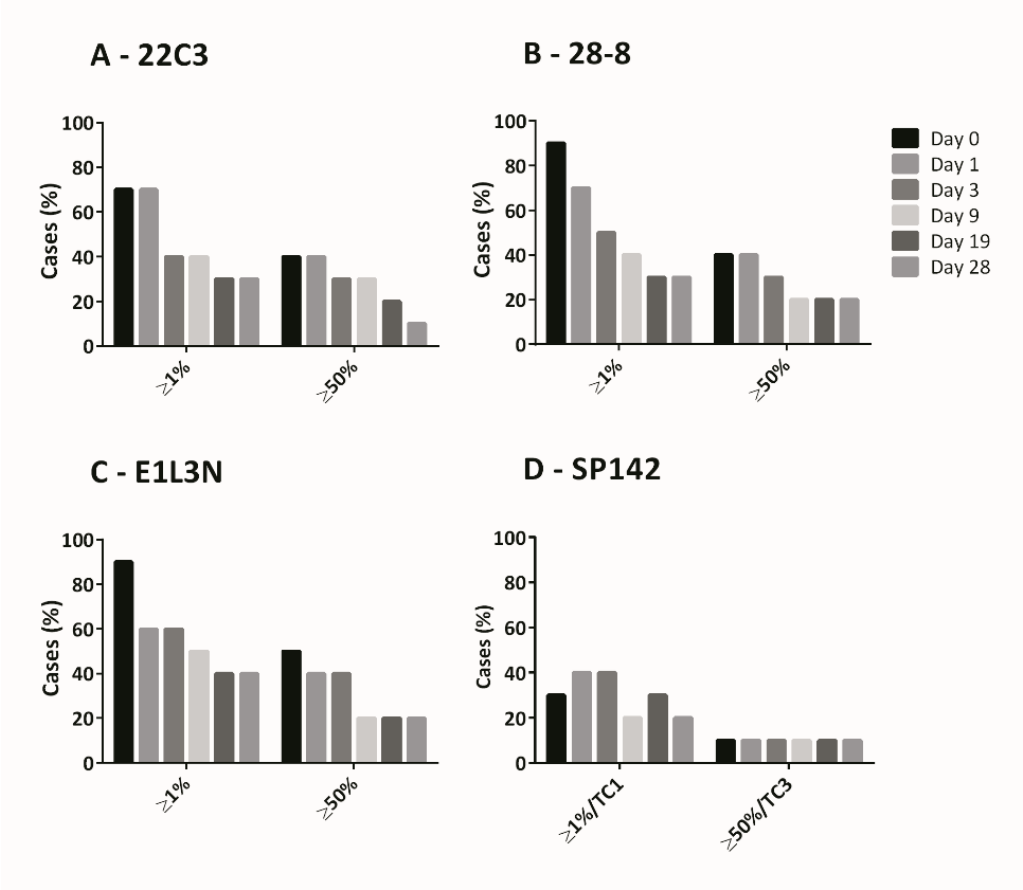


Fig 4.2.8 PD-L1 expression by clinical cut-offs in NSCLC sections over time in the acceleration chamber with conditions of 100% oxygen, 80% humidity, and 37 °C for 22C3, 28-8, E1L3N, and SP142 PD-L1 clones. Bars represent number of cases in series with PD-L1 expression equal or above TPS clinical cut-off thresholds for all positive cases (≥1%) and strongly positive cases (≥50%). SP142 specific TC scoring is included.

4.2.4 Effect of humidity and temperature on immunoreactivity loss by IHC for PD-L1 and pan-CK

In order to further understand the relative contribution of major environmental conditions on tissue immunoreactivity, the effect of oxygen, humidity, and temperature on PD-L1 (E1L3N) and pan-CK (AE1/AE3) IHC expression was assessed on tonsil and placenta tissue sections stored in the acceleration chamber.

First, the baseline acceleration chamber conditions were used to study the effect of these IHC assays in these tissues. By day 28 significant loss of PD-L1 and pan-CK positivity was seen in both placenta and tonsil [average positivity day 0 vs day 28: PDL1 in placenta 0.528 vs 0.088 (100% vs 17%); $p < 0.001$, tonsil 0.123 vs 0.018 (100% vs 15%) $p < 0.001$; CK in placenta, 0.626 vs 0.259 (100% vs 41%); $p = 0.05$, tonsil 0.319 vs 0.219 (100% vs 69%); $p < 0.05$]. Control slides kept at normal ambient

conditions had no significant loss of either PD-L1 or pan-CK expression by day 28 for both placenta and tonsil tissue.

Changing the temperature in the acceleration chamber had a significant impact on the rate of IHC signal loss. Increasing the temperature to 60°C (in the context of elevated oxygen and humidity) resulted in extremely rapid loss of PD-L1 expression: by day 7 PD-L1 expression was reduced to 8% and 3% positivity in placenta and tonsil respectively. These conditions degraded PD-L1 so quickly that the experiment was terminated before day 28. Conversely, decreasing the temperature to 20°C reduced immunoreactivity loss, resulting in no statistically significant loss of PD-L1 in placenta tissue by day 28 [average positivity, 0.538 vs 0.257 (100% vs 48%); $p = 0.174$] or pan-CK in both placenta and tonsil tissue [average positivity, placenta 0.615 vs 0.215 (100% vs 35%); $p = 0.284$, tonsil 0.293 vs 0.247 (100% vs 84%); $p = 0.423$] with a significant reduction seen only for PD-L1 expression in tonsil by day 28 [average positivity 0.135 vs 0.07 (100% vs 52%), $p < 0.05$] though this was significantly less than the loss seen under 37 °C conditions [average positivity of PD-L1 in tonsil by day 28, 20 °C vs 37 °C: 0.07 vs 0.02 (52% vs 14%) $p < 0.05$], results summarized in Fig 4.2.9.

Reducing humidity had a significant impact on the rate of IHC signal loss. Decreasing humidity to 45% (in the context of elevated oxygen and temperature) resulted in no significant loss of PD-L1 expression in tissues, such that the experiment was significantly extended. At 28 weeks a significant loss in tonsil was seen to occur [0.183 vs 0.021 (100% vs 10%); $p < 0.05$], but even at 39 weeks no significant loss in placenta was seen [0.639 vs 0.408 (100% vs 64%); $p = 0.201$], and the experiment was terminated at this point. The rate of PD-L1 expression loss in reduced humidity conditions was slowed to the extent that both placenta and tonsil tissue demonstrated loss by 39 weeks at 45% humidity similar to, or less than, 1 week at 80% humidity [average positivity for placenta 0.41 vs 0.17 (64% vs 27%); $p = 0.13$, and tonsil 0.018 vs 0.019 (10% vs 10%); $p = 0.93$]. Average positivity of PD-L1 in placenta and tonsil at 28 days under 45% and 80% humidity are shown in Fig. 4.2.9.

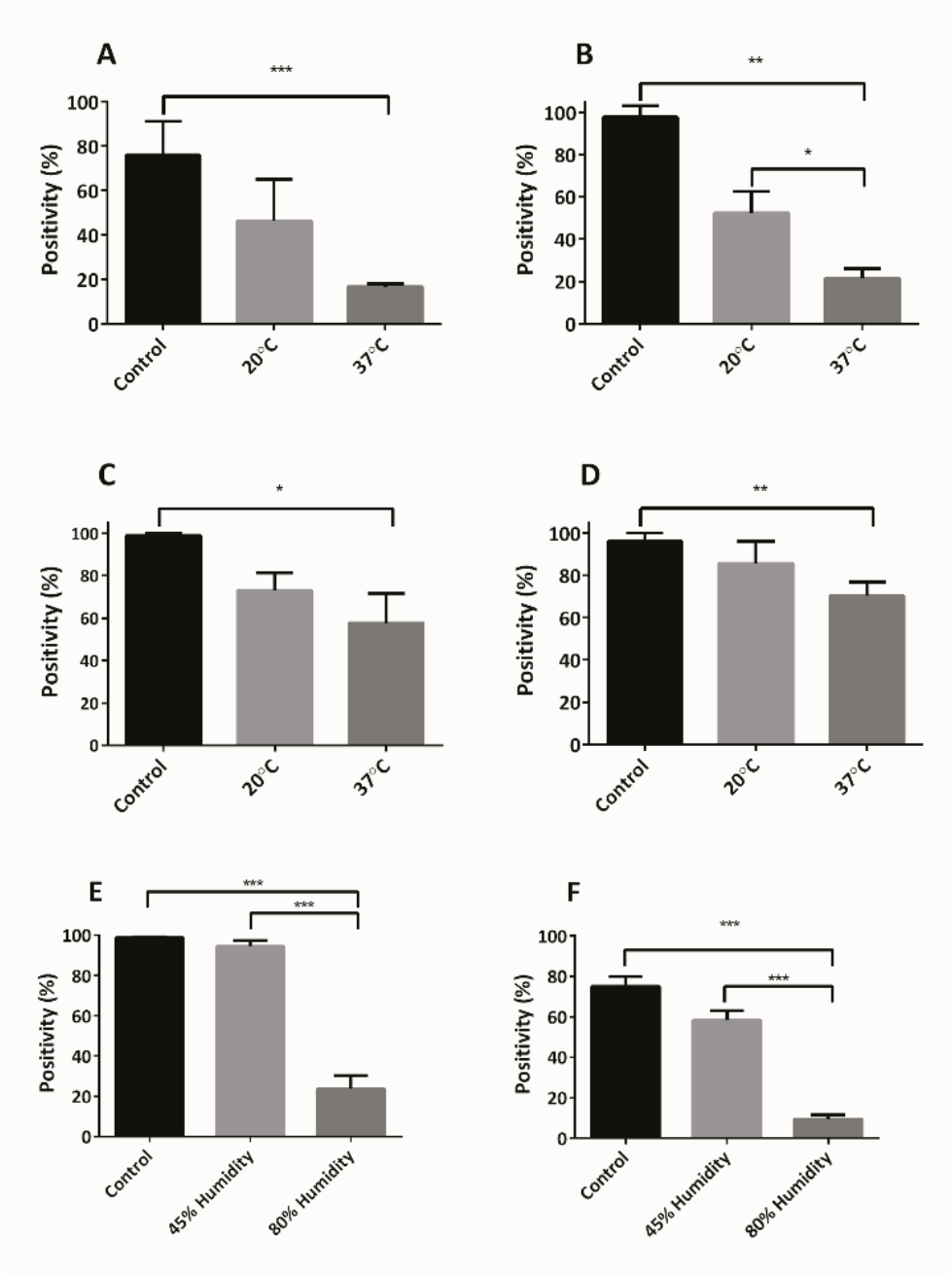


Fig 4.2.9 PD-L1 expression in placenta and tonsil FFPE sections incubated in the acceleration chamber under different environmental conditions at day 28; a–d: 100% oxygen and 80% humidity at either 20 °C or 37 °C, then stained for PD-L1 (E1L3N) or pan-CK (AE1/AE3): **A** Placenta PD-L1, **B** Tonsil PD-L1, **C** Placenta pan-CK, **D** Tonsil pan-CK. 100% oxygen and 37 °C at either 45% or 80% humidity at day 28, **E** Placenta PD-L1, **F** Tonsil PD-L1. Control conditions: 20 °C, atmospheric humidity and oxygen. Bar represents mean ± SEM. *p < 0.05, **p < 0.01, ***p < 0.001.

4.2.5 Effect of oxygen on PD-L1 expression

Reducing oxygen levels in the incubation chamber to 20% (in the context of elevated temperature and humidity) had no significant effect on the rate of loss or total quantity of either PD-L1 or pan-CK

expression measured in either placenta or tonsil tissue [day 28 positivity, 100% vs 20% oxygen: PD-L1, placenta 0.074 vs 0.08 (14% vs 16%); p = 0.918; tonsil 0.0079 vs 0.016 (7% vs 14%); p = 0.937; pan-CK, placenta 0.43 vs 0.45 (68% vs 71%) p = 0.918; tonsil 0.21 vs 0.22 (61% vs 64%); p = 0.937]. Results shown in Fig 4.2.10.

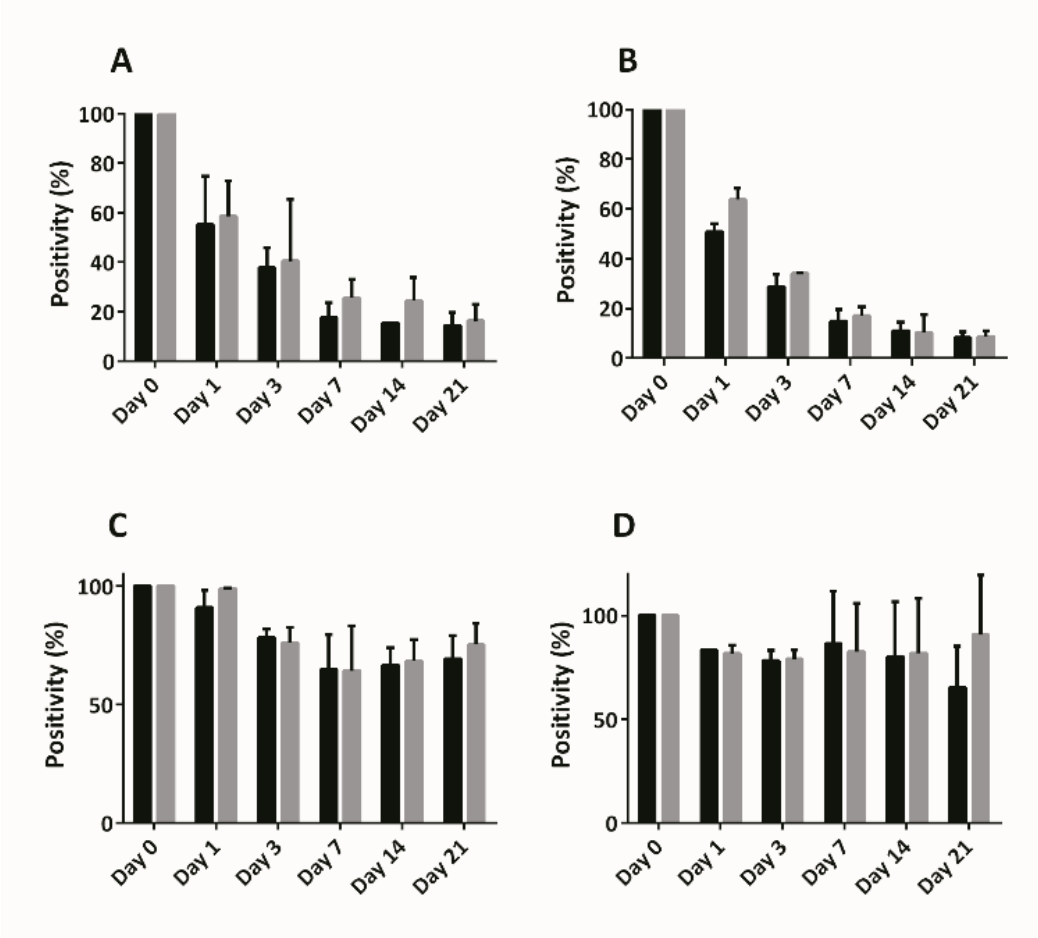


Fig 4.2.10 PD-L1 expression in placenta and tonsil FFPE sections incubated in the acceleration chamber under different environmental conditions at day 28; a–d: 100% oxygen and 80% humidity at either 20 °C or 37 °C, then stained for PD-L1 (E1L3N) or pan-CK (AE1/AE3): **A** Placenta PD-L1, **B** Tonsil PD-L1, **C** Placenta pan-CK, **D** Tonsil pan-CK. 100% oxygen and 37 °C at either 45% or 80% humidity at day 28, **E** Placenta PD-L1, **F** Tonsil PD-L1. Control conditions: 20 °C, atmospheric humidity and oxygen. Bar represents mean ± SEM. *p < 0.05, **p < 0.01, ***p < 0.001.

Furthermore, there was minimal evidence of oxidation by MS. Detection of LQD was comparable to PD-L1 TPS at day 0 ($r^2 = 0.7436$, $p<0.0013$, Fig 4.2.11 A), but whereas decrease in PD-L1 levels was seen overtime by IHC in the standard acceleration conditions, at day 9 and day 28 no loss of either LQD or AEV was seen, indicating no genuine loss of PD-L1 protein, but also no oxidation of PD-L1 proteins. (Fig 4.2.11 B) Increased variation in measured PD-L1 abundance was seen at day 9 and day 28 for both peptides, relative to day 0, which may be an artefact of the storage conditions, but

comparison of the MS peak areas did not show any apparent loss, again indicating no genuine loss of PD-L1 protein (Fig 4.2.11 B).

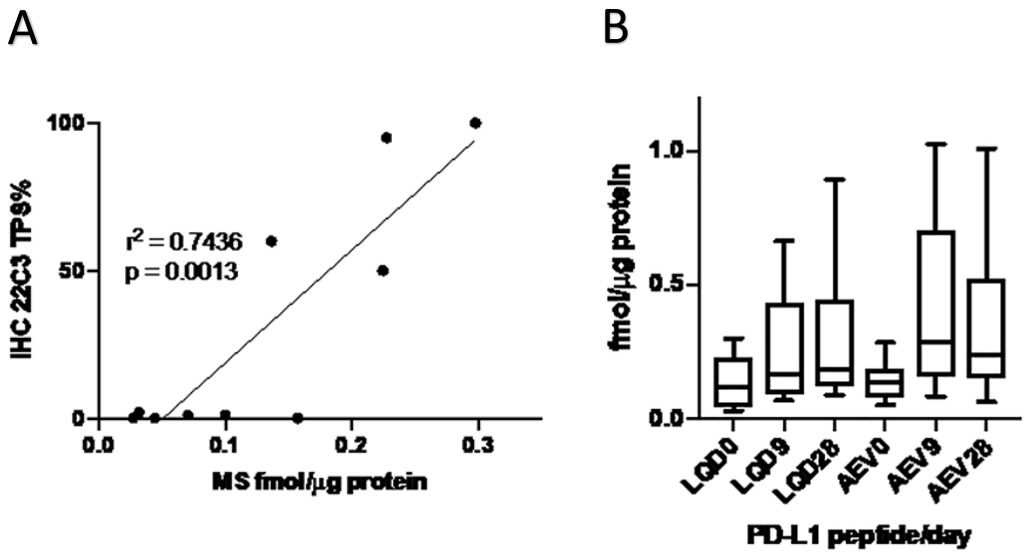


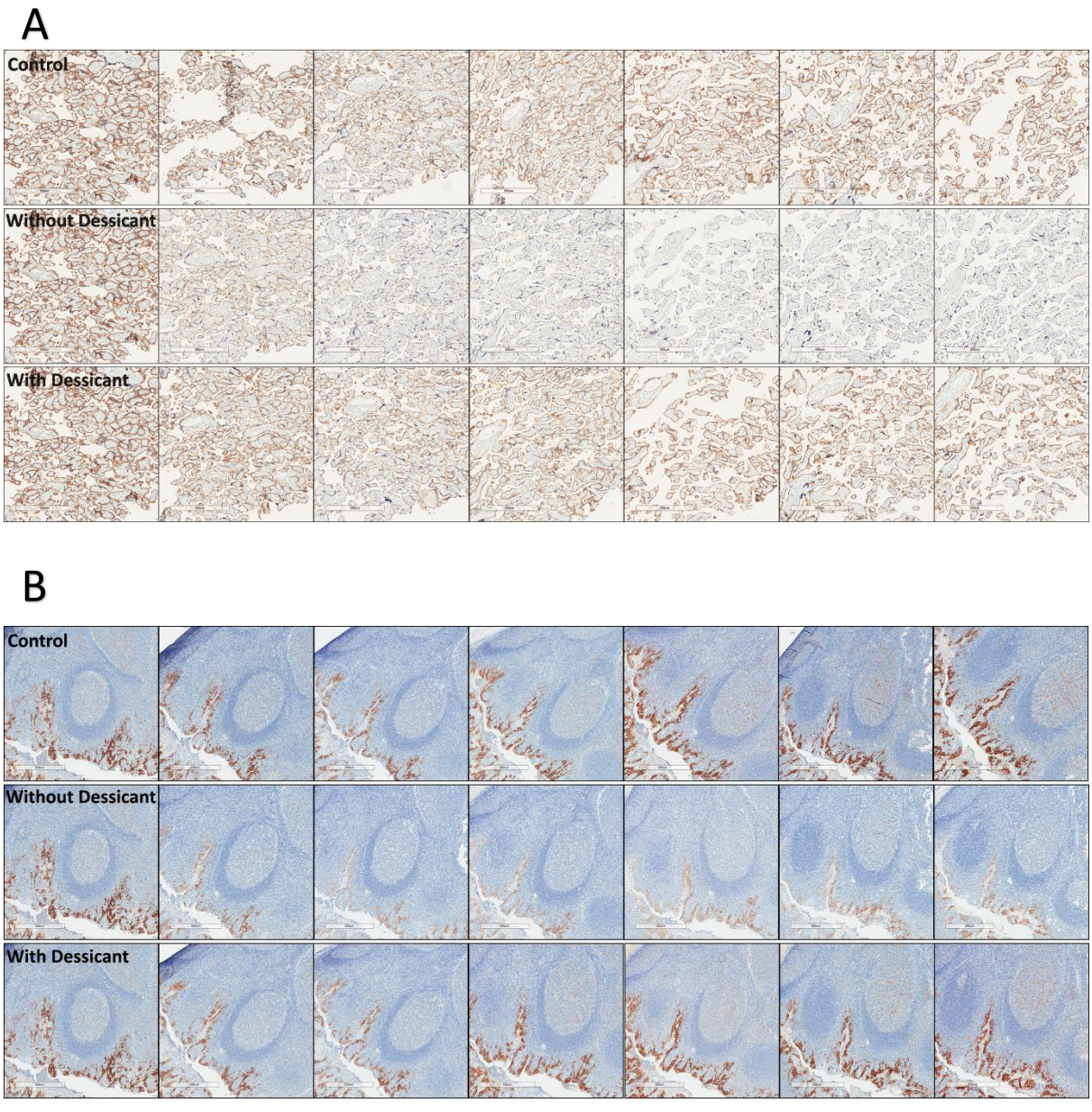
Fig 4.2.11 PD-L1 protein by MS. **A** – Correlation of PD-L1 peptide LQD with PD-L1 22C3 expression by IHC at day 0. **B** – PD-L1 peptides LQD and AEV at days 0, 9 and 28 in NSCLC sections stored in the acceleration chamber at 100% oxygen, 80% humidity and 37 °C. No significant change in the PD-L1 peptides with no easily oxidisable residues (LQD) or with an oxidisable tryptophan (AEV) implying no genuine loss and no significant oxidisation of the PD-L1 protein.

Having confirmed PD-L1 protein was detected by MS and was stable at 28 days, evidence of oxidation was looked for detection of AEV oxidation products, however, none were found, nor was there any change in AEV, suggesting no evidence of PD-L1 protein oxidation. Global proteome analysis did detect some significant levels of oxidation in the methionine peptides at day 28 in the accelerated condition NSCLC samples, as well as in naturally aged placenta and tonsil at 2 years (Dunn's test, NSCLC adj. p = 9 × 10−14; placenta adj. p = 1 × 10−34; tonsil adj p = 8 × 10−58) but despite being significant, the measured change was only modest and could not account for the near total loss of PD-L1 IHC expression seen in paired samples.

4.2.6 The impact of desiccant on preventing immunoreactivity loss

Because humidity was determined to be a major factor in loss of tissue immunoreactivity, the effect of storing slides with or without desiccant in the acceleration chamber set to baseline conditions of 100% oxygen, 37 °C and 80% humidity over 28 days was examined. This was accomplished by measuring the expression of PD-L1 (E1L3N) and pan-CK IHC using positive pixel count scoring (positivity) on placenta and tonsil tissue sections stored over 28 days. Upon removal of slides from the sealed container stored within the acceleration chamber, the humidity level was recorded as

<30% using the enclosed indicator card. Slides stored with desiccant showed no significant loss of either PD-L1 or pan-CK expression in any tissue at day 28 [Average positivity day 0 vs day 28: PD-L1 in placenta 0.57 vs 0.53 (100% vs 93%); p = 0.083, tonsil 0.088 vs 0.073 (100% vs 83%); p = 0.555, pan-CK in placenta, 0.74 vs 0.72 (100% vs 97%); p = 0.311, tonsil 0.33 vs 0.30 (100% vs 91%); p = 0.185]. Qualitative assessment of sections showed slides stored with desiccant demonstrate expression loss similar to sections stored under normal atmospheric conditions, with appreciable loss of PD-L1 expression in sections stored without desiccant. To a lesser extent, loss of pan-CK was observed (Fig 4.2.12). Significant loss of PD-L1 immunoreactivity was seen for placenta tissue stored within the acceleration chamber, with appreciable, but non-statistically significant loss demonstrated in tonsil tissue. (Fig 4.2.13).



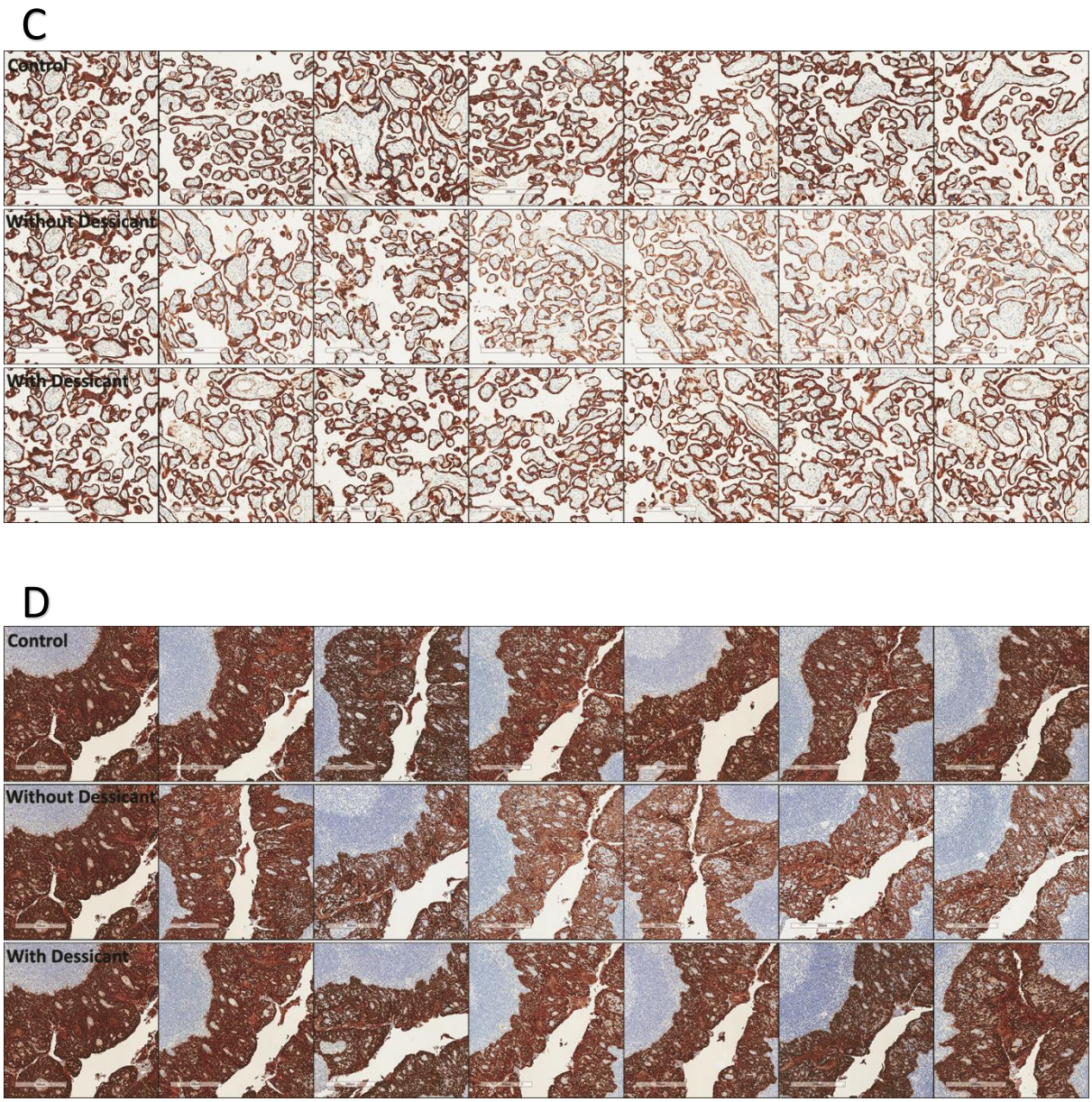


Fig 4.2.12 IHC expression in FFPE tissue sections at days 0, 1, 3, 7, 14, 21, and 28 for **A** – PD-L1 (E1L3N) in placenta, **B** - PD-L1 (E1L3N) in tonsil, **C** – Pan-CK in placenta, **D** – Pan-CK in tonsil. For each figure, the first row shows tissue sections stored under normal ambient conditions and the second and third row shows tissue sections within the acceleration chamber at 100% oxygen, 37 °C and 80% humidity both without (second row) and with (third row) desiccant. The reduction in staining loss overtime in the acceleration chamber with the use of desiccant is seen for both IHC assays and both tissues.

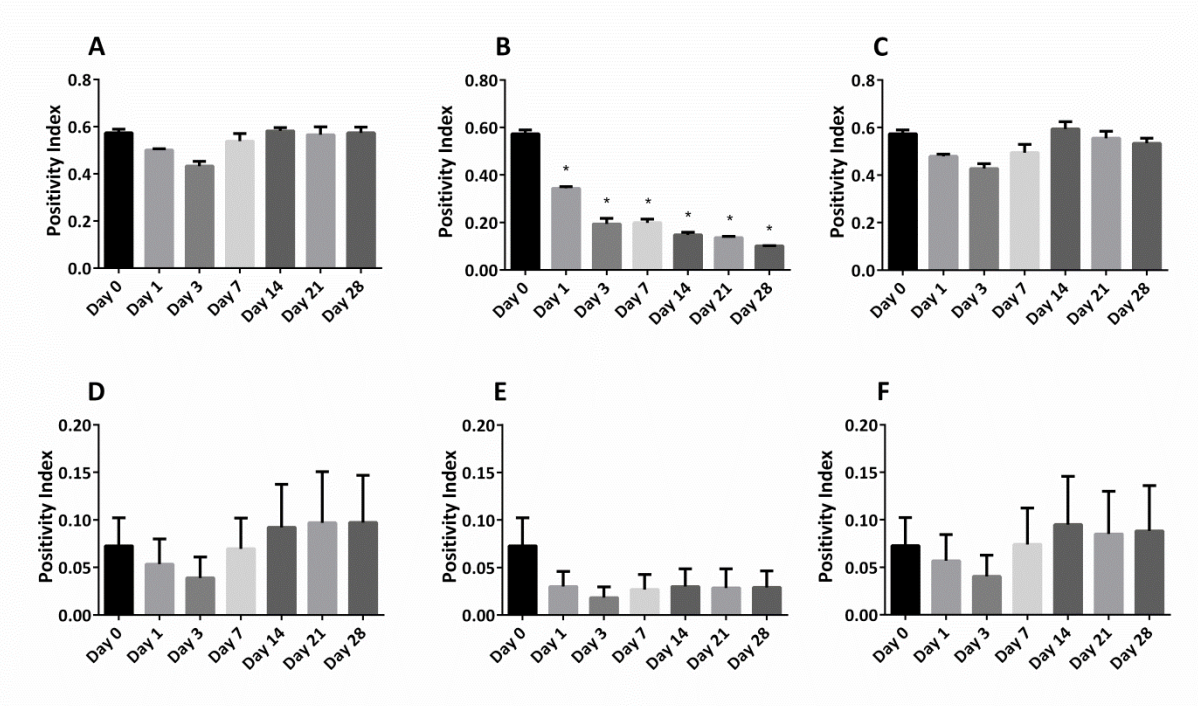


Fig 4.2.13 PD-L1 expression by positive pixel count in placenta (A-C) and tonsil (D-F). Control conditions (A + D): 20°C, atmospheric humidity and oxygen. Case sections stored in the acceleration chamber with conditions of 100% oxygen, 80% humidity, and 37°C without (B + E) or with (C + F) desiccant. B and E represent sections stored without desiccant and demonstrate a noticeable loss of PD-L1 signal, significant for placenta and non-statistically significant for tonsil (likely due to small sample size). Similar patterns are seen between the control sections and those stored with desiccant for both placenta and tonsil. Significant reduction in PD-L1 expression is measured against day 0 values. Bars represent mean ± SEM. *p<0.05.
PD-L1, programmed-death ligand-1.

4.2.7 PD-L1 immunoreactivity loss on specific cell types

Tonsil tissue stained for PD-L1 with E1L3N stored in the acceleration chamber with conditions of 100% oxygen, 80% humidity and 37°C over 28 days was assessed for expression loss difference between crypt epithelium cells and germinal centre immune cells. Day 0 sections showed strong staining within the crypt epithelium, and weaker staining within the germinal centres, in keeping with known PD-L1 expression variation between these cell types.[282] Loss is seen in a stepwise fashion over time, with no difference in the rate of loss seen between the crypt epithelia and the germinal centres. Due to the weaker immunoreactivity associated with these immune cells at baseline, however, signal loss in this cell type was appreciated earlier in the time course and was particularly noticeable when scanning tissues at lower magnification. This observation suggests clinical scoring guidelines that rely upon characterization of weaker staining cells may be impacted differentially by storage conditions. These results are illustrated in Fig 4.2.14.

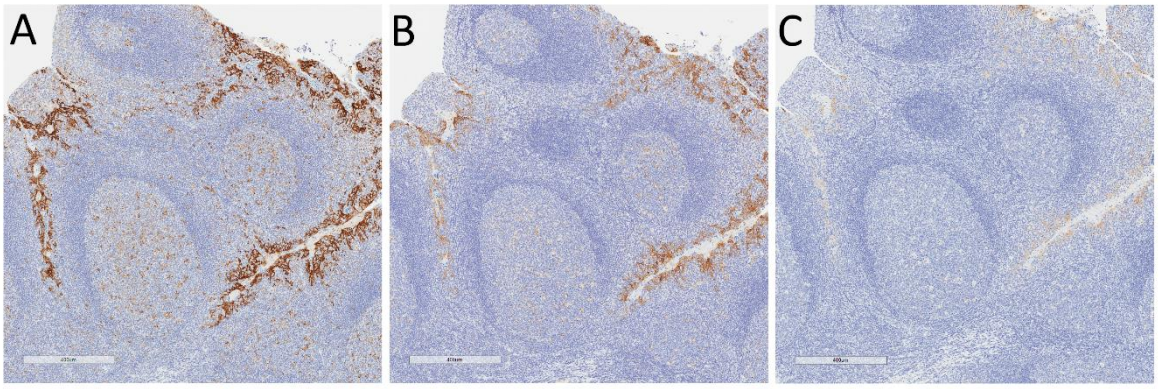


Fig 4.2.14 Tonsil tissue stained for PD-L1 with E1L3N in sections stored in the acceleration chamber with conditions of 100% oxygen, 80% humidity and 37°C at day 0 (**A**), day 14 (**B**), and day 28 (**C**). Loss of both crypt epithelial and germinal centre staining is seen, with more rapid loss of the weaker staining of cells within the germinal centres.

4.3 Discussion

Pre-analytical conditions play an important role in ensuring accuracy and consistency in IHC assays; changes in antigenicity and expression levels as a result of tissue storage or processing will return false positive or false negative results, which could potentially render patients ineligible for treatments they may benefit from, or subjected to treatments which provide no benefit but expose them to a risk of toxicity. Predictive IHC assays that have a quantitative element are particularly vulnerable to loss both in terms of the number of cells deemed 'positive' and the intensity of staining and result in these undesirable clinical situations.

4.3.0 Impact of sampling methods on PD-L1 expression

The impact of sampling methods can be difficult to fully explore due to confounding factors, such as tumour heterogeneity or other pre-analytical factors, but remains a potentially important source of variation in IHC staining. As a result, the question of differing sampling methodologies on predictive biomarkers has been previously studied, including IHC for ER, PR and HER2 in breast cancer, that have been in clinical use for many years longer than PD-L1 IHC has. A meta-analysis looking at 21 papers found a generally good correlation between surgical resections and core biopsies for ER - 92.8% (K0.78) and PR 85.2% (K0.66),[465] with comparable results for both biomarkers also seen when comparing FNAs to core biopsies.[466] Meric-Bernstam *et al*. 2014 looked at a panel of IHC antibodies as well as by protein arrays to compare breast resections and core biopsies, and concluded that some proteins are more sensitive to variation and loss than others,[467] and indeed where ER and PR might remain similar regardless of sampling technique, Ki67 and HER2 often see much more variation.[466, 468, 469] This probably reflects the underlying biology to some extent: Ki67 as a measure of proliferation can have a spectrum of expression patterns, whereas ER tends to be either strongly or

weakly expressed in most tumour cells within a specimen. None the less, this reinforces the importance that pre-analytical conditions may have on proteins with a broad range of expression, including PD-L1. This review of 2,016 cases of PD-L1 expression by SP263 illustrated that there was no significant difference between samples by sampling method or by division into cytology and histology. Indeed, in keeping with the overall findings of the literature review, sampling technique *per se* does not seem to effect PD-L1 IHC, and this cohort is a typical one for PD-L1 expression: approximately a third of samples negative, a third weakly positive and a third strongly positive for PD-L1. Sampling technique may be an important factor for certain predictive biomarkers, but PD-L1 appears to be reasonably resistant to this variation. The most striking difference was dividing the samples by anatomical site, with pleural/pericardial fluids consistently having more cases strongly positive for PD-L1. This will be explored in more detail in Chapter 8 (Predicting response to PD-1/PD-L1 immunomodulatory therapy), but, in the context of the wider analysis, this is more likely to be a result of biological heterogeneity rather than sampling artefact.

4.3.1 Impact of fixatives on PD-L1 expression

Having established that sampling methodology is not a significant cause of PD-L1 variation, the use of differing fixatives was considered. This is particularly important for cytology specimens as many cytology fixatives include the use of alcohol and have been noted to alter antigenicity compared with formalin in a wide variety of tissues and IHC targets.[470] A number of studies have looked at tissue specimens fixed in various fixatives prior to being assessed for PD-L1, with some inconsistent results surrounding the use of alcohol based fixatives in that their use may cause a loss of PD-L1 staining intensity[272, 275] or have no effect.[274, 290] However, other studies studying the impact of fixation or the use of cytology specimens for PD-L1 expression typically have variation in anatomical sites, variation in the nature of the specimens or even the use of unmatched pairs, and with so many potential confounding factors it is difficult to appreciate the true scale of any variation when seen.[272, 274, 275, 290, 303, 307-309, 311, 315, 316, 319, 471, 472] In contrast, this study used matched pairs taken in an identical fashion, from the same site at the same time and processed as cytology specimens. Our study found there was no difference between 10% NBF fixed or alcohol fixed NSCLC samples, either by average TPS, clinical groupings, or indeed by failure rate due to insufficient viable tumour cells, and the intensity and quality of staining was maintained regardless of fixative used. One explanation is that post-fixation in formalin may reverse the deleterious effects of alcohol fixation on protein denaturation[441, 473] though it has been shown that alcohol fixed smears, which have no exposure to formalin, can also be suitable substrates for PD-L1.[274, 315, 316] From a practical perspective, it would be reasonable to state that cytology specimens processed to form cell blocks are suitable for PD-L1 IHC, but it may be prudent to ensure a post-fixation step with formalin when doing so.

4.3.2 Impact of storage conditions on PD-L1 expression

The impact of storage conditions on FFPE tissue antigenicity is an important pre-analytical consideration in NSCLC, as patients with recurrent or advanced disease may not be fit for re-sampling, and so a previously gained sample might be the only tissue on which to perform predictive profiling – particularly crucial if their initial sample was not tested at the time of diagnosis for PD-L1 (for example, their centre did not practice reflex testing or their diagnosis was prior to local use of PD-1/PD-L1 IM). The impact of temporal or iatrogenic heterogeneity resulting in a genuine change in PD-L1 status is considered later in this thesis, but storage conditions may impact antigenicity such that a false change is seen. Generally speaking FFPE blocks are fairly robust for most IHC assays, with blocks as old as 60 years able to produce interpretable IHC stains.[294, 474-477] However, whilst the general observation that certain antibody targets are more prone to loss than others, any given protein of interest varies between studies: IHC for ER has been shown to have no loss in FFPE blocks even 40 or 60 years of age,[475, 476] but also found to have considerable loss, even at 6 months, with total loss seen by 1 year,[294, 474, 477] with mixed observations of loss seen for PR, HER2 and other IHC stains. As previously detailed, the loss of PD-L1 in FFPE blocks is reasonably low in blocks under 6 months, but blocks 3 year or older are at high risk of antigenicity loss and falsely low levels of PD-L1 expression.[283-285, 287, 291] Even more fragile are stored FFPE sections, with loss of a wide panel of IHC markers seen at 6 months.[5] What is particularly striking across these studies is the range of antibodies effected: both polyclonal and monoclonal antibodies may be affected, and loss may occur in assays that target the nucleus, cytoplasm, or membranes of cells and specific IHC signal loss can vary between studies depending on factors such as fixation and the antibodies used.[5, 295, 341, 456, 474, 478, 479]

The mechanisms of antigen degradation have been explored previously, with a variety of potential factors thought to influence the loss, including oxidation, humidity, high temperature and variation in tissue processing.[292, 293, 295, 456-459] Our initial thought that oxidation could potentially have an impact on antigenicity has been seen previously[295, 459] and indeed the MS global proteome analyses indicated that conditions that facilitate accelerated wet-air oxidation caused a statistically significant degree of oxidation that was similar to that seen in placenta and tonsil samples stored under ambient conditions for 2 years. However, this extent of oxidation was likely insufficient to account for the reduction in PD-L1 immunostaining by IHC either in the acceleration experiments, or under normal ambient storage conditions. Furthermore, we found no evidence that changing the storage conditions from low to high oxygen content affected PD-L1 or pan-CK detection by IHC to any significant degree. Previous studies have considered the impact of minimising the effects of oxidation on storage tissue through the use wax-coating unstained sections for the duration of

storage, with some studies finding this had no effect on preventing IHC staining loss[478, 480, 481] and others finding it was preventative, particularly in the context of increasing the beneficial effects of cold storage.[295, 479, 482] Blind *et al.* 2008 considered the role of oxidation through the use of hydrogen peroxide but found this has no effect on immunogenicity.[459] Oxidation undoubtedly appears to play a role in stored tissue, but its impact on antigenicity is variable, and seemingly minimal on PD-L1 IHC.[459]

Previous work suggested that humidity or the fixation process can have an effect on antigen degradation.[292, 442, 483, 484] Aldehyde induced crosslinks between proteins, DNA or chromatin are reversible by AR. In 1991 the first article on AR was published,[485] and it has been more recently acknowledged by a major pioneer of this approach that despite AR not being fully understood, (probably as a result of incomplete knowledge surrounding aldehyde induced crosslinking), it is still widely accepted and used because it delivers reliable results.[442, 486] However, what is generally agreed is that a step is required to break or reverse the crosslinks, and this is typically heat-induced or via proteolytic means (such as by broad proteolytics as proteinase K, trypsin or pepsin).[442, 487] There is also evidence that aldehyde induced crosslinks can also undergo spontaneous hydrolysis, a process that is catalysed by higher temperatures[488, 489] and our results are consistent with previous reports that the presence of water and high temperature are a major cause of antigen loss in FFPE tissue.[292, 456, 480] Indeed temperature appears to be a reasonably consistent factor in contributing to antigenicity loss in stored tissue: storage at high temperatures results in increased loss of IHC staining,[456, 459] and storage at cold temperatures (~4°C) generally results in a reduced loss,[456, 457, 459, 479-482, 490] with further reduction in loss seen if using even colder storage (-20°C)[481] such that storage at -80°C results in no loss of IHC staining intensity of quantity even after 9 years or more of storage.[491] Water appears to have a more complex relationship with antigenicity: excessive dehydrating or rehydrating of tissues during pre-analytical processing could negatively impact IHC staining, as could storage in humid conditions.[292, 295, 455, 484, 492] Therefore, one mechanism by which antigen expression may be lost is heat catalysed hydrolysis of susceptible protein–protein crosslinks, resulting in a change in crosslinked protein structures and loss of discontinuous epitope sites or masking of linear sites. The crosslinking process increases accessibility to some antigens and renders others inaccessible through masking of epitopes.[483, 486] The mechanism of crosslinking and antigen masking is not fully understood, but the masking of antigens in FFPE tissue is more likely with specific amino acid sequences[493, 494] and, importantly, discontinuous epitopes,[494, 495] that are also particularly susceptible to loss in high temperatures.[484] Anti-PD-L1 IHC clones detect a variety of epitope regions, many of which are believed to be discontinuous (28-8, SP263, and SP142)[326, 327, 496] possibly explaining why PD-L1 IHC is particularly sensitive to loss of immunoreactivity. An alternative

explanation for this observation includes the possibility that extracellular epitopes recognized, for example, by 22C3 and 28-8 are particularly accessible, and therefore more susceptible to environmental humidity during tissue storage.

Interestingly, the effect observed with the acceleration chamber was prevented with the use of desiccant stored alongside the sections, to the extent that loss of immunogenicity was the same as tissue stored in normal atmospheric conditions. This suggests that humidity during storage is a major driving force behind immunoreactivity loss. This may be due to epitope conformational changes driven by hydration, which occurs over time during storage. The practical implications of this finding are significant: desiccant may provide an effective method of preventing antigen loss that could be immediately implemented into clinical or research protocols involving the storage and transportation of tissue. This would provide an attractive alternative to other more complicated, time-consuming, and expensive methods of preventing loss such as microwave heating, recoating in paraffin wax, storage under vacuum, or the use of nitrogen chambers.[293, 295, 453, 457, 497]

4.3.3 Limitations

An important note throughout all of these studies is that, regardless of the environmental impact of storage on IHC staining, different antibodies were variably effected by these conditions. For example, Ramos-Vara *et al.* 2014 found the almost total loss of IHC for 3 markers when stored in light (CD45, CD68 and TTF-1) could be largely reduced by storage in darkness, but other markers (p63, PR) suffered no loss in either condition.[457] Some antibodies appear to be consistently more robust or more fragile than others, but crucially, the clinically relevant markers of ER, PR and HER2 had considerable variation between papers. Therefore, a limitation in this study is that not all the clinical PD-L1 clones were subjected to conditions in the acceleration chamber, and the variation between PD-L1 clones and tissues may vary considerably in terms of the magnitude of antigenicity loss.

Other limitations in the acceleration chamber study include sample sizes: while over a thousand tissue sections were included in the analysis, sample sizes were small for some conditions, perhaps accounting for non-statistically significant trends in certain experiments. Positivity as defined by pixel counting was complemented by TPS/CPS scores in tumours to give clinical relevance, but the equivalent is not possible in placenta and tonsil; therefore, although significant loss can be quantitatively demonstrated with good reproducibility the point of clinically relevant loss does not translate for these tissues. Although PD-L1 expression by IHC did broadly correlate with MS findings, not all cases did, and these discrepancies could be accounted for by PD-L1 protein post-translational modifications including glycosylation which may not be detected by IHC but are detectable by MS.[360]

Labelled internal peptide standards used in the targeted analysis of PD-L1 indicated that peak areas were reduced in some, but not all, samples. While the cause of this is uncertain, the possibility cannot be ruled out that the accelerated degradation conditions may have altered those specimens in a manner that reduced recovery of the labelled standards. Inspection of the MS peak areas for the endogenous PD-L1 peptides indicated no loss of protein during accelerated incubation. Finally, the acceleration incubator has demonstrated that select environmental conditions reproducibly affect the loss of IHC expression, but this has focused on PD-L1 and pan-CK in specific tissues. The application of this approach as a wider tool in understanding antigen loss under ambient conditions, and the optimal conditions to predict storage effect on novel biomarkers in development requires further study.

In summary, pre-analytical conditions are important factors to consider in ensuring accurate and consistent PD-L1 IHC stains. Sampling methods and fixation should not be considered major barriers to the use of PD-L1 IHC *per se*, and the inclusion of cytology specimens for predictive IM therapy is therefore wholly appropriate for NSCLC, and most likely other tumour types. FFPE sections should not be left for any significant period of time prior to staining for PD-L1, but if necessary, they should be submitted to the same optimised storage conditions that FFPE blocks are, which this study would suggest is cold, dry conditions, with the use of desiccant.

Chapter 5 – Analytics and post-analytics: Digital pathology and PD-L1 interpretation

5.0 Introduction

In the previous chapter, the effect of pre-analytics on PD-L1 expression by IHC was explored in order to ascertain the impact of tissue handling and storage prior to the assay of choice being applied. This chapter will focus on the use of differing assays for PD-L1 assessment, and the subsequent interpretation of them with digital pathology and image analysis assistance.

5.0.0 Analytics

There are many different IHC assays for assessing PD-L1 expression, each with their own specific properties and characteristics. In addition to the 4 clinically validated clones (SP263, SP142, 28-8 and 22C3) there are a number of other clones used in research or clinical trials, including 73-10, E1L3N, E1J2J, 5H1 and others. There are also numerous platforms on which PD-L1 assessment can be performed (e.g. Ventana Benchmark Ultra, Dako Omnis, Leica Bond-III etc). As a result, and as explored extensively within the literature review, even the 4 clinically validated clones cannot be considered equivalent to each other. The IASLC guidelines and manufacturers recommendations are that the clones are used as part of an approved IVD assay, with specific parameters determined on a specific platform. Even when adhering to these consistent approaches however, there is significant variation between clones, and even subtle variation can be important when tumours are close to critical cut-offs for clinical grouping, (e.g. 1% or 50% TPS) and can be enough to categorise samples into different categories depending on which clone has been used.[297, 304, 305, 353, 498, 499] Several studies have considered the impact of LDTs, in which various PD-L1 clones are used on different platforms with modifications of assay parameters to best suit the individual laboratories requirements. With sufficient standardisation and optimisation, LDTs can be suitable for providing equivalent tests to the validated IVD assays, but inconsistencies and inadequacies are common. Furthermore, IVD assays, in addition to a potentially more rigorous and comprehensive development than LDTs, are often the only assays that have undergone a formal clinical performance evaluation. [500] Thus whilst comparison of LDTs to IVD assays is a reasonable approach in order to save money and time, and broaden the use of a biomarker, it is generally accepted that an approved IVD assay will result in the highest accuracy.[266, 305, 328, 331, 332, 353]

Throughout this thesis, the majority of work has used the SP263 clone on the Ventana Benchmark Ultra following strict IVD assay protocols. This has been beneficial due to its validation as a clinical complementary diagnostic, my familiarity with scoring it due to my clinical experience, its widespread use in research, and my ongoing collaboration with Roche Ventana (whom manufacture

the clone) that has allowed for applications of novel technologies to the stained tissue. However the 22C3 clone on the Dako ASL48 was used for the CCC cohort (as this was the assay of choice in routine clinical practice at our laboratory at the time of these patients' diagnosis), and so the concordance of these two clones are of particular relevance. Studies that compared 22C3 and SP263 generally find them to be similar[288, 304, 305, 328, 343, 344, 499, 501], but not necessarily equivalent.[297, 305, 353, 498] Comparing different studies against each other is challenging, as they vary in regards to the platforms used, the tissue samples involved and other pre-analytical and analytical factors, including different pathologists' interpretation. This project will consider the use of different PD-L1 clones tested in the same laboratories, in order to minimise pre-analytical variables, and scored by the same pathologist (myself) in order to reduce inter-observer variation, with a view to providing some data to the applicability of the SP263 specific findings of this thesis to other PD-L1 clones.

5.0.1 Post-analytics: Interpretation of PD-L1 expression

PD-L1 expression is one of the most difficult IHC assays to interpret. IHC stains are typically interpreted in a qualitative fashion: the presence or absence of staining, the intensity of staining and the location of staining in regards to the cells of interest and the cellular components of interest. Interpreting PD-L1 expression by IHC requires all of these, but also has a quantitative requirement to produce a percentage score of relevant cells staining for PD-L1. The clinical implications of scoring PD-L1 can render patients eligible or ineligible for specific treatments, and at crucial clinical categories the difference in treatment decisions can be made by a difference of just 1%.[297, 304, 305, 353, 498, 499] As a result, advice is for pathologists to be trained in the interpretation of specific clones and for a regular case-load involving PD-L1 interpretation to ensure ongoing clinical competency.[266, 282, 323]

Challenges involved in PD-L1 expression include the expression of PD-L1 by immune cells, particularly macrophages that can be difficult to distinguish from tumour cells on non-H&E sections, the aberrant expression by necrotic, apoptotic and other non-viable tumour cells, and the interpretation of the cellular location of the stain (Fig 5.0.1) These largely qualitative features can be overcome by an experienced pathologist whom can reliably differentiate these findings with the aid of H&E sections and interpretation guides, but an arguably greater challenge is that of the quantitative nature of PD-L1 IHC interpretation. PD-L1 scoring in NSCLC requires a percentage score of the number of positive tumour cells over the total number of tumour cells – the TPS – and is a requirement for guiding clinical management of patients.[47, 48, 254, 255] Whilst various guides have been developed to aide pathologists with this, it remains a challenging area. The use of novel technologies to help assist the pathologist with the quantification of PD-L1 IHC would likely prove to be an

invaluable tool and one that may help better predict response to PD-1/PD-L1 IMs. The use of digital pathology and image analysis software to augment pathologists' quantitative interpretation has seen success in more straight forward IHC stains[502, 503] but applying these techniques for the more complex PD-L1 expression is likely to be significantly more challenging.

5.0.2 Digital pathology and image analysis

The use of digital pathology, that is, the scanning of glass slides to produce images viewable on a computer screen, has been recognised as being equivalent to conventional pathology in a wide array of pathological areas,[504-507] including general IHC[508] and PD-L1 expression analysis,[304] and also provides an array of practical advantages such as ease of sharing cases over wide geographical distances, the permanent storage of images and a more convenient platform for a variety of uses including education EQAs and MDTs (Multi-disciplinary teams).[509-513] The digitisation of pathology slides allows for ambitious novel applications, including the use of 'big data', and machine learning to produce predictive models or be combined with other data sets such as radiology or 'omics' data.[512, 514] The complexity of these methods means most are still in development, although the limited use of supervised deep learning software has been approved by the FDA.[515, 516]

A more immediately applicable use for digital pathology is its use as an augmentation of the routine pathologist's workload with digital image analysis. Current approaches include assessing nuclear pleomorphism and mitosis, or measuring distances for tumour depth or surgical margins.[517, 518] The use of digital pathology in interpreting IHC stains has been plausible for some time, as the nature of the staining makes it easier for a computer to interpret than H&E,[519-521] and can be sufficiently accurate to quantify predictive markers such as ER and PR in breast carcinomas.[502, 503] The use of these approaches to assisting PD-L1 interpretation is an exciting prospect, and would help to combine pathologists' expertise in PD-L1 expression qualification (i.e. what should be counted) with the computers' ability in PD-L1 expression quantification (i.e. how much there is).

5.0.3 Inter-observer concordance of PD-L1 interpretation

Inter-observer disagreement between pathologists is a widely recognised phenomenon in a variety of areas, including morphological diagnostics[522] predictive IHC stains such as Her-2,[523] and classifying immune-infiltration patterns[524]and it has been suggested that inter-observer discordance in interpreting PD-L1 might account for even more variability than the use of different PD-L1 clones.[335] The use of digital pathology to assist in inter-pathologist concordance has been attempted in a variety of areas,[525, 526] and the use of digital pathology and image analysis to improve inter-observer concordance in interpreting PD-L1 would be a valuable tool. If digital pathology slides and

algorithmic assistance can be used to reliably score PD-L1 expression and provide a more accurate 'absolute value' for a TPS, it would be hoped that improved inter-observer scoring would also be seen, and therefore minimise the impact of a considerable source of potential error.

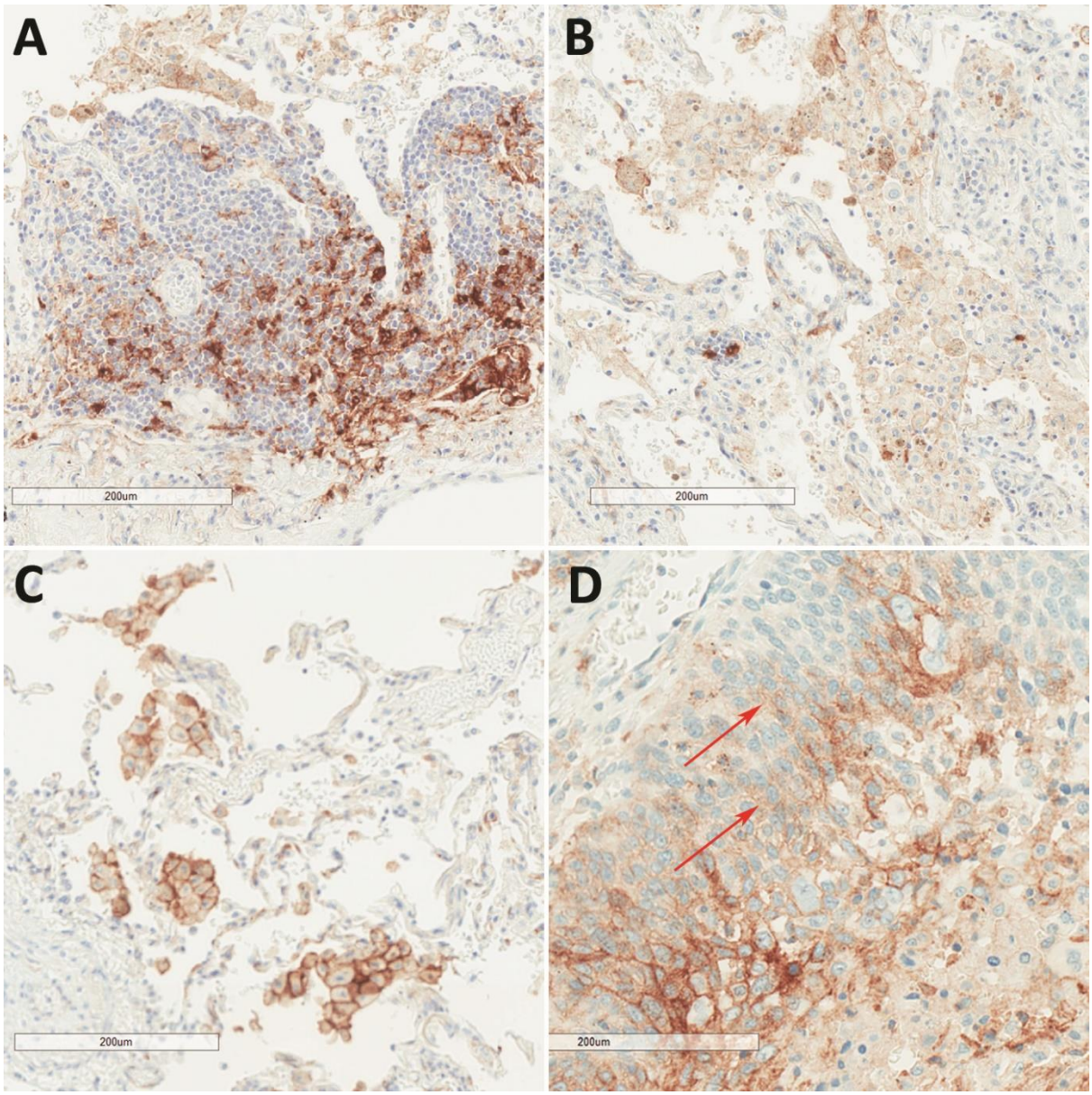


Fig 5.0.1 Challenges in assessing PD-L1 expression to generate a TPS. **A** – TIL expressing PD-L1, **B, C** – Macrophages expressing PD-L1. **D** – Tumour cells with faint cytoplasmic staining for PD-L1 (red arrows)

This chapter therefore sets out to address the issues of differing PD-L1 clones and their equivalency to each other when scored by a single pathologist. The accuracy in interpreting PD-L1 through the use of digital pathology and algorithmic image analysis will be explored as well as its ability to impact inter-pathologist concordance scored by multiple pathologists. Finally, the benefits and limitations of digital pathology and the image analysis tools utilised in this project will also be considered in their potential application to routine pathology workloads.

5.1 Methods

5.1.0 Comparing PD-L1 22C3 and SP263 IHC antibody clones

A retrospective review of 1,400 consecutive NSCLC cases scored for PD-L1 by a single pathologist in the routine clinical setting was performed to collect data on tumour type, specimen type and PD-L1 TPS. Data was fully anonymised and divided according to the PD-L1 clone used, so that a comparison between pathological metrics, including TPS, specimen type and anatomical location of specimen could be made to look for differences in the quantity of staining between the 22C3 and SP263 clones. Average TPS and clinical categories (<1%, 1-49% and ≥50% TPS) were used to compare the clones.

5.1.1 Comparing PD-L1 22C3, 28-8, E1L3N and SP142 IHC antibody clones

10 cases of NSCLC from the Eli Lilly cohort underwent staining for 4 different PD-L1 clones as described in Chapter 4. In Chapter 4, these findings mostly pertained to the variable loss of PD-L1 expression over time due to storage conditions, but the differences between the clones will be described in more detail here. A TPS for each section and clone (on unaged tissue) is scored for each case alongside a positive pixel count using the Aperio ImageScope integrated image analysis 'Positive Pixel Count v9' algorithm[338] and compared to each other using TPS values, clinical categories and positive pixel counts.

5.1.2 uPath and use of the PD-L1 IHC interpretation algorithm

Glass slides of tissue sections stained for H&E and PD-L1 (SP263) as per main methods section were scanned using the Roche DP200 scanner at 20x magnification to produce *.BIF images. Regions of interest (ROI) were applied to the glass slide snapshots to ensure all tissue of interest was included before being scanned at full magnification (Fig 5.1.0). Images were digitally transferred to the uPath system where each case was manually assigned their respective H&E and PD-L1 tag. The uPath system is a Roche developed digital pathology platform that allows for viewing of digital images of scanned slides and the correlation of these data with other clinical details and demographic details, with an ultimate objective of being developed into a tool for use by all aspects of the MDT.[432] Each

PD-L1 section can be viewed in uPath alongside its H&E and assessed to give a manual TPS (Fig 5.1.1). The PD-L1 tagged images could then be analysed by the Roche-PD-L1-algorithm (RPA).[527] The RPA is a Roche developed image analysis tool that identifies and distinguishes PD-L1 positive tumour cells from PD-L1 negative tumour cells and ignores other cell types. It is currently for use in research only and is being improved and developed in a continuous fashion. Due to collaboration with Roche as part of the Northern Pathology Imaging Consortium (NPIC), Liverpool was the first site in the UK to have access to the RPA and my research the first UK based project to involve its use. Images of PD-L1 sections viewed on uPath can then have annotations drawn around the tissue of interest so as to include these areas for PD-L1 assessment and have the RPA applied to these regions. (Fig 5.1.2) Each annotation returns a total tumour cell count, a positive for PD-L1 cell count and a negative for PD-L1 cell count to give a TPS, which would then be averaged to include all annotations across the section, to give an unadjusted RPA TPS (Fig 5.1.2). Each case was then reviewed with the corresponding unadjusted RPA TPS to reach a final agreed TPS for each section based on the RPA and pathologists' assessment. Alternatively, the RPA can be applied to the entire section without the use of annotations. Four TPSs could thus be generated for each case:

- A manual TPS (no use of the RPA)
- Whole-section TPS (RPA applied to the entire slide)
- Unadjusted RPA TPS (RPA applied to annotated regions)
- RPA assisted TPS (Combined score using RPA and pathologists' judgement)

Whenever the RPA is used throughout this project, both an unadjusted and RPA assisted TPS is generated as a minimum.

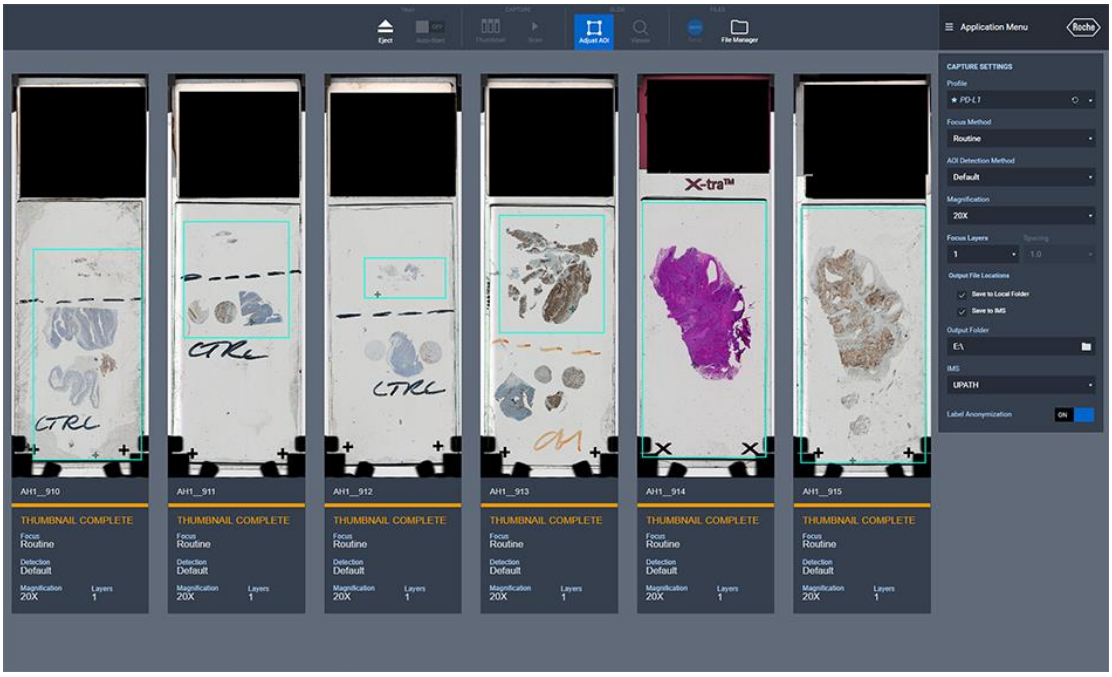


Fig 5.1.0 DP200 slide scanner user interface. Regions of interest (turquoise) are applied around the tissue of interest. Control tissue can be included or ignored as per user preference.

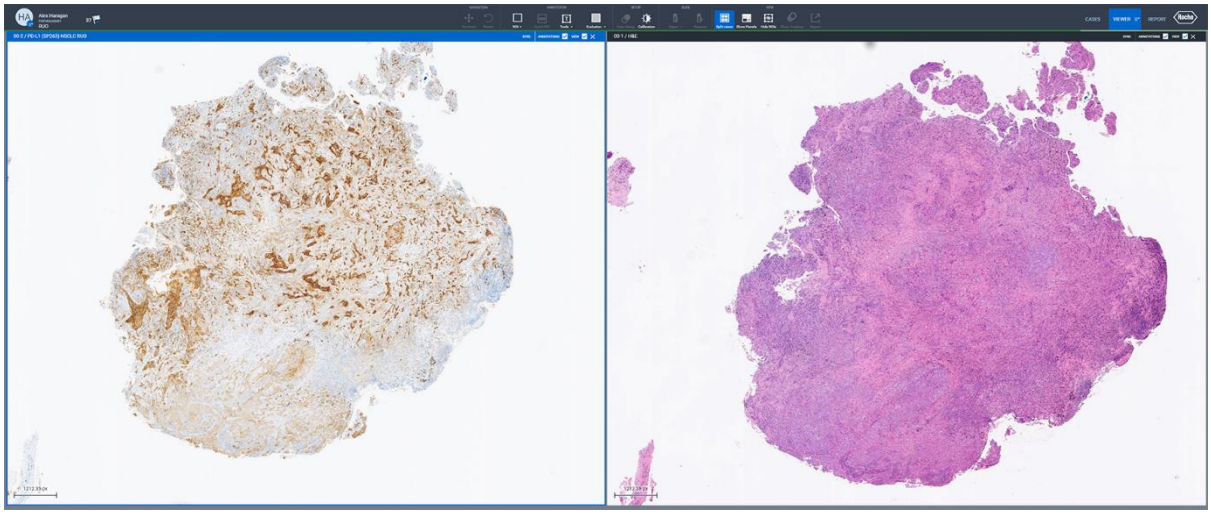


Fig 5.1.1 uPath view of NSCLC stained for PD-L1 and H&E. Images are registered so as one image is moved, the other is moved synchronously. This allows for manual scoring of PD-L1 alongside the relevant area of tumour on H&E.

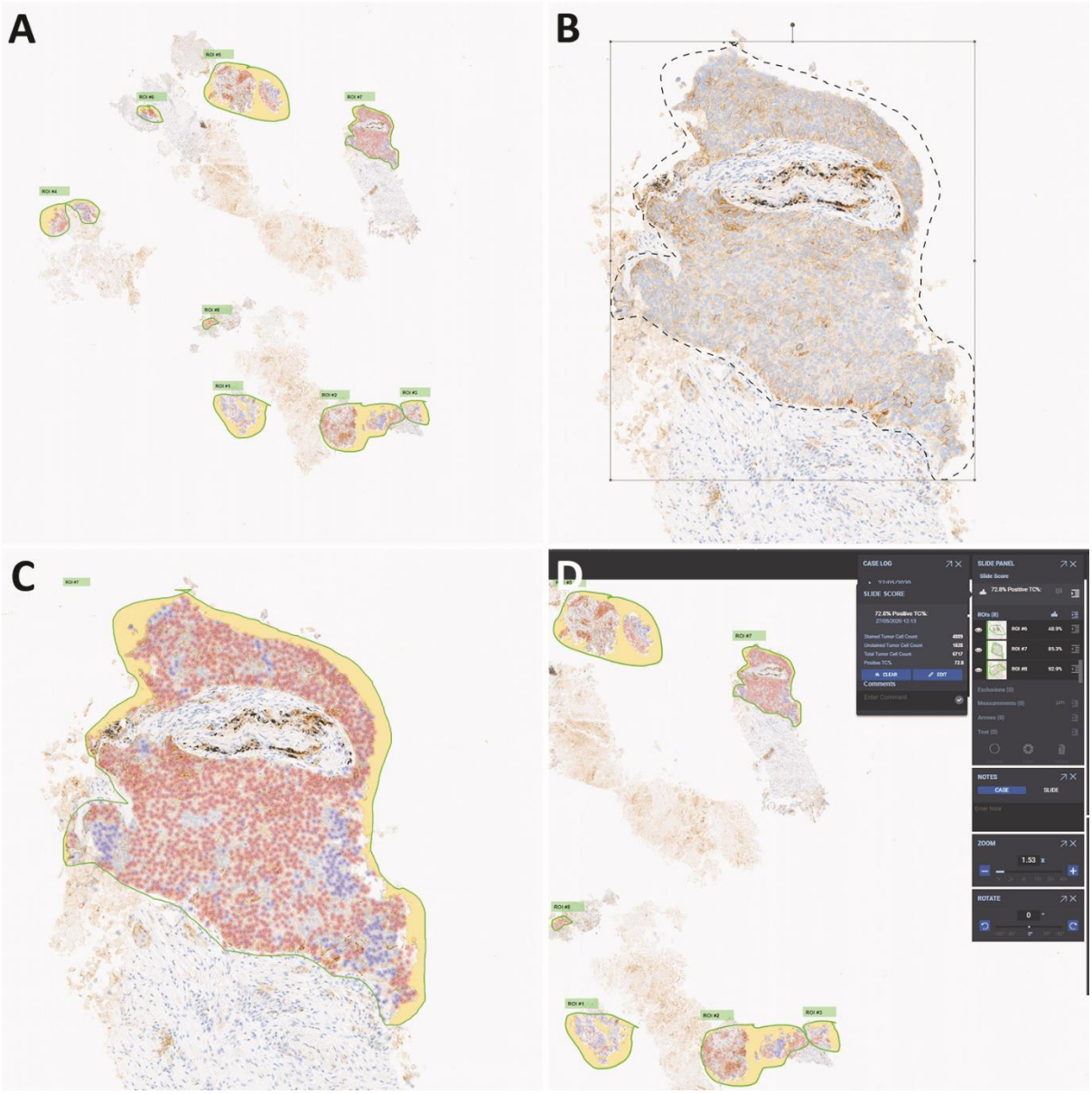


Fig 5.1.2 Applying annotations to NSCLC tissue in uPath. **A** - Multiple annotations are drawn around each area of interest. **B** – Higher power view of annotated region. **C** – Annotated region with RPA applied; note the PD-L1 positive cells (red) and PD-L1 negative cells (blue). The stroma and anthracotic pigment in the central area is ignored. **D** – Final data for all annotations: 8 ROI totalling 6717 tumour cells, 4889 PD-L1 positive cell, 1828 PD-L1 negative cells to give an unadjusted RPA TPS of 72.8%.

5.1.3 RPA in the LLP cohort

Cases from the LLP cohort were assessed for PD-L1 using uPath and RPA to act as a training set in order to provide familiarity with the uPath system and RPA platforms. A manual TPS of glass slides via traditional microscope assessment was compared to manual scoring by the uPath system and

both compared to an unadjusted RPA TPS and an agreed RPA TPS. Average TPS and clinical categories (<1%, 1-49% and ≥50% TPS) were used to compare scoring methods by Spearman's correlation, ICC, Cohen's Kappa and predictive values. Discrepancies between manual TPS and RPA TPS were reviewed to provide qualitative data on the limitations of the algorithm and to understand specific pathological features that presented the RPA with an interpretive challenge. Based on these findings, a training set was developed, with examples of strengths and weaknesses of the RPA, in a bid to develop an approach that would allow for other users to be introduced to the uPath system and RPA.

5.1.4 Training to use the RPA

Two pathologists from the RLUH (Professor John Gosney and Dr Piya Parashar), both of whom report pulmonary pathology and are trained in and regularly report PD-L1 assessments in NSCLC, agreed to participate in this study. Using the initial training set developed above, I introduced them to the use of uPath and RPA. They then participated in an inter-observer and intra-observer project as detailed in the following section, during which I collated their qualitative feedback about the algorithm and their suggestions for use of the RPA and the limitations of the RPA. A final training approach was then developed that incorporated the views of all pathologists and their experience over the course of this project, to produce a potential training series and webinar designed to introduce any pathologist to the use of RPA within uPath.

5.1.5 RLUH Cohort – Intra-observer concordance

107 cases from the RLUH cohort were assessed for PD-L1 expression intra-observer concordance in the following way. 83 cases were scored by myself manually using the uPath viewer. After a 6 week washout period these were re-scored using the RPA to give both unadjusted and agreed RPA TPSs. A subsequent manual rescoring of the specimens was performed following another six week washout period, and final rescoring with the RPA after another six week washout period resulted in six scores across four assessments by the same pathologist of the same 83 cases: two manual, two unadjusted RPA and two RPA agreed TPSs. Separately, 25 cases were scored independently by pathologist A or pathologist B and re-scored by the same pathologist using the RPA after a 6 week washout period. Intra-observer concordance using manual and/or RPA TPSs for each of the three pathologists could thus be recorded.

5.1.6 RLUH Cohort - Inter-observer concordance

50 cases from the RLUH cohort were scored for a PD-L1 TPS by two pathologists independently of each other in the following way. 25 cases were scored manually for a TPS by either pathologist A or

pathologist B, and re-scored manually by Pathologist C. The other 25 cases were scored with the RPA by either pathologist A or pathologist B, and re-scored with the RPA by Pathologist C. In this way inter-observer concordance between two pathologists could be compared using either manual scoring or RPA scoring for PD-L1 expression.

5.1.7 Eli Lilly Cohort – Inter-observer concordance

To establish an idea of inter-observer variation when using digital pathology and annotations in a relatively simple fashion, 3 colleagues (Dr Aaron Gruver, Ms Dimple Das, Mr Michael Soper) applied the positive pixel count algorithm to 30 TMA cores from the Eli Lilly cohort stained for PD-L1 (E1L3N clone). In so doing the variation between 4 observers, including two non-pathologists, when using a digital system could be calculated.

Finally, all 122 cases from the RLUH cohort were assessed for PD-L1 expression using the RPA applied to the entire section and compared to the agreed RPA TPS and the manual TPS for each case.

5.1.8 Statistics

Comparison of clones for PD-L1 was performed using Wilcoxon signed-rank test for TPSs as a continuous variable and Chi Squared Goodness of Fit test for TPSs divided into clinical categories of negative, weak and strong positive (<1%, 1-49%, ≥50% TPS). Comparison between scoring techniques for PD-L1 TPS was performed using Spearman's correlation, and ICC for raw TPSs and Cohen's kappa for TPS scores divided into clinical categories. Negative predictive value (NPV), positive predictive value (PPV) and OPA was calculated for differing PD-L1 scoring approaches using a dichotomous division of ≥1% or ≥50% TPS, with an OPA also calculated for both cut-offs. All significances are taken as $p<0.05$.

5.2 Results

5.2.0 22C3 vs SP263

A total of 1,400 cases scored for PD-L1 expression were identified in this retrospective review. 1000 cases were scored by the SP263 clone and 400 by the 22C3 clone, examples shown in Fig 5.2.0. There was no significant difference in mean TPS 32 vs 30 ($p=0.235$), and no significant difference in the categorisation of tumours into clinical categories ($p=0.115$), results shown in Table 5.2.0. Distribution of specimens by anatomical locations and sampling technique are similar for both clones. Data shown in Table 5.2.1

Clone	SP263	22C3
Total	1000	400
<1%	332 (33)	119 (30)
1-49%	304 (30)	117 (29)
>50%	306 (31)	140 (35)
Inadequate	58 (6)	24 (6)

Table 5.2.0 PD-L1 expression by either SP263 or 22C3 clones across 1,400 NSCLC specimens separated into clinical categories of negative, weak or strong positive expression (<1%, 1-49%, ≥50% TPS).

.

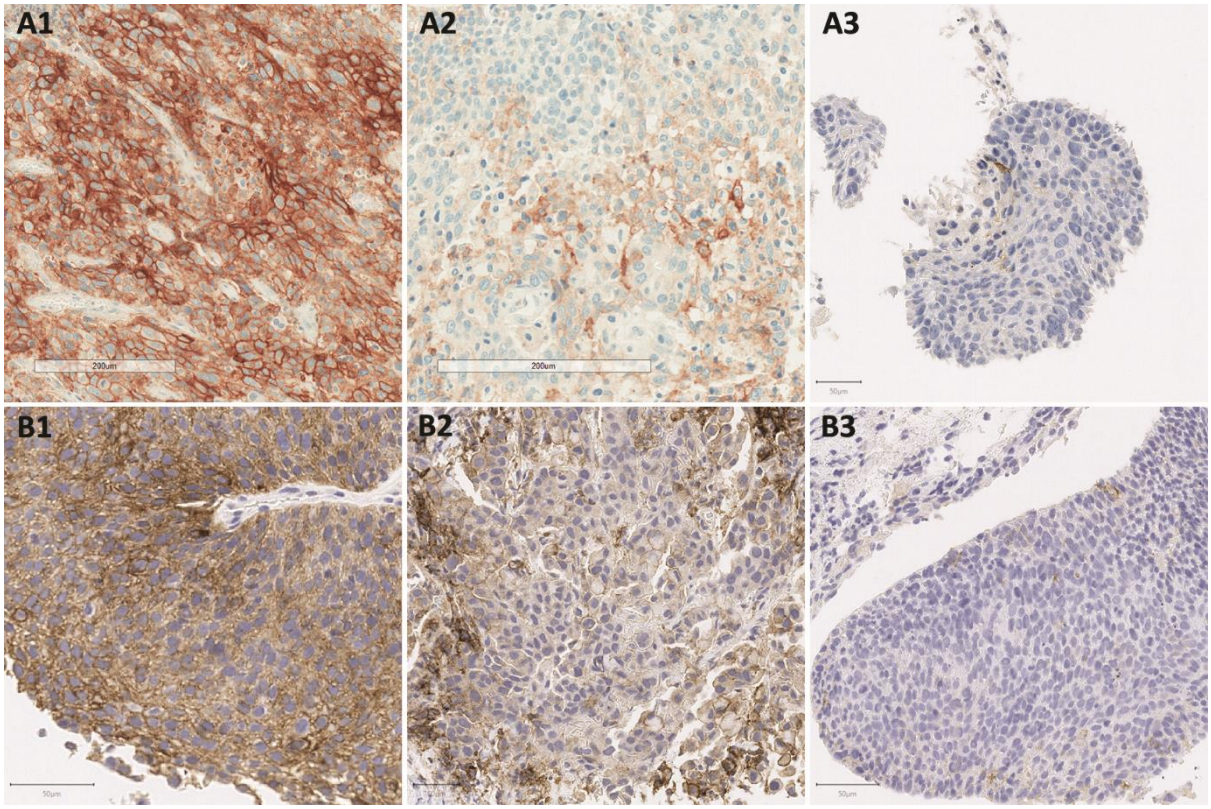


Fig 5.2.0 Examples of PD-L1 expression by SP263 (**A1-A3 A1-A3**) or 22C3 (**B1-B3 A1-A3**) **A1/B1 –** Strong (≥50%) TPS **A2/B2 –** Weak (1-49%) TPS **A3/B3** – Negative (<1%) TPS. SP263 generally produces a sharper and more intense stain than 22C3.

			SP263 (n)	(%)	22C3 (n)	(%)
Specimen						
Biopsies			308	31	156	39
Aspirates			659	66	226	57
Resections			26	3	15	4
BWs/BBs			7	1	3	1
Site						
Primary			559	56	179	45
Regional LN			271	27	134	34
Pleura			74	7	37	9
Distant Mets	96	10	50		13	

Table 5.2.1 PD-L1 expression by either SP263 or 22C3 clones across 1,400 NSCLC specimens separated into specimen type and anatomical site of sample.

5.2.1 22C3 vs 28-8 vs E1L3N vs SP142

10 cases of NSCLC from the Eli Lilly cohort were scored for PD-L1 by 4 different clones to give a TPS score and a positive pixel count for each case. (Table 5.2.2).

Case	TPS				Positive pixel Count			
	22C3	**28-8**	**E1L3N**	**SP142**	**22C3**	**28-8**	**E1L3N**	**SP142**
1	50%	60%	80%	2%	0.183	0.523	0.480	0.024
2	100%	99%	100%	60%	0.567	0.523	0.646	0.199
3	95%	90%	95%	<1%	0.531	0.395	0.538	0.023
4	60%	70%	80%	20%	0.250	0.277	0.312	0.106
5	1%	2%	1%	0%	0.135	0.110	0.151	0.019
6	0%	0%	0%	0%	0.015	0.044	0.025	0.009
7	<1%	1%	50%	0%	0.035	0.083	0.149	0.024
8	2%	2%	2%	0%	0.091	0.123	0.122	0.050
9	<1%	1%	1%	0%	0.109	0.121	0.086	0.021
10	1%	1%	1%	0%	0.164	0.157	0.176	0.029
Average	**31%**	**33%**	**41%**	**8%**	**0.21**	**0.24**	**0.27**	**0.05**

Table 5.2.2 PD-L1 expression by 22C3, 28-8, E1L3N or SP142 in 10 matched NSCLC specimens scored by TPS or per Aperio Positive Pixel Count algorithm.

There was a significant reduction in PD-L1 TPS of SP142 compared to 22C3 (8% vs 31% p=0.008), 28-8 (8% vs 33% p=0.007) and E1L3N (8% vs 41% p=0.007) with no significant difference seen between 22C3, 28-8 and E1L3N (31% vs 33% vs 41% p=0.093). A significant reduction in staining intensity by positive pixel count was also seen for SP142 against 22C3 (0.05 vs 0.21 p=0.005), 28-8 (0.05 vs 0.24

p=0.005) and E1L3N (p=0.005), with no significant difference seen between 22C3, 28-8 and E1L3N (0.21 vs 0.24 vs 0.27 p=0.082). Placing the tumours into clinical categories for PD-L1 expression also showed a significant reduction in strong and weak positive cases for SP142 compare to 22C3 (p=0.02) 28-8 (p=0.007) and E1L3N (p=0.008) but no significant difference between 22C3, 28-8 and E1L3N (p=0.156). Results are summarised in Fig 5.2.1. An illustrative example of staining for each PD-L1 IHC clone is shown in Fig 5.2.2.

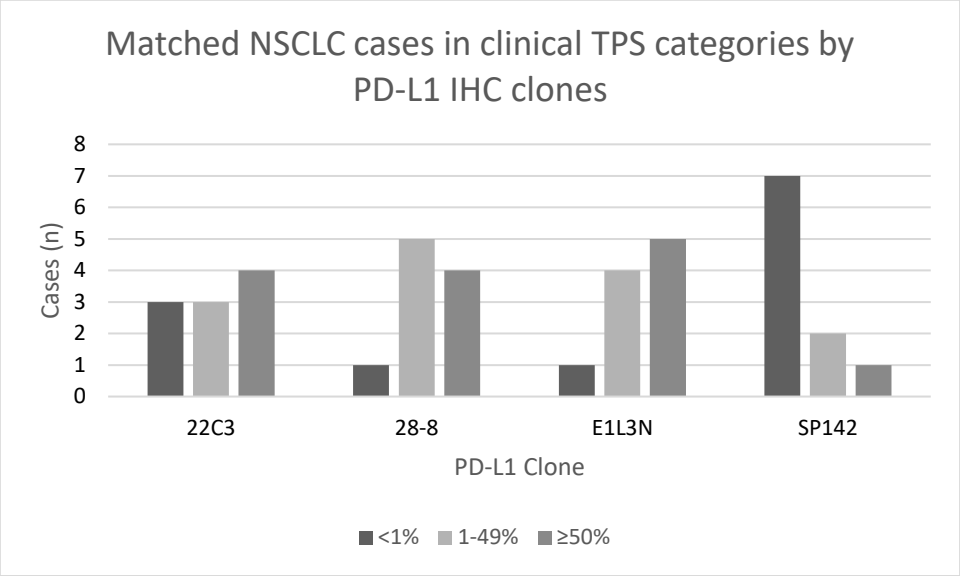


Fig 5.2.1 PD-L1 expression by 22C3, 28-8, E1L3N or SP142 in 10 matched NSCLC specimens scored by TPS and separated into clinical categories of negative, weak or strong positive expression (<1%, 1-49%, ≥50% TPS).

5.2.2 Scoring PD-L1 expression with the RPA in the LLP cohort

113 cases from the LLP cohort were assessed for PD-L1 (SP263) expression manually and with the RPA with a minimum of a 6 week washout period between each scoring attempt. Four scores were thus acquired: the glass slide (conventional microscopy) manual TPS, the digital manual (using uPath) TPS, the unadjusted RPA score, and the RPA assisted TPS. Manual scoring by microscope and digital pathology correlated well (ICC 0.981 p<0.0001, Fig 5.2.3) with a high level of agreement for placing tumours into clinical categories (Cohen's Kappa 0.828 p<0.0001). Use of the unadjusted RPA TPS generally resulted in poorer levels of correlation and agreement for clinical categories, but use of the RPA assisted score showed good correlation with both microscope and digital manual scoring for overall TPS (ICC 0.935, 0.948 p<0.0001 respectively) and placement into clinical categories (Cohen's Kappa 0.72, 0.789 p<0.0001 respectively). Results are shown in Table 5.2.3 and Fig 5.2.3.

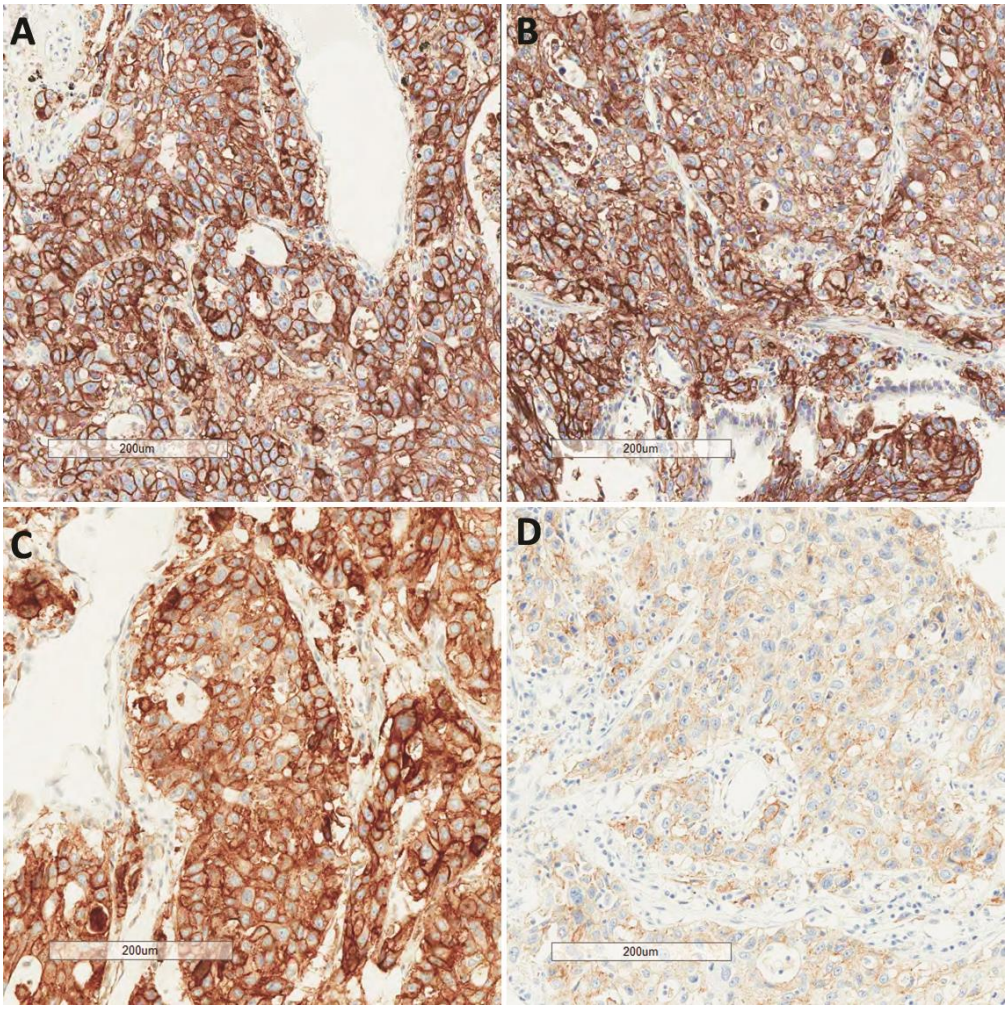


Fig 5.2.2 Examples of PD-L1 expression by 22C3 (**A**), 28-8 (**B**), E1L3N (**C**) or SP142 (**D**) in a matched NSCLC specimens

	Correl.	Sig?	ICC	Sig?	Kappa	Sig?
Glass vs Dig. Man	0.97	Y (p<0.0001)	0.981	Y (p<0.0001)	0.828	Y (p<0.0001)
Glass vs unadj. RPA	0.875	Y (p<0.0001)	0.876	Y (p<0.0001)	0.479	Y (p<0.0001)
Dig Man vs aunadj. RPA	0.876	Y (p<0.0001)	0.885	Y (p<0.0001)	0.497	Y (p<0.0001)
RPA assist vs unadju. RPA	0.883	Y (p<0.0001)	0.897	Y (p<0.0001)	0.464	Y (p<0.0001)
Glass vs RPA assist	0.935	Y (p<0.0001)	0.956	Y (p<0.0001)	0.72	Y (p<0.0001)
Dig Man vs RPA assist	0.948	Y (p<0.0001)	0.968	Y (p<0.0001)	0.789	Y (p<0.0001)

Table 5.2.3 Correlation and scoring agreement of assessing 113 cases of NSCLC for PD-L1 (SP263) by manual glass slide (glass), manually via uPath (Dig. Man), or with the aid of RPA to give an unadjusted RPA TPS (unadj. RPA) and a RPA assisted TPS (RPA assist).

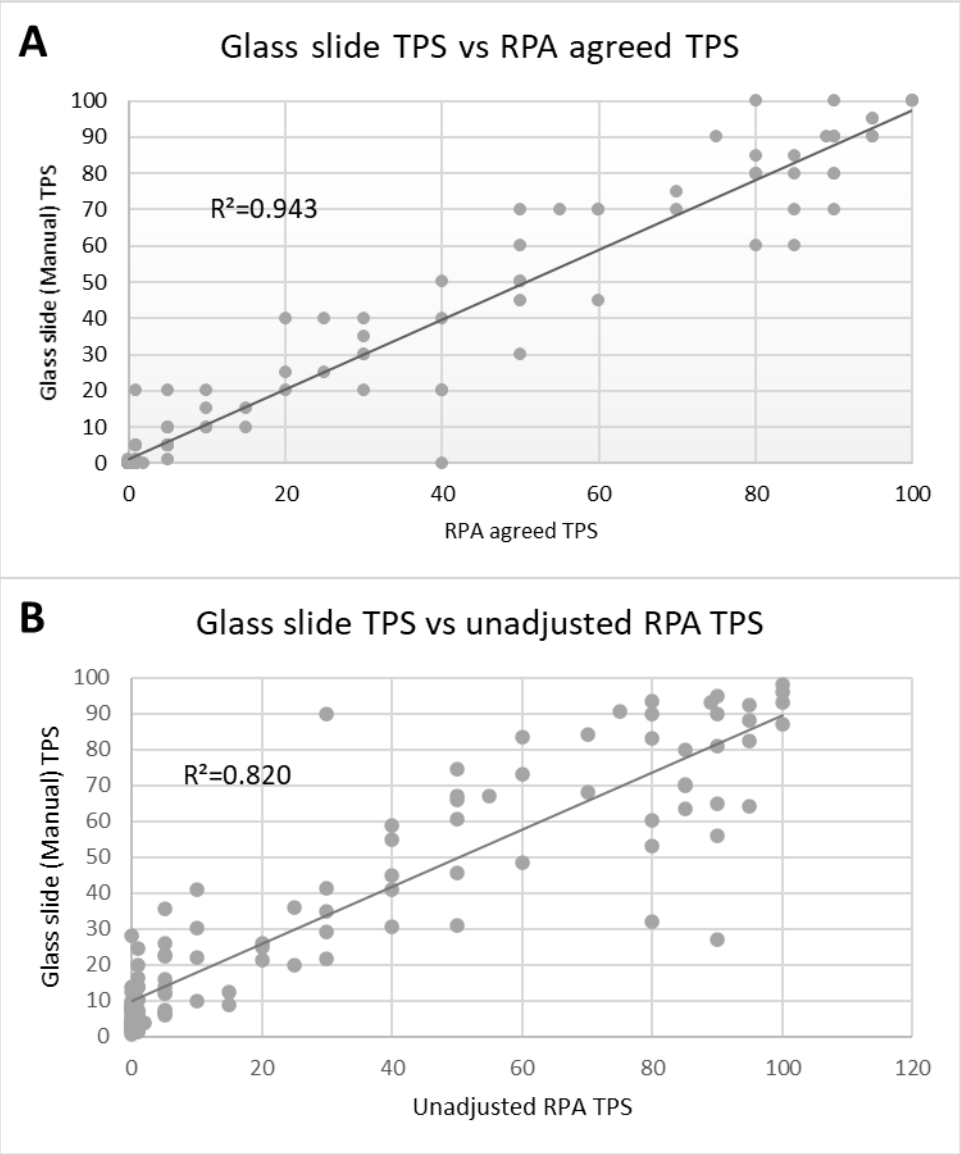


Fig 5.2.3 Linear regression for scoring 113 cases of NSCLC for PD-L1 (SP263) comparing manually via conventional microscopy on glass slides to **A**) the agreed score of the RPA and **B**) the unadjusted RPA

The predictive values of unadjusted RPA and RPA assisted TPSs, using digital manual scores as the 'ground truth' are shown in Table 5.2.4. Cohen's Kappa was generally poor for the unadjusted RPA scores largely as a result of misclassification of <1% TPS tumours as ≥1% TPS. Only 1 case (4%) was scored as <1% TPS when using unadjusted RPA TPSs compared to 24 cases (96%) using the RPA assisted TPSs.

	OPA	NPV	PPV
Unadj. RPA 1%	0.79	1	0.78
Unadj. RPA 50%	0.93	0.94	0.95
RPA assist 1%	0.9	0.73	0.98
RPA assist 50%	0.96	0.96	0.97
Unadj. RPA Both	0.71		
RPA assist Both	0.86		

Table 5.2.4 Predictive power of using either unadjusted RPA (Unadj.RPA) or RPA assisted (RPA assist) to score PD-L1 relative to original manual assessment.

5.2.3 Qualitative data from the LLP cohort

The LLP cohort is composed entirely of resection cases, and was used to understand the strengths and limitations of the algorithm. Careful application of the annotations was required to achieve accurate and helpful guidance by the RPA. For certain tumours this became time-consuming and difficult to the extent it defied realistic expectations of RPA use. For example, wide-spread punctate necrosis, tumour infiltrating immune cells, and acellular debris were virtually impossible to avoid annotating incorrectly in areas for some tumours, (Fig 5.2.4, Fig 5.2.5 B1-B2). The RPA itself is not perfect; it would occasionally not include bland looking tumour cells or include stromal cells in scoring, but the RPA algorithm could reliably ignore anthracotic pigment, and generally managed well with differing tumour cell types (e.g. acinar adenocarcinoma, keratinising squamous cell carcinoma etc.) and was generally consistent and accurate when applied to carefully drawn annotations (Fig 5.2.5 A1-A2, Fig 5.2.6). There were 15 occasions when the RPA assisted score and the manual score classified the cases into different clinical categories. All 15 cases were re-reviewed, and in 14 cases (12% of the LLP cohort) the RPA assisted score was felt to be more accurate than the original manual score. The unadjusted RPA for these 15 cases was typically very poor. The major difference was felt to be related to the algorithm helping to more accurately quantify patchy and multi-focal staining across large sections of tissue in cases close to both the ≥1% and ≥50% TPS cut-off.

Learning curve data was not formally assessed, but it was agreed by all pathologists that there was a period of learning the basic strengths and limitations of the RPA before the user felt confident in its application. Time to report was also not formally captured, but it was felt that it took considerably longer to report a case during this learning period than manual scoring would take. Furthermore, the learning period was also required for the users to identify which cases were not suitable for use with the RPA. For example, a cytology specimen with relatively intact morphology may seem like a suitable candidate for use, but if a significant degree of single tumour cells admixing with

macrophages has occurred, it may render the specimen, at least in part, less accurate for assessment with the RPA. It was also felt that the initial learning phase combined with an introductory lesson or tutorial on the use of the RPA would minimise this time period. However, once the RPA had been used on a number of cases (agreed to be between 10-20 cases, but not formally assessed) time reporting was felt to be likely only moderately longer than conventional reporting at worst and equivalent at best.

Once the learning period had elapsed, and the appropriate use of the RPA established, it was felt that the RPA provided an increased measure of 'confidence' to the user: when scoring the specimen for a TPS, if the algorithm assisted score was similar to the value identified as the probable TPS by the pathologist, it conveyed an increased sense of confidence that this is the 'true' TPS value. Conversely, if the RPA assisted score was significantly different to the TPS identified as a likely value, it triggered a more thorough re-evaluation of the case. Outcomes of re-evaluation included a dismissal of the RPA assisted score as a result of algorithm error, specimen unsuitability, inaccurately drawn annotations and so forth, or an agreement to change the TPS as a result of the RPA.

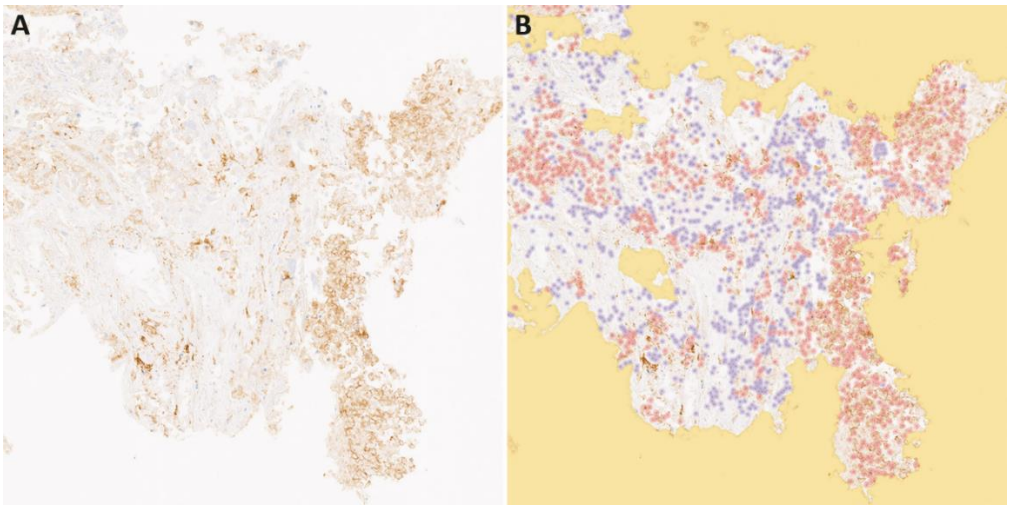


Fig 5.2.4 False scoring of PD-L1 (SP263) by the RPA. Acellular debris (**A**) can cause false positives (red) and false negatives (blue) to be scored by the RPA, as seen in **B**. Care needs to be taken when drawing annotations to avoid excess inclusion of these areas.

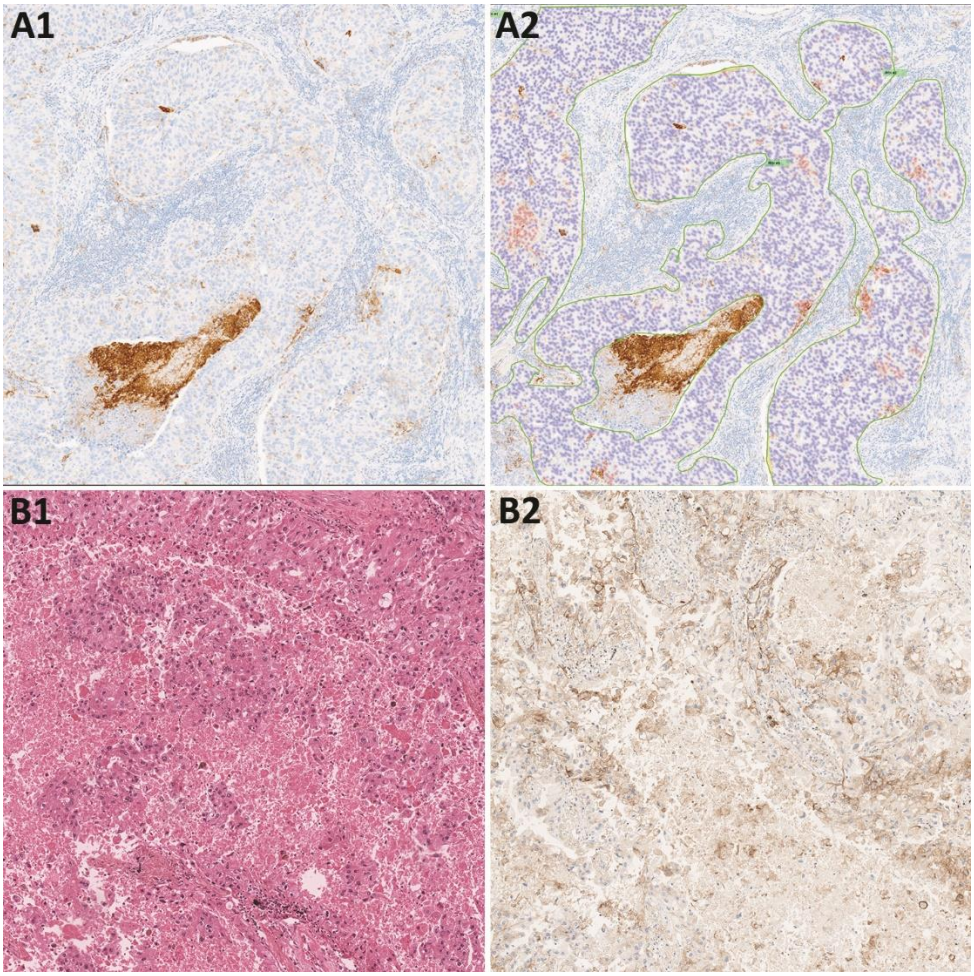


Fig 5.2.5 Necrosis and PD-L1 (SP263) positive TILs in NSCLC sections. **A1-A2 –** Area of well circumscribed necrosis and TILs (**A1**) can be annotated around accurately (**A2**). **B1-B2 –** Area of tumour admixed with necrosis and apoptotic tumour cells (**B1**) which is nearly impossible to annotate accurately (**B2**).

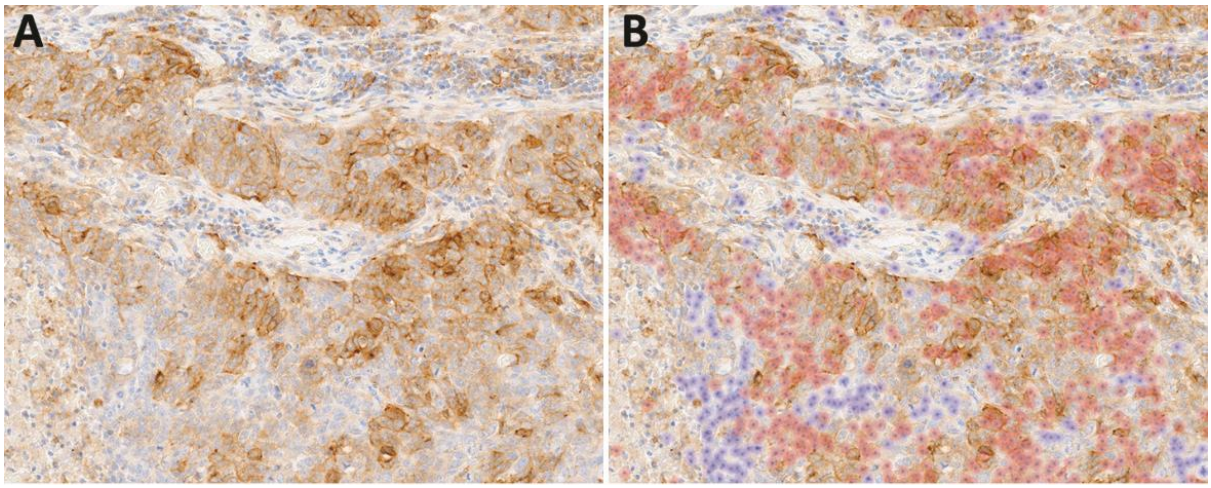


Fig 5.2.6 NSCLC tumour stained for PD-L1 (SP263) with TILs and necrosis (**A**) with RPA applied to the entire area (**B**). Despite inclusion of TILs and necrosis, PD-L1 positive areas (red) and PD-L1 negative areas (blue) are generally accurately confined to tumour cells.

5.2.4 Intra-observer concordance for scoring PD-L1

Intra-observer concordance for PD-L1 for three separate pathologists was scored using the RPA and manual approaches across 107 cases. There was variation in ICC (0.832-0.956) and Cohen's Kappa (0.455-0.719) between the pathologists. Intra-pathologist concordance was improved for scoring PD-L1 if using the RPA both times (ICC 0.993, Kapp 0.98). Results shown in Table 5.2.5.

	Correl	Sig	ICC	Sig	Kappa	Sig
Path 2 Man vs RPA	0.889	Y (p<0.0001)	0.832	Y (p<0.0001)	0.719	Y (p<0.0001)
Path 3 Man vs RPA	0.881	Y (p=0.002)	0.905	Y (p<0.0001)	0.455	Y (p=0.029)
Path 1 Man vs RPA	0.925	Y (p<0.0001)	0.872	Y (p<0.0001)	0.675	Y (p<0.0001)
Path 1 Man vs Man	0.98	Y (p<0.0001)	0.956	Y (p<0.0001)	0.842	Y (p<0.0001)
Path 1 RPA vs RPA	0.985	Y (p<0.0001)	0.993	Y (p<0.0001)	0.98	Y (p<0.0001)

Table 5.2.5 Intra-pathologist correlation and scoring agreement of 107 NSCLC cases for PD-L1 (SP263) by manual scoring on uPath (Man) or by RPA assisted (RPA) for a TPS.
Correl., Correlation; ICC, intraclass correlation co-efficient; Sig, significance. RPA, Roche-PD-L1-algorithm

Variation between pathologists for their respective unadjusted RPA scores compared to their agreed RPA adjusted scores are shown in Table 5.2.6.

RPA assisted vs unadjusted RPA	Correl	Sig	ICC	Sig	Kappa	Sig
Path 1	0.93	Y (p<0.0001)	0.892	Y (p<0.0001)	0.88	Y (p<0.0001)
Path 2	0.833	Y (p<0.0001)	0.799	Y (p<0.0001)	0.689	Y (p<0.0001)
Path 3	0.908	Y (p<0.0001)	0.919	Y (p<0.0001)	0.818	Y (p<0.0001)

Table 5.2.6 Intra-pathologist correlation and scoring agreement of 107 NSCLC cases for PD-L1 (SP263) comparing unadjusted RPA TPSs to RPA assisted TPSs for each pathologist.

5.2.5 Inter-observer concordance for scoring PD-L1

Inter-observer concordance for PD-L1 between three pathologists was compared for both manual scores and RPA assisted scores for 50 matched cases. Overall inter-observer agreement was good for both scoring techniques, with an improvement in both ICC (0.962 vs 0.986) and Cohen's Kappa (0.808 vs 0.878) when using the RPA to assist scoring. Results shown in Table 5.2.7

	Correl	Sig	ICC	Sig	Kappa	Sig
Manual	0.945	Y (p<0.0001)	0.962	Y (p<0.0001)	0.808	Y (p<0.0001)
RPA assisted	0.962	Y (p<0.0001)	0.986	Y (p<0.0001)	0.878	Y (p<0.0001)

Table 5.2.7 Inter-pathologist correlation and scoring agreement of 50 NSCLC cases for PD-L1 (SP263) comparing manual scoring on uPath (Manual) to RPA assisted for a TPS

Each pathologist drew annotations independently and to differing numbers and sizes (Mean annotations for each pathologist: (5 (range 1-10), 10 (range 1-26), 23 (range 1-40)). Example shown in Fig 5.2.7.

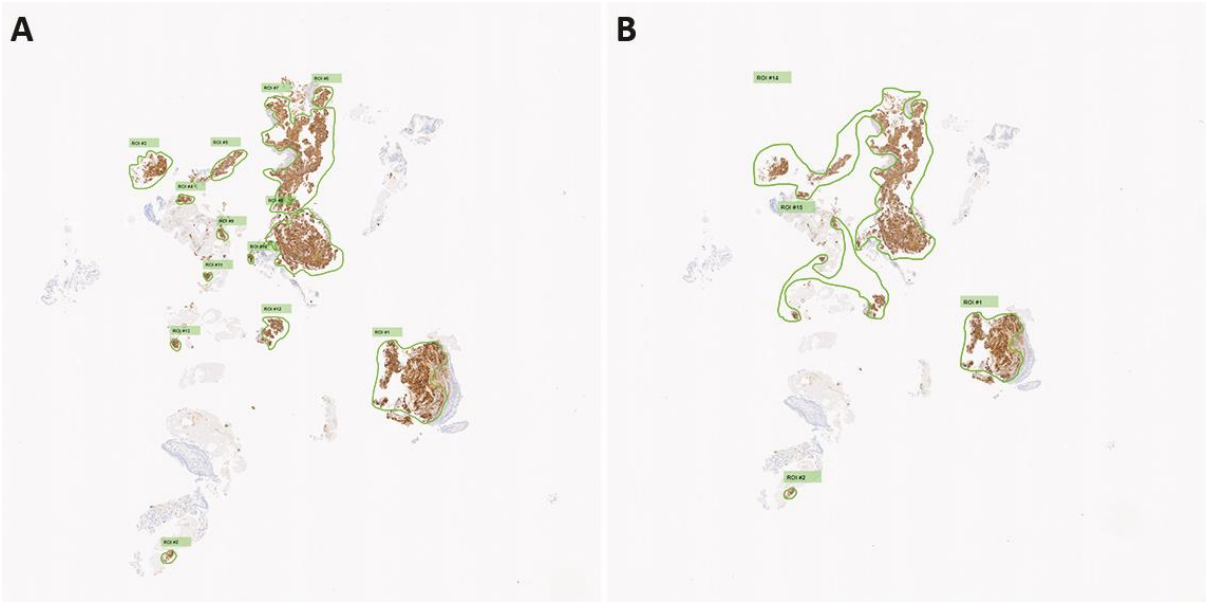


Fig 5.2.7 Case assessed for PD-L1 (SP263) using uPath to draw annotations for RPA assistance. **A** – 13 annotations by pathologist C **B** – 4 annotations drawn by pathologist B.

5.2.6 Inter-observer concordance for positive pixel count algorithm

Four users applied the positive pixel count algorithm on 30 TMA cores from the Eli Lilly cohort to assess for inter-user variability in simple specimens. Concordance between the scorers was nearly perfect. Results shown in Table 5.2.8.

	Correlation	Sig	ICC	Sig
Scorer 1 vs 2	0.991	Y (p<0.0001)	1	Y (p<0.0001)
Scorer 1 vs 3	1	Y (p<0.0001)	1	Y (p<0.0001)
Scorer 1 vs 4	1	Y (p<0.0001)	1	Y (p<0.0001)
Scorer 2 vs 3	0.993	Y (p<0.0001)	1	Y (p<0.0001)
Scorer 2 vs 4	0.991	Y (p<0.0001)	1	Y (p<0.0001)
Scorer 3 vs 4	1	Y (p<0.0001)	1	Y (p<0.0001)
All			1	Y (p<0.0001)

Table 5.2.8 Inter-observer correlation and scoring agreement of 30 TMA cores stained for PD-L1 (SP263) when using the Aperio Positive Pixel count algorithm.

5.2.7 Whole section scoring with the RPA

122 cases were scored using a manual TPS, an RPA assisted TPS or with the RPA applied to the whole section (WS) with no adjustment by a pathologist. Agreement for WS TPSs was generally poor compared to either manual or agreed TPSs. Results summarised in Table 5.2.9 and Fig 5.2.8.

	Correl	Sig	ICC	Sig	Kappa	Sig
RPA vs Man	0.965	Y (p<0.0001)	0.954	Y (p<0.0001)	0.818	Y (p<0.0001)
RPA vs WS	0.797	Y (p<0.0001)	0.762	Y (p<0.0001)	0.446	Y (p<0.0001)
WS vs Man	0.777	Y (p<0.0001)	0.698	Y (p<0.0001)	0.353	Y (p<0.0001)

Table 5.2.9 Correlation and scoring agreement of 122 NSCLC cases for PD-L1 (SP263) comparing manual scoring on uPath (Man) to RPA assisted (RPA) and whole-section (WS) TPSs.

The predictive value of WS and agreed (RPA adjusted) compared to original manual scores are shown in Table 5.2.10.

	NPV	PPV	OPA
WS 1%	0.75	0.91	0.76
WS 50%	0.72	0.96	0.78
WS Both			0.54
RPA assist 1%	0.94	0.94	0.94
RPA assist 50%	0.9	1	0.93
RPA assist Both			0.88

Table 5.2.10 Predictive power of using the RPA applied to either the whole section (WS) or to annotations reviewed by a pathologist (RPA Assist) to score PD-L1 relative to original manual assessment.

Specimen Type	All	Path 1	Path 2	Path 3
Biopsy	72	12	13	47
Cytology	40	9	10	21
Resection	10	4	2	4

Table 5.2.11 Specimen types in the RLUH cohort assessed for PD-L1 expression.

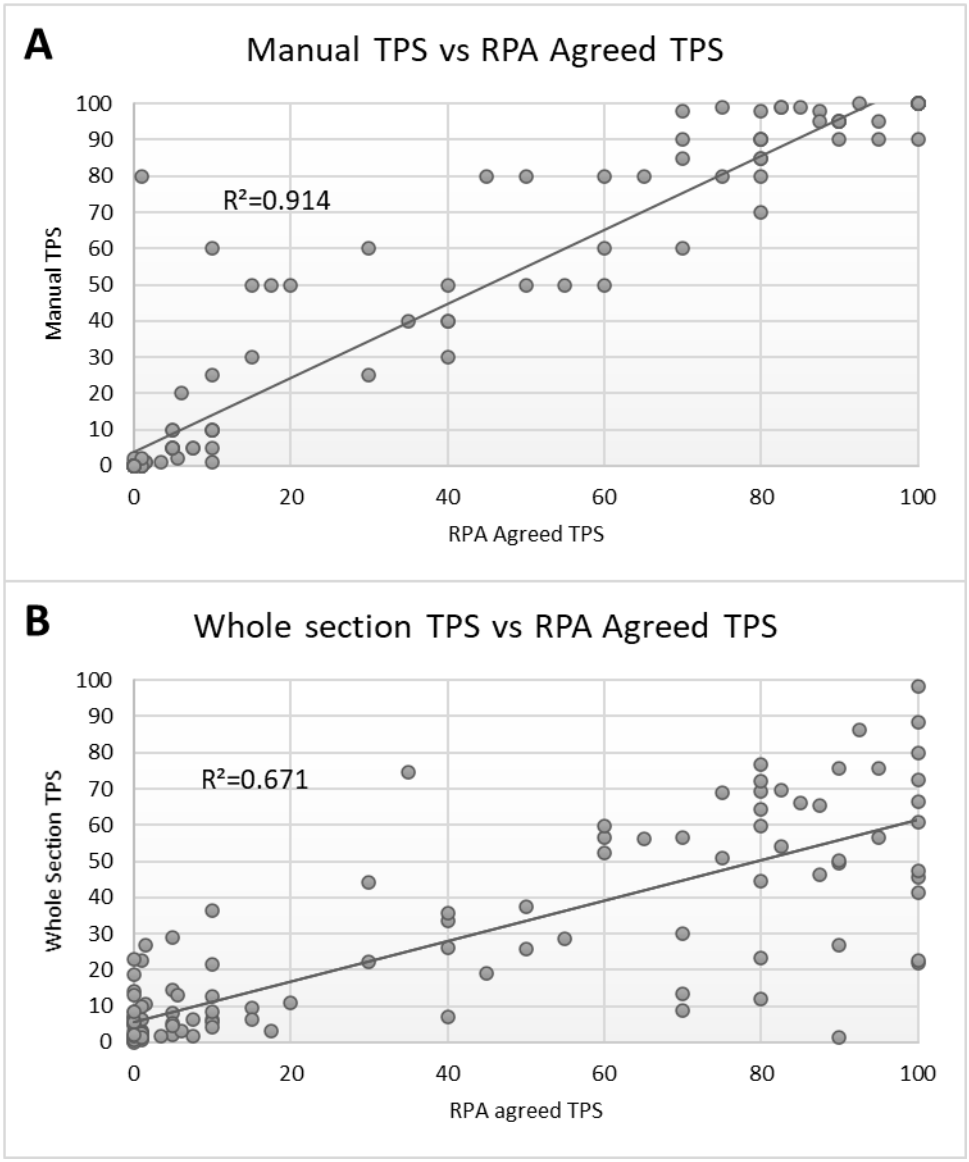


Fig 5.2.8 Linear regression for scoring 122 cases of NSCLC for PD-L1 comparing TPSs from **A)** manual scoring, the RPA applied to the whole section (WS) or **B)** the RPA applied to annotations reviewed by a pathologist (RPA Assisted).

5.2.8 Qualitative feedback from the RLUH cohort.

The RLUH cohort involved biopsies, cytology specimens and resections, reflecting a typical array of specimens used in routine diagnosis of NSCLC cases (Table 5.2.11). Certain specimens were not suitable for drawing accurate annotations; for example, cytology specimens such as pleural fluid aspirates which were composed of single tumour cells or only small groups of tumour cells, admixed with mesothelial cells, macrophages, background debris and so forth were virtually impossible to annotate in a timely or accurate fashion. (Fig 5.2.9) Other specimens did not require the assistance

of the RPA for an accurate interpretation; for example specimens with no staining for PD-L1 in any of the test material, or for small biopsies with homogenous high PD-L1 positivity. (Fig 5.2.10) Other specimens presented a challenge in drawing annotations for the RPA depending on the type of PD-L1 staining and the nature of the morphology and other features within the tissue. For example, as with the LLP cohort, very large pieces of tissue could be heavily infiltrated by TILs and macrophages expressing PD-L1, or with large amounts of punctate necrosis (Fig 5.2.11). However, the application of the RPA to certain cases proved invaluable; when there were multiple fragments of tissue that could not be visualised on screen at the same time, when there was significant heterogeneity of PD-L1 expression and/or when there were large quantities of tissue without the previously mentioned confounding features (Fig 5.2.12). Despite the large list of specimens and indications that warrant caution, it was generally felt the RPA added benefit if applied correctly, and that the majority of specimens would be suitable for this use.

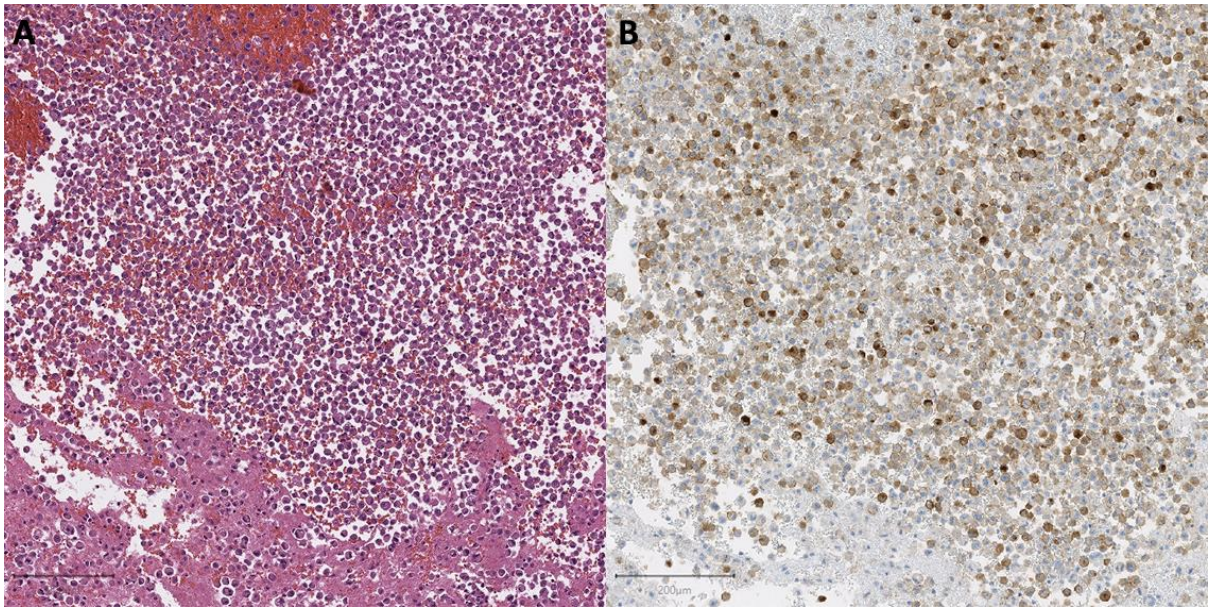


Fig 5.2.9 Samples not suitable for RPA **A**+**B** – Pleural fluid aspirates with tumour cells admixed with mesothelial cells, macrophages, lymphocytes and cellular debris stained for H&E (**A**) and PD-L1 (SP263) (**B**) Drawing annotations for RPA use is unlikely to yield helpful data in these samples.

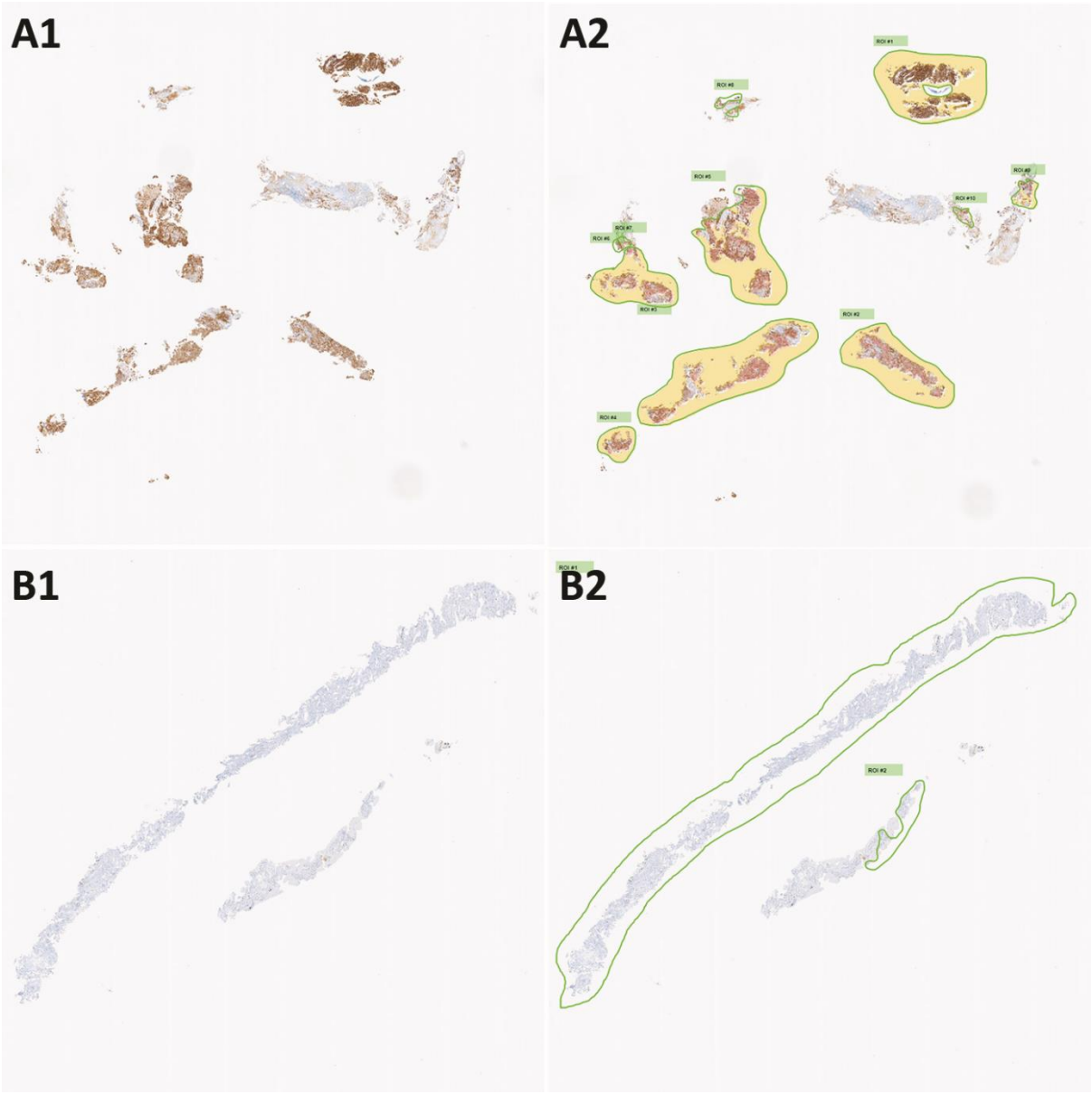


Fig 5.2.10 Samples not requiring RPA **A1/A2** – A biopsy of NSCLC with homogenous strong staining for PD-L1 (SP263) scored manually as 95% TPS (**A**). **B** - Application of RPA returns a TPS of 92.5%. **B1/B2** – A core biopsy of NSCLC pan-negative for PD-L1 scored manually as 0% TPS (**A**). **B** – Application of RPA returns TPS of 0.1%. In both instances, RPA can be applied and returns an accurate result, but the added benefit is minimal.

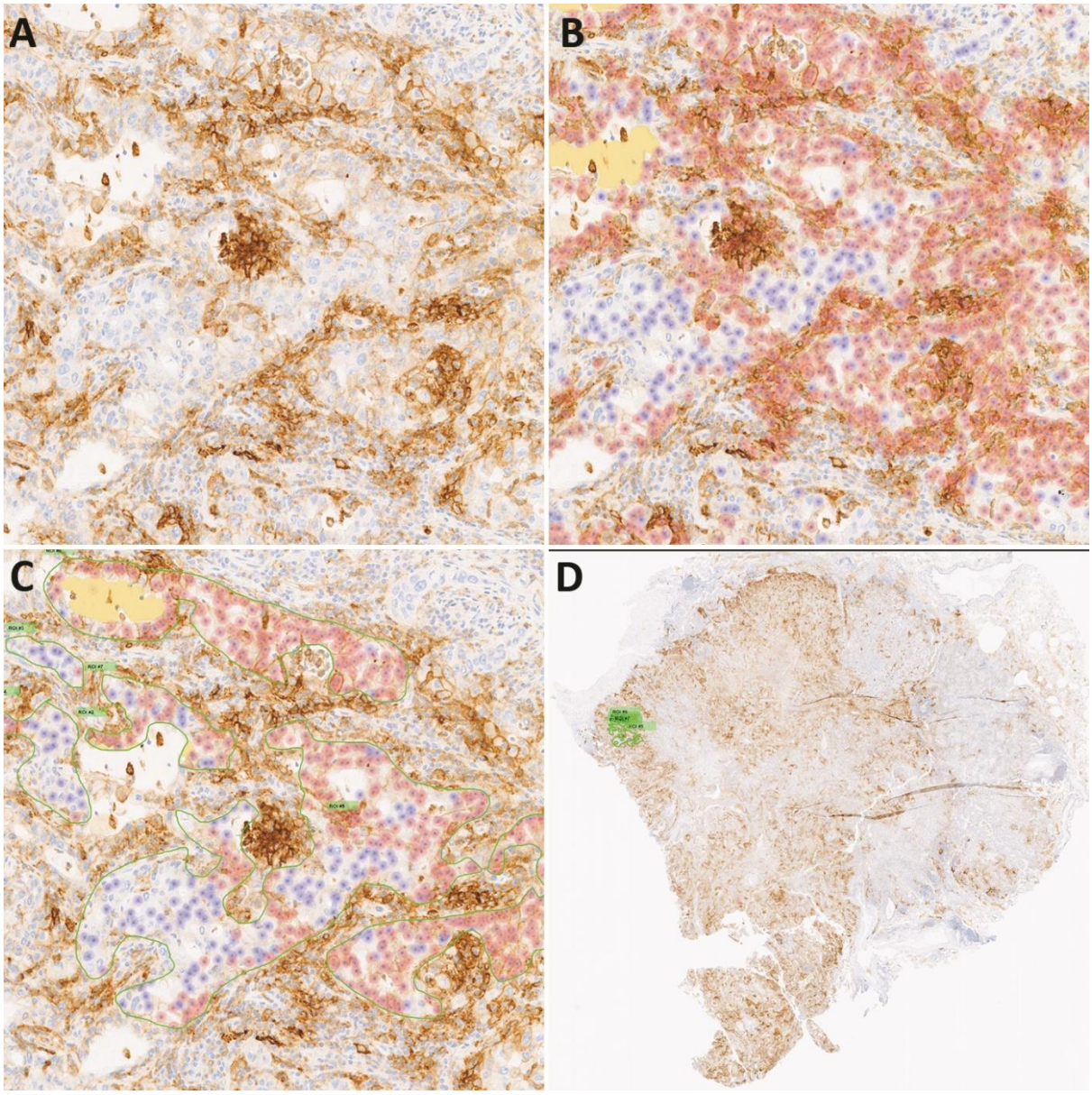


Fig 5.2.11 Example of when accurate application of RPA is excessively time consuming in NSCLC stained for PD-L1 (SP263). **A** – Tumour with positive TILs and areas of necrotic cells. **B** – Application of RPA to entire area includes positive TILs and necrotic cells. **C** – Careful annotation results in a more accurate RPA but only a very small area of the tumour is included (**D**). To annotate the entire tumour accurately would be very time consuming and would take far longer than an accurate manual review.

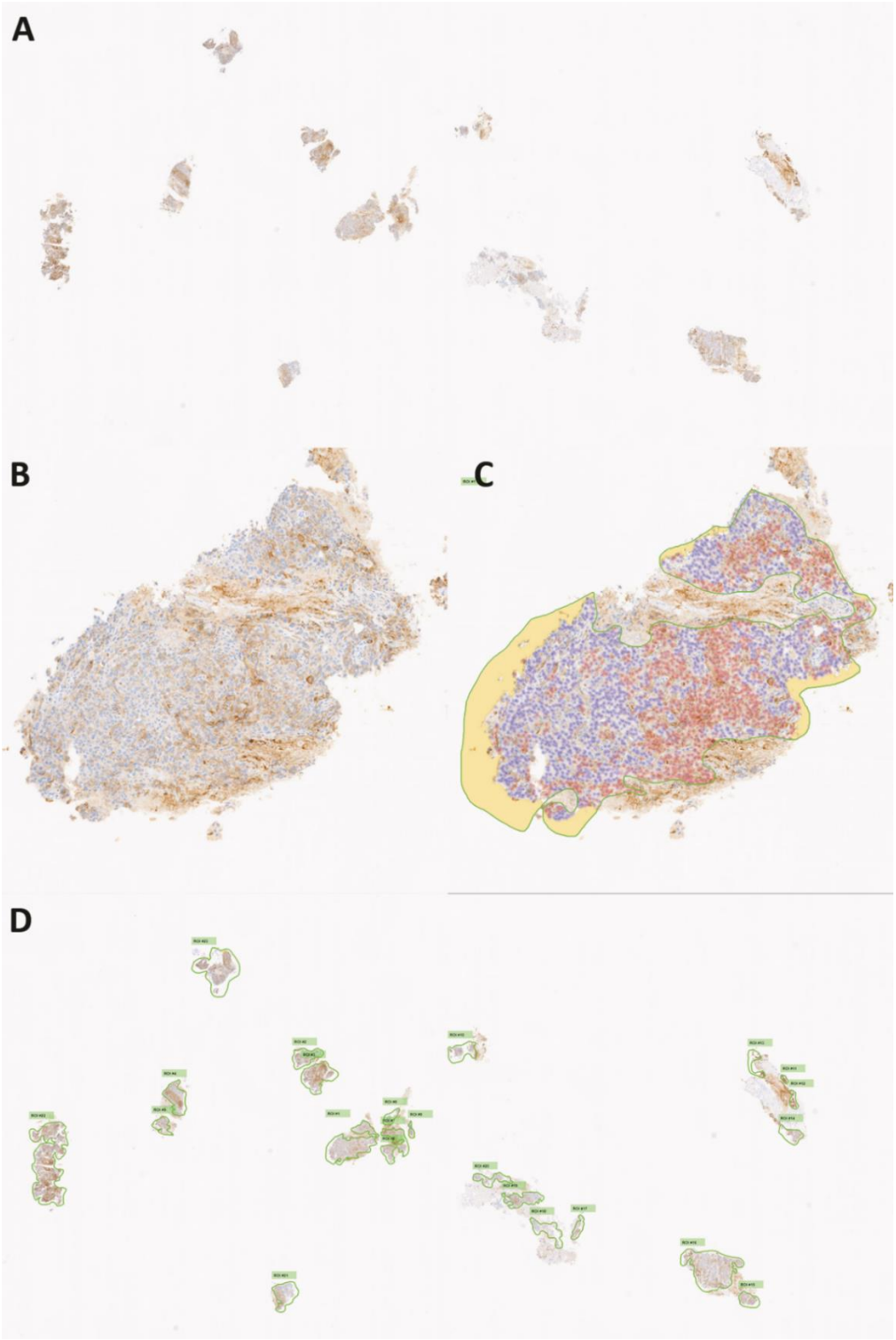


Fig 5.2.12 Example of when application of RPA adds benefit. **A** – Biopsy of NSCLC with multiple fragments of tissue which cannot all be viewed at high power simultaneously. Manually scored as 50% TPS for PD-L1 (SP263). **B** – Higher power view of a fragment showing tumour with heterogenous expression of PD-L1. **C** – RPA applied to annotated regions shows scoring of heterogenous area. **D –** RPA applied to multiple annotations returns a total of 19,965 tumour cells scored, with 10,615 PD-L1 positive and 9,350 PD-L1 negative tumour cells to give a RPA TPS of 53.2%, confirming this case is ≥50% TPS.

5.2.9 Training webinar to introduce new users to the RPA

A copy of a training webinar produced as part of this work is attached to the thesis. In summary it documents the basics of how to open and view images in uPath with a brief overview of the main functions and advantages of digital pathology (Fig 5.2.13). It then focuses on the application of the RPA scoring approach, and demonstrates a number of potential pitfalls and benefits outlined in Table 5.2.12, as well as examples of worked cases. This is also available online at https://www.brighttalk.com/webcast/12161/431368.

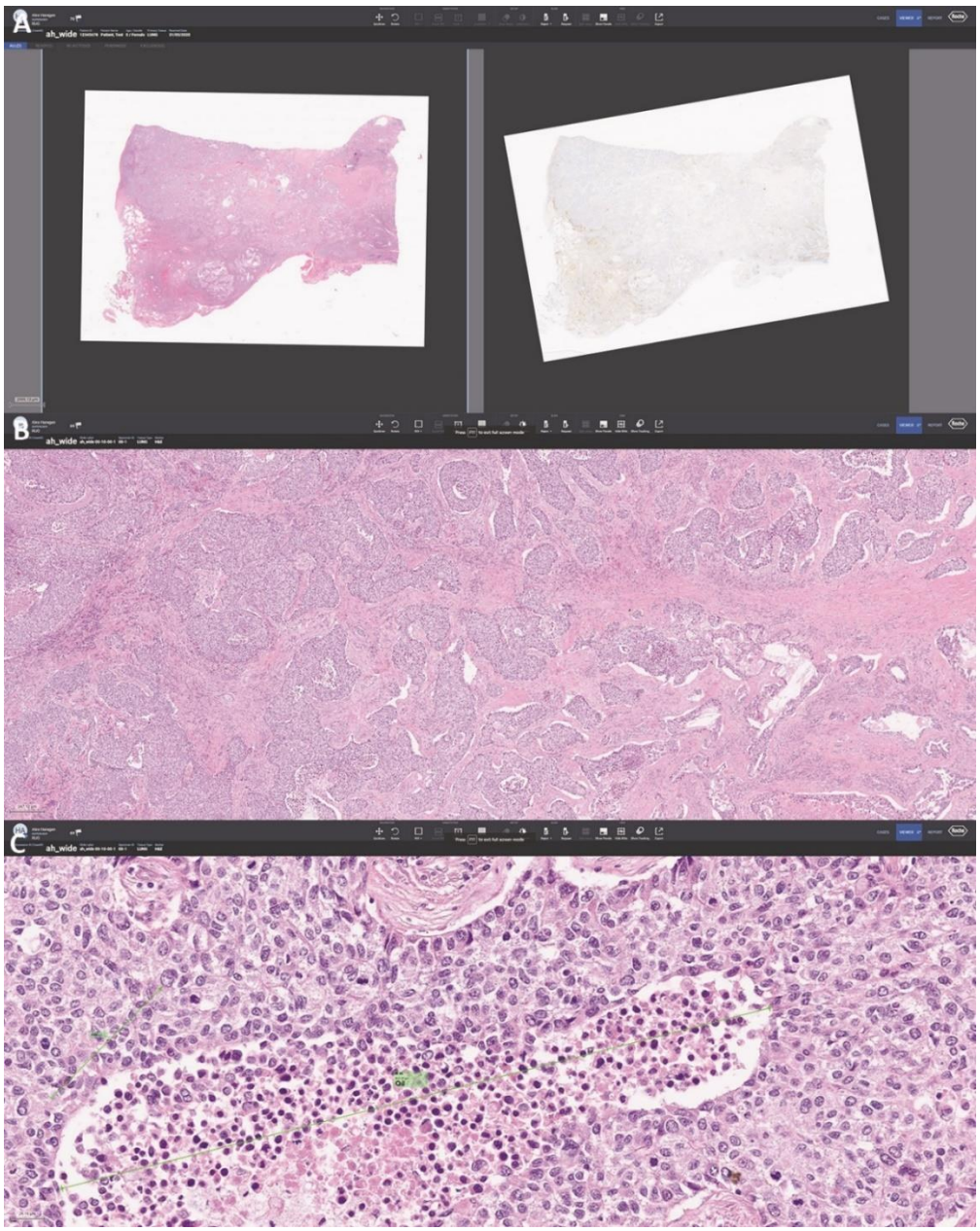


Fig 5.2.13 Benefits of digital pathology. **A** – Parallel viewing of H&E and PD-L1 IHC. **B** – Increased field of vision allows for considerably more tumour to be viewed at once than conventional microscopy. **C –** Annotations, arrows and measurements are all easily applied to the image and can be shared with other viewers. (Green arrows/lines)

Benefits	Pitfalls	Consider not using if:	General Advice
May increase intra-observer concordance	Admixed cells (e.g. cytology)	Pan-negative cases	RPAA is only as good as the annotations drawn
May increase inter-observer concordance	Punctate necrosis	Very high TPS cases	
Functions like a '2nd Opinion'	Large resection cases	Very small tissue samples	
Helpful when not all tissue can be visualised at once	**Hard cases/poor annotation may**	Annotations likely to be inaccurate	RPAA requires experienced pathologist to maximise benefit
Good quantification of heterogenous PD-L1 expression	**A) Falsely lower TPS by:**	Poorly differentiated tumour	Training and learning period to familiarise oneself with RPAA is advised
With practice is quick and easy to apply	Including stroma	Tumour is densely infiltrated by lymphocytes	
Does not need to be applied to all cases	Excluding positive tumour cells	PD-L1 interpretation is particularly tricky	
Pathologist judgement can be guided by RPAA	**B) Falsely raise TPS by including:**		
	Necrosis		
	Apoptotic/crushed tumour cells		
	Macrophages		
	Tumour infiltrating lymphocytes		

Table 5.2.12 – Advice to a new user of uPath wanting to apply the RPA highlighting some of the strengths and pitfalls of it use.

5.3 Discussion

5.3.0 Variation of PD-L1 IHC clones

In NSCLC, a TPS will determine eligibility for PD-1/PD-L1 IM treatment or help to guide treatment decisions, but the use of differing PD-L1 IHC clones, TPS cut-offs and IM agents between clinical trials has resulted in four different PD-L1 IHC antibodies being clinically validated to assess PD-L1 expression in a wide variety of indications. Attempts to rationalise this complex situation has involved numerous studies that assess the equivalency of different PD-L1 IHC clones, on the basis that clones that are concordant could be used in place of each other. In keeping with previous papers, this study has shown that SP142 is considerably less sensitive than other clones for staining PD-L1 positive tumour cells,[288, 297, 304, 330, 335, 336, 342, 345, 350-352, 501] and that 22C3, 28-8 and SP263 are similar[304, 317, 330, 334, 335, 337, 343, 344, 347, 499, 501]. There are appreciable differences in the intensity and sharpness of staining, with SP263 generally producing a cleaner and brighter stain (Fig 5.2.0), but overall TPSs were essentially similar in distribution between 22C3 and SP263 in this study. These were not matched samples, however, and though over 1,400 cases were included, the differences between 22C3 and SP263 is sometimes only appreciable on matched cases very close to critical clinical cut-offs of ≥1% or ≥50%, and indeed even in papers where concordance of these two clones are very high, they fall beneath a 90% concordance rate that is generally considered a minimal level of agreement for *in vitro* assays to be considered equivalent.[304, 498, 528]

The underlying cause for differences between PD-L1 IHC clones maybe a result of inherent properties of the clones or as a result of analytical variations between the assays. A recent study has shown that SP263 and SP142 share the same epitope region, which is cytoplasmic and distinct to the extracellular binding regions of the 22C3 and 28-8 clones, which are also distinct from each other. E1L3N, a research only clone, has a cytoplasmic epitope region that overlaps considerably with the SP263/SP142 site, but is still distinct. Considering the shared region of SP263 and SP142, but the consistent observation that SP142 is discrepant for IHC staining, and the similarity of staining for SP263, 28-8 and 22C3, that have no epitope region in common, the authors conclude that differences between clones is more likely a result of assays, platforms, or indeed tumour heterogeneity than variation in antibody binding sites.[327] It should be noted that affinity for a binding site(s) is typically considered to be more important than the binding site itself,[529] so there may simply be limitations inherent to each clone that the assays cannot overcome, although with sufficient alteration of the SP142 clone assay, staining equivalent to other clones can be achieved.[336, 337, 530]

In addition to their similar staining and similar TPSs, 28-8, SP263 and 22C3 also share similar predictive power for response to PD-1/PD-L1 IMs, whereas SP142 is again an outlier for this purpose.[501] The difference in PD-L1 IHC staining between clones therefore appears to be driven more by analytical variations than inherent properties of the clones themselves. Although the difference in staining between clones carries important clinical ramifications, where even subtle changes can classify patients differently, the underlying biology is perhaps better represented as spectrum of PD-L1 expression. For example, two tumours with a TPS of 0.5% and 1% would potentially classify patients into different clinical groups, but the TPS values suggest these two tumours are very similar from a PD-L1 expression perspective. Conversely, two tumours scoring a TPS of 1% and 49% would potentially classify the patients into the same clinical group, but clearly have a much larger variation in terms of PD-L1 expression between them. As the majority of the work performed in this thesis is with the SP263 clone, but other studies and clinical laboratories may use other validated clones, it is important to consider how relevant findings within this thesis maybe to other cohorts of patients and/or other PD-L1 IHC clones. It is reassuring that there is generally a minimal difference between the clones in this study as a result of interpretation, and the consistent observation that SP142 stains fewer tumour cells is in keeping with other studies. Furthermore, as variation between IHC stains seems to be a result of analytical and technical factors, rather than inherent properties of the clones themselves, it seems reasonable to suggest that findings from this thesis could be applicable to cohorts using clones other than SP263, particularly when using clones with similar staining patterns.

5.3.1 Digital pathology and image analysis for scoring PD-L1

The use of digital pathology has already been shown to be equivalent to conventional microscopy for routine clinical workloads and PD-L1 IHC interpretation specifically.[304, 504, 505] This study has shown a high concordance between matched cases scored using either conventional microscopy or scanned images (Fig 5.2.3) and has inherent advantages for routine work that includes larger fields of view easier ability to change magnification and ease of highlighting specific features on a slide. (Fig 5.2.13). More ambitious uses of digital pathology include automation, machine learning and AI approaches, all of which are at various stages of progression.[512, 514, 515] Several approaches to semi-automated or fully-automated interpretation of PD-L1 have been attempted, with generally encouraging results seen in malignant melanoma,[531] breast cancer[532] and NSCLC.[533, 534] Concordance between fully automated approaches and manual scoring are good, but not yet good enough to be applied without pathologist input, with correlation of scores generally around 0.7-0.9.[532-534] In this study the application of the RPA to the entire section without any annotations, as surrogate measure of a fully-automated approach, resulted in similarly good but imperfect concordance of results.

The application of semi-automated approaches requires an experienced pathologist to draw appropriate annotations around ROIs, and to have the background knowledge of the inherent pitfalls of PD-L1 staining to know when an image analysis solution is providing incorrect or inaccurate results. Semi-automated approaches are not a panacea for complex interpretation of IHC stains; the algorithm is only as good as the annotations drawn, and whilst unadjusted RPA TPSs varied considerably between pathologists, the final agreed scores were generally in high levels of concordance. The high levels of inter-pathologist concordance in this study is likely a result of the generally experienced pathologists involved with PD-L1 scoring, and it is possible that a less experienced pathologist may not achieve these levels. However, both intra-pathologist and inter-pathologist concordances were improved using the RPA to assist scoring, and previous studies have shown that improved inter-pathologist concordance is achievable using image analysis techniques.[531] The levels of concordance seen in this study when using image analysis are as high or higher than many other studies comparing PD-L1 expression interpretation.[304, 330, 342, 347, 379, 535] Using image analysis to augment and assist pathologists in a semi-automated fashion appears at present to be the best way to ensure accuracy equivalent to a pathologist, as they ultimately decide the final TPS to be reported.

In future developments, it would be hoped the most arduous step of this process, the drawing of ROI, can be minimised through the use of image-analysis assisted classification of tumour cells. Regrettably the ability of AI to identify morphological distinctions on H&E is limited and the subtle differences between a dysplastic cell and a tumour cell is likely to require many years of development before it can be routinely relied upon. [536, 537]

5.3.2 Training to use digital image analysis

A particularly interesting aspect of this study was the opportunity to train experienced pathologists in a novel technique. Specific training to introduce pathologists to assessing PD-L1 expression has been in place for several years, and previous studies have shown there is a period of learning required for using digital pathology in general,[266, 323, 504] but there is limited data on the combination of these techniques. Correct application of the RPA can result in a helpful, fast and reliable aide in scoring PD-L1 that can reduce discrepancies between scorers and provide confidence in the quantification aspect of its interpretation. Conversely poor application may be time consuming and lead to inaccurate and confusing results. Appropriate training for use of the RPA is therefore an important process, and as with conventional PD-L1 interpretation a teaching session to inform new users of the limitations and pitfalls of interpretation is a useful place to begin. Likewise, ongoing experience is necessary to maintain confidence and competence. The webinar written as part of this

project is designed to outline the strengths and weakness of the RPA, but it is up to the individual to decide when and if they wish to use it. It is the experience of this project that, despite the long list of potential pitfalls to be aware of when using the RPA, the majority of NSCLC specimens requiring PD-L1 expression would be suitable to apply the RPA and to benefit from its use, but requires experience to be able to rapidly determine the appropriate specimens for this.

This study is limited by the small number of pathologists and modest sample sizes used, as well as having no predictive data for response to PD-1/PD-L1 IMs in these cohorts. However, it has provided pilot data to ascertain scenarios where the benefit of digital pathology and image analysis assistance via the RPA should or should not be used. IHC for assessing PD-L1 expression remains an important predictive biomarker, and whilst imperfect, is quick, effective and relatively easy to perform. Concordance of interpretation of PD-L1 IHC between pathologists can be improved with the use of digital pathology and semi-automated algorithms, and with care and experience can be used to augment the pathologists' interpretative power.

Whilst there were no predefined success criteria for this study, this initial project has provided the opportunity to develop a learning programme and tutorial to understand the RPA, and future work will focus on the use of this with a larger number of pathologists whom have a broad range of experience in scoring PD-L1 in NSCLC. The current regulatory status of the algorithm is currently research only, but it is being developed with an aim to be used as a clinical assisted tool in the UK and Europe (as SP263 is not at present validated for use in NSCLC in the USA).

One critically important aspect of digital pathology is that image analysis is only as good as the quality of the image it is being performed on; which means the questions pertaining to pre-analytics, analytics and interpretation are all just as relevant to digital pathology as to conventional pathology. PD-L1 expression interpretation is also hampered by heterogeneous expression. In certain cases this study has demonstrated the benefits of using digital pathology to help interpret heterogeneous staining, but if the algorithm is only as good as the annotations, and the annotations only as good as the image, and the image only as good as the specimen, it requires the specimen itself to be representative of the tumour, a challenge the next chapter will explore in depth.

Chapter 6 – Heterogeneity of PD-L1 expression

6.0 Introduction

In the previous two chapters, the issues surrounding pre-analytics and analytics led to findings that could help ensure that sampled tissue from NSCLCs provide an accurate and consistent representation of PD-L1 expression. In this chapter, the issues surrounding aspects of tumoural heterogeneity will be considered. This is an important next step in optimising the use of PD-L1 IHC; the impact of pre-analytics and analytics are only beneficial as far as the sample available is concerned. For a specimen to be truly useful, it must represent the tumour burden of a patient as a whole, and insufficient or inappropriate sampling will render all efforts to optimise the pre-analytics and analytics meaningless. This chapter will therefore consider the scale of PD-L1 heterogeneity specifically, in a bid to understand how representative specimens are in regards to PD-L1 expression and other relevant biomarkers.

6.0.0 Tumour Heterogeneity

Heterogeneity of malignant tumours is a well-recognised phenomena that is being increasingly understood as having clinical and biological significance in a number of solid cancers, including lung, colorectal, breast and renal cancers, which can be measured by a variety of metrics including pathological and radiological methods.[538-541]. Lung cancers, especially ADCs, are particularly heterogeneous for both morphology and immune markers.[378, 380, 386, 542, 543] Molecular heterogeneity also exists, with studies such as TracerX and other multi-region sequencing approaches providing in-depth information pertaining to how the genomics of tumour cells can change over space and time, and how various metastatic deposits can be mapped out to define specific clones and sub clones and their genetic history.[544-547] Variation in clinically relevant mutation expression rates between and within tissues, including *EGFR*, *ALK* and *KRAS,* as a result of tumour heterogeneity has been previously demonstrated, and this heterogeneity could potentially misclassify patients and render them ineligible for a beneficial treatment option.[548-552]

The clinical impact of tumour heterogeneity varies between each target of interest, with appropriate measures usually put in place to attempt to account for potential issues. For example, although testing for *EGFR* is generally robust enough to be applied to a variety of tissue types, tumoural heterogeneity is a significant concern in that it could lead to false negative results.[553-555] Approaches to help combat the effect of *EGFR* heterogeneity include the use of more sensitive methods, re-biopsy of patients, and the use of 'liquid biopsies'.[551, 556, 557] The use of IHC as a diagnostic aid is perhaps less affected by tumoural heterogeneity: it can be consistently applied regardless of the

anatomical location of the tumour site; an ADC of the lung may express TTF-1 if sampled from lung parenchyma or from a metastatic site, and indeed this consistency is why IHC stains can be helpful in diagnosing poorly differentiated metastatic cases. However, it is well recognised that no IHC stain is 100% specific or sensitive, so even morphologically convincing tumours may not always express the range of proteins in a 'textbook' manner, often as a result of pre-analytical factors, the assay or indeed tumour cell heterogeneity.[550, 558-561] To help combat this, IHC as a diagnostic aid is typically applied as part of a panel of targets which can help ratify diagnostic decisions even if one of the assays is equivocally staining. Though this approach has its limitations, IHC stains applied in this fashion can help decide the cell type or types present, even if heterogeneous expression is a factor. To consider what appropriate measures may be required to counter PD-L1 expression heterogeneity by IHC, it is necessary to likewise first determine the scope of the problem, and then attempt to suggest and apply solutions to combat it.

6.0.1 Heterogeneity of PD-L1 expression

The heterogeneous expression of PD-L1 is a significant challenge to ensure accurate and relevant prediction of response to IM therapy. As PD-L1 is not used as a diagnostic aid, but as a predictive tool, the underlying biology that leads to PD-L1 expression is perhaps a more important factor than for other IHC assays.

If PD-L1 expression is upregulated as part of an active immune-escape process by the tumour in response to immune activity targeted against it,[562-565] heterogeneity may represent distinct areas of relevant biological and immunological changes across a tumour and between tumour sites. However, a tumour expressing PD-L1 constitutively as a result of genetic changes that are independent of immune activity[189, 563-566] may demonstrate variable loss or gain of PD-L1 expression driven by background genomic variation, which is potentially less relevant to guiding therapy. The distinction of reactive or constitutive expression is therefore likely an important aspect in regards to PD-L1 heterogeneity, but is not ascertainable by PD-L1 IHC alone. Furthermore, there is no panel of predictive biomarkers for PD-L1 IMs: the determination of PD-L1 expression is typically made on a single assay alone.

As discussed previously, I have defined the four main types of PD-L1 protein expression heterogeneity as intra-tumoural (variation within a single tumour site), inter-tumoural (variation between tumour sites), temporal (variation of PD-L1 expression over time) and iatrogenic (PD-L1 expression changes as a result of treatments). PD-L1 expression is routinely performed on new diagnoses of NSCLC and in many centres this is performed as 'reflex' testing; meaning that PD-L1 status will be tested on the first available diagnostic tissue.[282, 567] As the initial diagnosis of NSCLC is

typically made on small tissue biopsies or cytology specimens, heterogeneity of PD-L1 expression represents a practical problem in that these small samples potentially do not fully represent the entire tumour mass they are sampled from, and therefore may under or overestimate total PD-L1 expression levels. Subsequent samples are often not acquired (for example patients may be too unwell to undergo another procedure or the tumour is not appropriate for surgical resection), resulting in the use of a limited quantity of tissue with which to test for PD-L1. As such, the use of a small tissue sample, at a single point in time to make decisions about treatment options is potentially affected by all types of heterogeneity. Though these have been considered in more depth in the literature review, a brief overview of these types of heterogeneity will be considered here.

6.0.2 Intra-tumoural and inter-tumoural heterogeneity

The biggest issue of intra-tumoural heterogeneity is sampling error – if PD-L1 expression in a tumour is as shown in Fig 6.0.1 one can easily imagine how a biopsy, which will sample a small area, may be taken from different regions, each with vastly differing PD-L1 TPSs (e.g. a biopsy around the yellow arrow may return a TPS of 0%, whereas a biopsy from around the red arrow might return a TPS of 100%), and each, relative to the whole section image, might be wrong. It is well established in clinical trial and real-world data that PD-L1 negative patients can still respond to PD-1/PD-L1 IM therapy, [47, 49, 50, 234] and it is worth considering that in at least some cases the tumours from these patients are not wholly negative for PD-L1; instead tumoural heterogeneity resulted in just the available tissue being negative for PD-L1. An important question to address, therefore, is what quantity of tumoural tissue is required to be fully representative of a tumour mass, especially as many previous studies have demonstrated that sampling by small biopsies can result in discrepancies of PD-L1 expression.[388, 390-392, 568]

Many diagnostic specimens in NSCLC are from regional lymph node metastases, or distant metastases, and the question of how representative they are of the primary lesion is also an important one.[299, 569] Indeed if molecular, morphological and diagnostic IHC stains can vary between anatomical sites, it seems reasonable to presume that a marker of immune activity will also vary between sites. The predictive power of PD-L1 IHC from Phase III clinical trials for PD-1/PD-L1 IMs was largely based on the use of primary tissue.[46-50] The use of metastatic NSCLC tissue in testing for PD-L1, however, is a clinical reality, but these are not always concordant with primary tumours,[355, 542, 570] therefore the impact of inter-tumoural heterogeneity is important for ascertaining the relevance of clinical trial data, and the impact of sample site selection on deciding treatment.

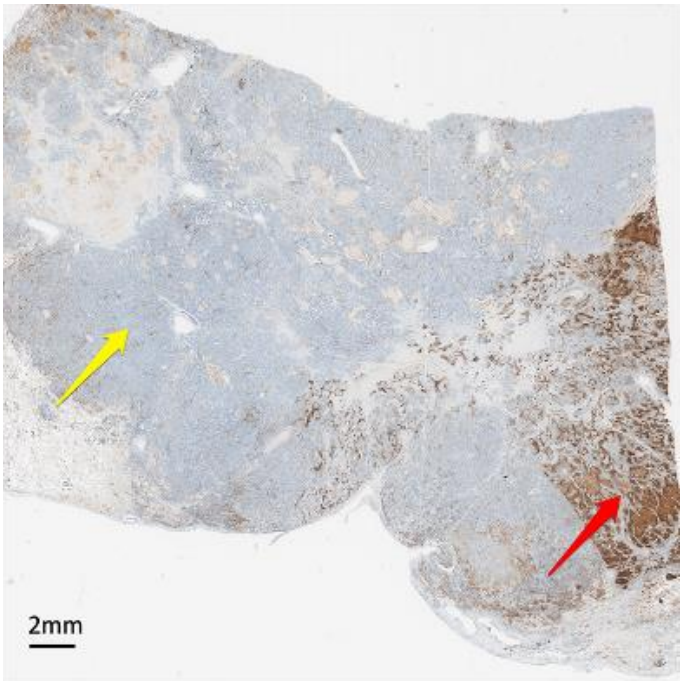


Fig 6.0.1 Intra-tumoural heterogeneity of PD-L1 (SP263) expression in a whole section of surgically resected NSCLC. Yellow arrow indicates area of PD-L1 negative expression, Red arrow indicates areas of high PD-L1 expression. Biopsies from each site may therefore dramatically under or over represent PD-L1 expression as a whole.

6.0.3 Temporal and iatrogenic heterogeneity

PD-L1 expression has been demonstrated to change over time, and most patients will receive some form of treatment over the course of their disease, with previous work varying on the impact this has on PD-L1 expression. Some studies have suggested minimal effect on the expression of PD-L1 by either time or treatment[399, 405] though a much more consistent observation is that time and treatment both effect PD-L1 expression, though the effects of these may increase,[423-426] decrease,[418, 421] or both increase and decrease PD-L1 expression.[349, 419, 420, 422] The metastatic process occurs over time and may occur during the course of treatment regimens, and thus inter-tumoural heterogeneity may be impacted by both temporal and iatrogenic changes.

Therefore, a major challenge is in separating the impact of these different heterogeneity types from each other. This study hypothesises that a well-defined cohort is the best approach in attempting to quantify the issue. The LLP cohort is ideal for these purposes as it is a cohort of NSCLC patients, all of whom underwent surgical resection with no neo-adjuvant therapy. In addition, every selected patient had N1 or N2 nodal metastases at the time of surgery with the primary lung NSCLC and concurrent nodal metastases sampled, thus providing a 'snapshot' of each patient's tumour burden. This allows for the study of intra-tumoural heterogeneity of the primary and metastatic tumours,

and inter-tumoural heterogeneity between primary and metastatic sites without the impact of treatment or changes overtime acting as confounding factors.

This will allow for two main outcomes: quantification in a robust cohort of the scale of PD-L1 intra- and inter-tumoural heterogeneity, and to look for the association of PD-L1 expression heterogeneity with other pathological and clinical outcomes. In so doing the scale of PD-L1 expression intra-tumoural and inter-tumoural heterogeneity can be understood, and attempts can be made at what measures can be practically implemented to help minimise the impact of this on clinical decision making.

6.1 Methods

6.1.0 Selection of cohort

All patients from the LLP cohort were reviewed as per the main methods section to ensure suitable quantity and qualities of tissue, as well as to confirm a diagnosis of NSCLC. H&E sections prepared at the time of original pathological assessment, alongside the FFPE blocks, original reports and data from the LLP were used to collect pathological and clinical details for each case. A total of 113 patients were selected for use in this project. A summary of pathological, clinical and demographic findings are shown in table 6.1.0.

Each patient was assigned an arbitrary case number so that all blocks and sections could be pseudo-anonymised. All patients had 1-2 blocks of primary tissue, and 1-5 blocks of tissue containing concurrent metastases to regional lymph nodes sectioned, prepared and stained for SP263 as described in the main methods section, with a parallel section also taken for H&E and labelled with pseudoanonymous case numbers. Both H&E and PD-L1 sections were scanned via Aperio CS2 Scanscope slide scanner at 20x and viewed via Aperio ImageScope or the open source QuPath (v0.1.2) software package. Pseudoanonymised whole sections were scored to give a PD-L1 TPS for each case by myself, with a review by Professor Gosney to form a consensus score for all cases, which could subsequently be linked back to the relevant pathological and clinical details.

	N/Avg.	%
Total	113	
Pathology		
Specimen Age (months)	69	(range 39-113)
Morphology		
ADC	68	60
SCC	45	40
Morphology Subtype		
ADC - Acinar	33	29
ADC - Mucinous	17	15
ADC - Solid	16	14
ADC - Papillary	2	2
SCC - Poorly differentiated	6	5
SCC - moderately differentiated	39	35
Anatomical location of tumour		
LUL	28	25
Lingula	1	1
LLL	13	12
RUL	19	17
RML	6	5
RLL	29	26
Multiple Lobe Involvement	17	15
Tumour size	42 mm	(range 15-130)
Distance to bronchial resection margin	24 mm	(range 0 - 70)
Pleural Status		
PL0	81	72
PL1	12	11
PL2	11	10
PL3	9	8
Necrosis present	70	62
Stage* (at diagnosis)		
T1**	38	34
T2***	55	49
T3	15	13
T4	5	4
Nodal stage (at diagnosis)		
N1	64	57
N2	49	43
N1 & N2	33	29
***ALK* positive**	0	0
***EGFR* positive**	6	5
Pre-surgical specimen - cytology	18	16
Pre-surgical specimen - biopsy	14	12
Post-surgical specimen - cytology	6	5

Post-surgical specimen - biopsy	2	2
Clinical/Demographics		
Median Age (at diagnosis)	68	Range 46-84
Gender		
Male	67	59
Female	46	41
OS status - dead	64	57
COD		
Lung Cancer (C349)	53	47
Other	11	10
Smoking Status		10
Never	11	11
Light (≤20 CSMPYs)	12	80
Heavy (>20 CSMPYs)	90	
Adjuvant therapy		5
TKI	6	65
XCT	74	40
Radiotherapy	45	100

Table 6.1.0 Pathological and clinical characteristics of the 113 patient samples included in the study from the LLP cohort. * Staged according to TNM 7th edition. **Includes T1a-T1c, ***Includes T2a + T2b. ADC, adenocarcinoma; SCC, squamous cell carcinoma; LUL, left upper lobe, LLL, left lower lobe; RUL right upper lobe; RML, right middle lobe; RLL, right lower lobe; ALK, Anaplastic lymphoma kinase; *EGFR*, Epidermal growth factor receptor; OS, overall survival; COD, cause of death; CSMPYs, combined smoking pack years; TKI, tyrosine kinase inhibitor.

6.1.1 Intra-tumoural heterogeneity

Intra-tumoural heterogeneity was assessed on multiple levels. The difficulty presented by the presence of differing types of intra-tumoural heterogeneity and the absence of any formal categorisation of PD-L1 heterogeneity led to the creation of the following approaches that would allow for the objective assessment of small-scale, medium-scale and large-scale intra-tumoural heterogeneity.

First, small scale heterogeneity, defined as heterogeneity within an approximately 1 cm^2 area of tumour was assessed using a grid split into 1 mm squares that was overlaid on to the section (Fig. 6.1.0). Only sections containing a continuous area of viable tumour were assessed; zones of confluent necrosis or fibrosis were avoided and sections in which these were extensive were not used. The PD-1 TPS was assessed for every 1 mm square to give 100 readings for each $1cm^2$ area. The application of this was achieved using QuPath software as illustrated in Fig 6.1.0.

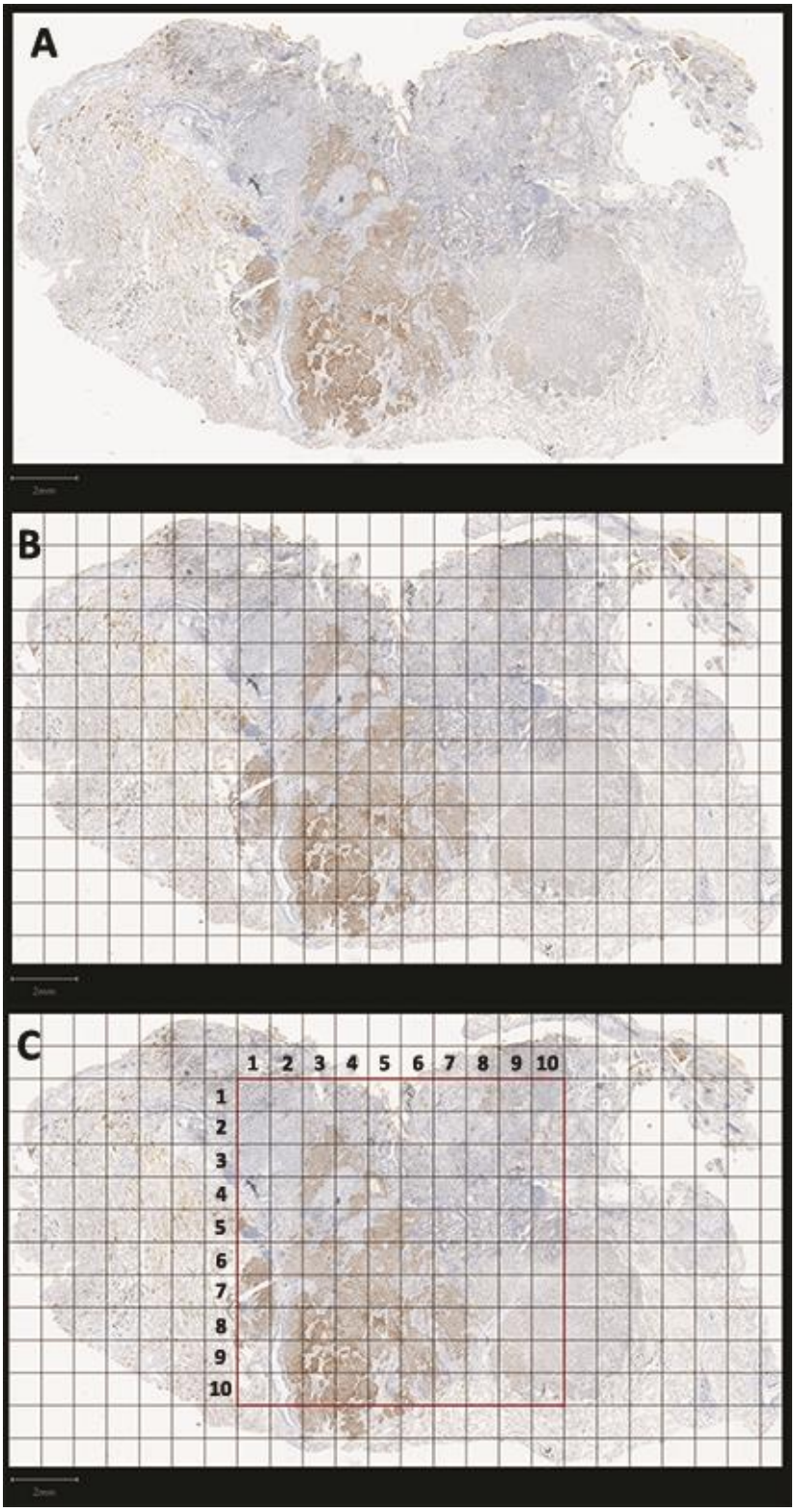


Fig 6.1.0 Creation of a labelled 1cm² area in Qupath by loading a digital image of NSCLC stained for PD-L1 (SP263) (**A**), and then overlaying a grid with spacing set to 1,000 microns (**B**), after which an annotated region can be overlaid to form a 1cm² area, and each mm square assigned co-ordinates (**C**). Individual 1mm squares, from (1,1) to (10,10) can then be scored for a PD-L1 TPS.

Data was captured in a spreadsheet to generate a digital database profile for each case, that could be shown as (X,Y) co-ordinate data. Co-ordinate data was used to form matrices which could be used in a variety of image viewing approaches, such as the R software Plotly Plugin to create heatmaps, with Z stacking based purely on TPS scores. This allows for equal weighting to be given to weak and strong staining, with background or immune staining ignored, which allows for easy visualisation of PD-L1 expression even at low power (Fig 6.1.1). The mean TPS and COV for each 1cm² area is calculated, with the former compared to the whole section TPS, and the latter used as a measure of PD-L1 expression variation (heterogeneity).

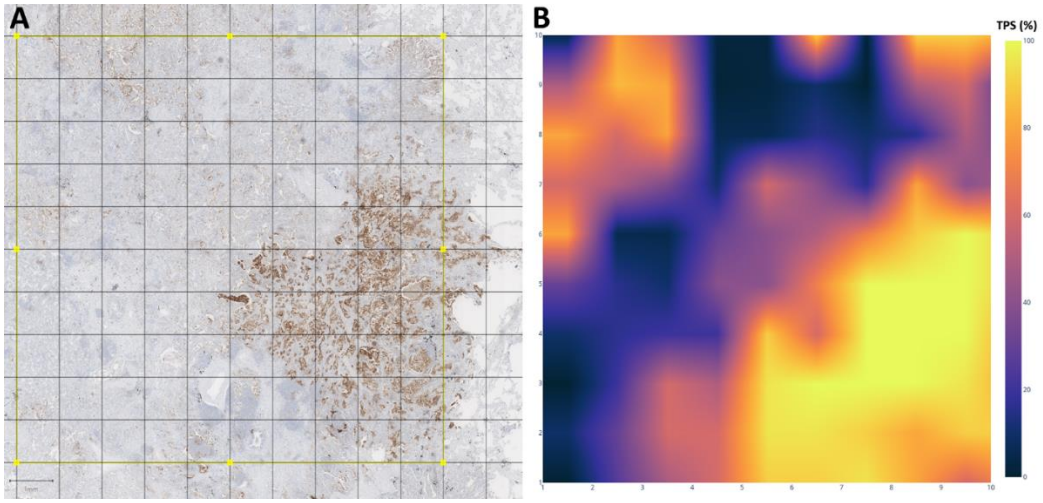


Fig 6.1.1 Data visualisation. **A** - NSCLC section stained for PD-L1 (SP263); at this low power weaker staining is difficult to appreciate and immune cell staining is difficult to differentiate from tumour cell staining. **B** – Heatmap of the same section using TPS scores only – the relative contribution of tumour cell PD-L1 expression from each part of the section is much easier to appreciate at low power.

Between one and three 1 cm squares were assessed in every section studied by this 'squares method' for primary tumours. (Fig 6.1.2). Medium scale heterogeneity, defined as heterogeneity between 1 cm squares, could thus be examined in cases with multiple 1cm squares, to give a broader assessment of intra-tumoural heterogeneity, by comparing mean TPS and COV for each 1cm² area.

In routine clinical practice the largest single piece of tissue used to assess PD-L1 IHC status would be a whole slide section from a surgical specimen. These are therefore used as the 'true' PD-L1 TPS, but to further explore the validity of this assumption, large scale heterogeneity, defined as heterogeneity between different tissue blocks, was assessed for primary tumours by scoring the entire viable tumour region within each section.

Intra-tumoural heterogeneity of nodal metastases was also assessed, but typically lymph node metastases have far less total tumour mass than primary tumour. Therefore 1mm squares were overlaid across the entire specimen, and any square with ≥100 tumour cells was scored for a PD-L1 TPS.

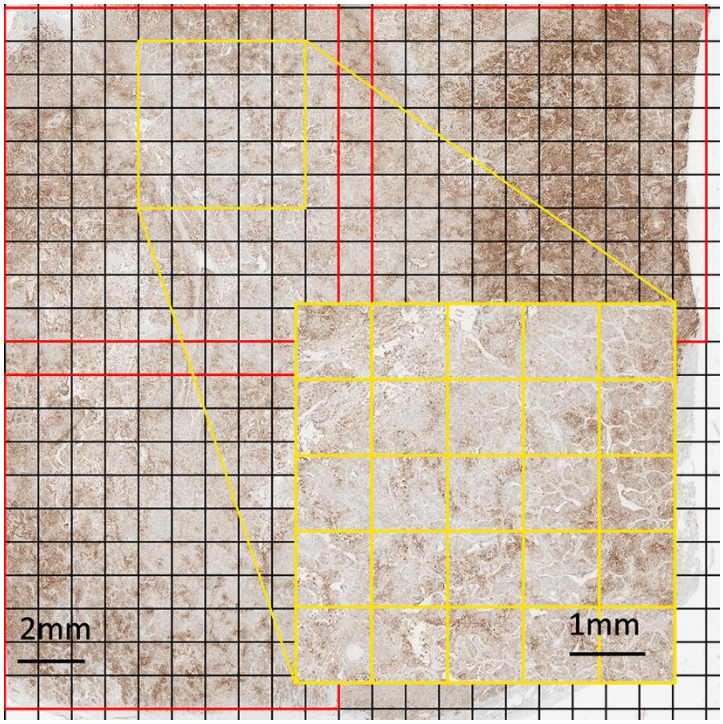


Fig 6.1.2 Medium-scale heterogeneity: NSCLC section stained for PD-L1 (SP63) with multiple 1cm² areas overlaid onto the tumour, for which each provides a measure of small-scale heterogeneity, but also provides data that can be compared between them for medium-scale heterogeneity.

6.1.2 Digital core biopsies

Having established regions of at least 100 mm squares for each case with each mm² area containing a PD-L1 TPS; the simulation of digital core biopsies (DCBs) could be achieved. A good quality needle core-biopsy will achieve an approximately 1x10mm² -1x20mm² area (Fig 6.1.3). By utilising the spatial co-ordinate data and the corresponding TPS at each co-ordinate, multiple mm squares can be combined to generate a mean TPS that simulates the TPS result a core biopsy would achieve if taken from that tumour. A simulated example is illustrated in Fig 6.1.4, showing parts of overlapping 3 DCBs. 10 'horizontal' and 10 'vertical' core biopsies can thus be taken from each square to generate a total of 20 digital core biopsies per 1cm² area. This is repeated for each 1cm² area in any given section, to give 20 to 60 digital core biopsies per section.

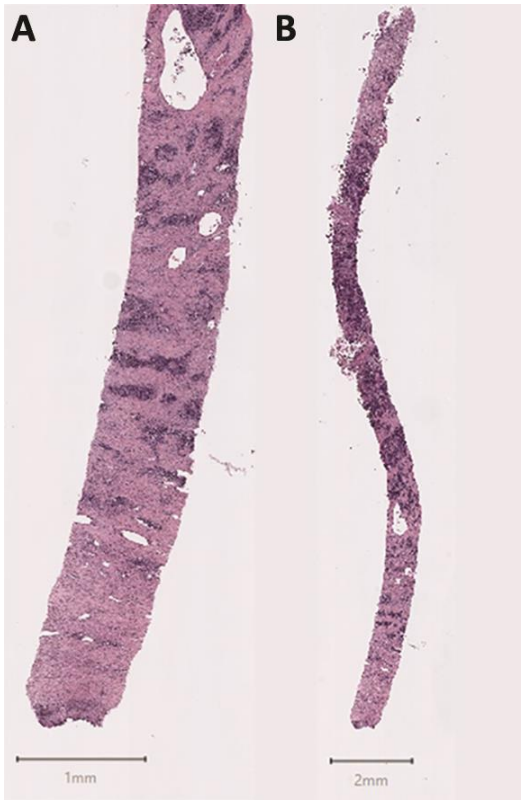


Fig 6.1.3 Example of a good quality core biopsy of a NSCLC. The core is approximately 1mm wide and 20mm long. Core biopsy needles can be pre-set to take 10mm or 20mm cores, but the diameter remains fixed at ~1mm, to produce 1x10mm² or 1x20mm² samples of tissue.

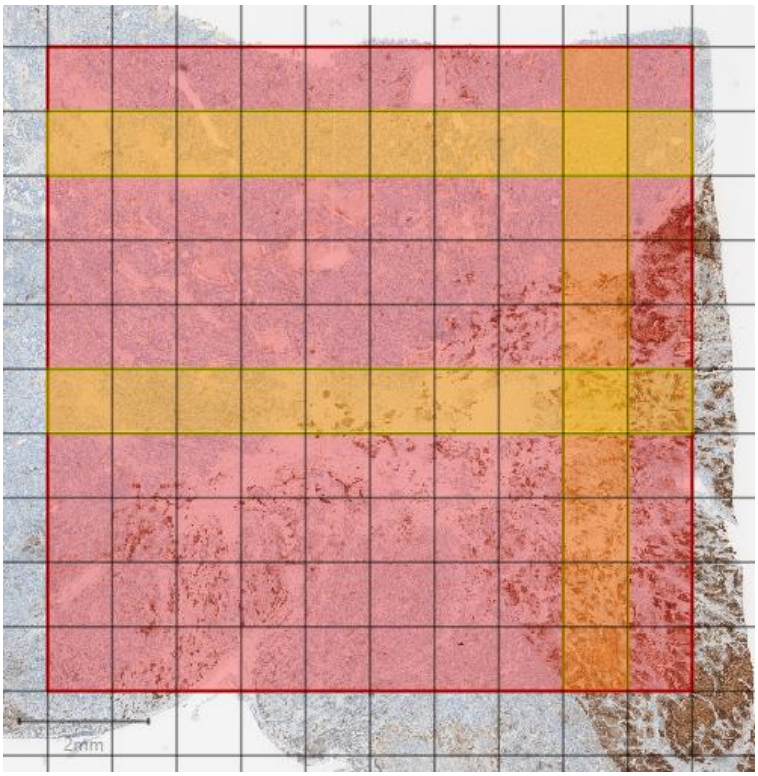
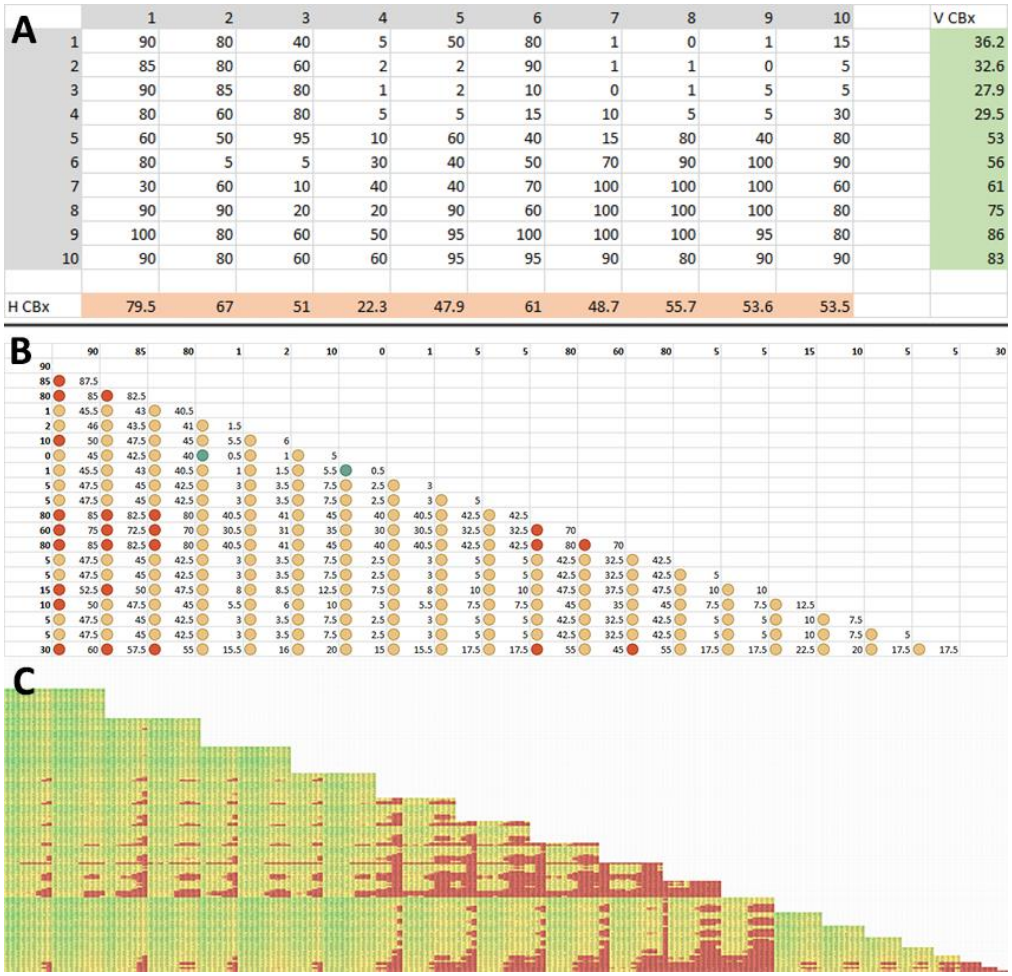


Fig 6.1.4 Simulation of DCBs in primary NSCLC stained for PD-L1 (SP263). A 1 cm² area (red) with visualisation of 3 simulated DCBs (yellow) equivalent to a 1x10mm² area each.

Comparison of a DCB's TPS is made to the relevant whole section TPS. To consider the accuracy of the DCB in scoring PD-L1 expression, the following approach was made: for every case the TPS score for every DCB was compared to the whole-section TPS by utilising the clinical categories of negative, weak and strong positive TPS (<1%, 1-49% and ≥50% respectively). If greater than 99% of cores are in concordance with the clinical category, it is assumed a single core biopsy is adequate for assessing PD-L1 in that tumour. All cases were then assessed for a combination of two DCBs in relation to the whole-slide TPS. This was achieved by creating a 20x20 matrix and an average for each two pairs taken to return a total of 190 DCBs. This can then be compared to the whole section TPS. In instances where two DCBs resulted in less than a 99% correct classification rate of the tumour in PD-L1 expression categories, the use of three DCBs is explored. All possible combinations of 3 DCBs are included (the equivalent of 16,815 averaged TPSs) and every combination score compared to the whole-section TPS to see how often it correctly classifies the tumour, again with an agreement in 99% or greater of instances taken to indicate sufficient tissue is included to correctly classify the tumour. This process was repeated for four and five DCBs in any case where fewer than 99% of simulated biopsies correctly classified the tumour. Example shown in Fig 6.1.5.

Metastatic deposits in lymph nodes underwent the same process. Lymph node metastases are much smaller than primary tumours, so overlaying a 1cm² grid is impractical. Instead any 1mm square with sufficient tissue is included, but co-ordinate data is still captured so that the 1mm squares can be known relative to each other. In this way the mean TPS and COV can still be calculated. Attempts at core biopsies was also made, albeit these will rarely have the full 10mm² area, instead nodes with moderate amounts of metastatic tumour are included, with whatever areas available sampled as for a digital core biopsy (Fig 6.1.6). LNs with small amounts of tumour are essentially 'fully sampled' even with just equivalent to a core biopsy or two, and therefore these have been excluded. For larger LNs, digital core biopsies to a maximum of two were considered in the same method as for primary tissue.



	1	2	3	4	5	6	7	8	9	10	V CBx
1	90	80	40	5	50	80	1	0	1	15	36.2
2	85	80	60	2	2	90	1	1	0	5	32.6
3	90	85	80	1	2	10	0	1	5	5	27.9
4	80	60	80	5	5	15	10	5	5	30	29.5
5	60	50	95	10	60	40	15	80	40	80	53
6	80	5	5	30	40	50	70	90	100	90	56
7	30	60	10	40	40	70	100	100	100	60	61
8	90	90	20	20	90	60	100	100	100	80	75
9	100	80	60	50	95	100	100	100	95	80	86
10	90	80	60	60	95	95	90	80	90	90	83
H CBx	79.5	67	51	22.3	47.9	61	48.7	55.7	53.6	53.5	

Fig 6.1.5 DCBs data collection. **A** – Matrix representing a 1cm² area with ten vertical (Green) and ten horizontal (orange) core biopsies, each biopsy taken as the average of 10 1mm² areas. **B** – Combination of any two possible core biopsies from **A** ; a traffic light system allows rapid visualisation of negative (<1% in Green), weak positive (1-49% in yellow) and strong positive (≥50% in red). **C** – Combination of any possible 3 core biopsies from **A**. Traffic light system allows for a general overview, though automated equations are used to harvest the data.

6.1.3 Digital fine needle aspirates

In addition to the DCBs, digital fine needle aspirates (DFNAs) were also explored in each case as an additional metric of ascertaining the quantity of tissue required to correctly classify each tumour's PD-L1 expression. DCBs provide a continuous piece of tissue across two axis to simulate biopsy sampling. However, as a section of tissue is a 2D representation of a 3D structure and as the reality of sampling is that it can occur across any plane, with movement of tumour likely during the process, it does not fully mimic the difficulties of sampling in actuality.

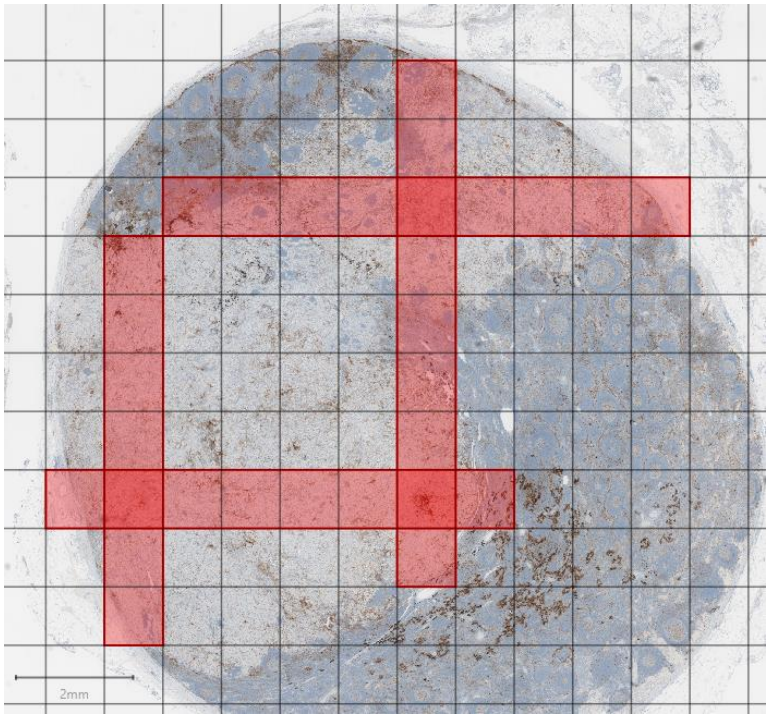


Fig 6.1.6 Simulation of DCBs in a lymph node stained for PD-L1 (SP263). Each core (Red) is smaller than the 1x10mm² used in primary tissue, with the potential for discontinuous areas to be included in a single core. This approach attempts to mimic the sampling of lymph nodes in reality.

In addition, some tumours, particularly metastatic deposits in lymph nodes accessed via EBUS, undergo needle aspiration to sample the tumour instead of a biopsy. This requires the movement of a needle in multiple planes with constant pressure, and will typically take single cells or small groups of cells with minimal architecture preservation. As a result, cytology aspirates might be considered to sample tumours in far more random fashion than core biopsies.

In a bid to replicate this digitally, DFNAs was used to compare multiple 1mm squares to the whole section TPS by selecting mm² areas by (X,Y) co-ordinate data at random. 10, 20 and 30, 1mm squares were taken at random for each primary tumour. For cases which had not achieved a ≥99% correct classification rate by the equivalent of a 30mm² area, DFNA was increased to 40 and 50 1mm squares. 10 mm squares at random is equivalent to 1 DCB, 20 mm squares to 2 DCBs and so forth, so a comparison between DCBs and DFNAs can also be made.

6.1.4 Inter-tumoural heterogeneity

The whole-section TPS from the primary tumour was compared to the TPS from lymph node metastases in each case. Lymph nodes identified as separate nodes are scored individually (thus a section may have two TPS results if it contains two distinct nodes), otherwise all lymphoid tissue

with metastatic deposits on a single slide are included together to generate an overall TPS. All cases had at least one N1 or one N2 node included, and for cases with both N1 and N2 disease at least one node from each category would be scored for TPS. In instances where any discrepancy between primary and metastatic tissue PD-L1 expression was found, all lymph nodes with metastatic tumour would subsequently be included. In so doing cases will have primary and metastatic tumour compared for PD-L1 expression to assess for inter-tumoural heterogeneity, and a subset of cases will have multiple metastatic deposits compared to each other as another metric of inter-tumoural heterogeneity.

6.1.5 Heterogeneity and pathological/clinical features

Intra-tumoural and inter-tumoural heterogeneity as defined in this project will allow the grouping of NSCLC cases by similar patterns of PD-L1 expression in addition to TPS. These groups were used to look for relationships to other pathological and clinical features as outlined in table 6.1.0. In so doing patient or tumour features that may predict or be affected by PD-L1 expression heterogeneity maybe identified as potential areas for further investigation.

6.1.6 Statistics

Data analysis, including testing of groups for significance is performed by the same statistical approach as outlined in the main methods section. TPS datasets typically failed to meet assumptions for parametric testing. Comparisons between primary and metastatic samples are independently tested, as though related they are not repeated measures. COV is calculated using the following equation: COV = (σ / μ)*100. Statistical significance for the comparisons of COV were performed as described by Forkam *et al* 2009.[571] DFNA was performed in Microsoft Excel using the following equation: N=INDEX(A:A,RANK(BN,B:B)) where A:A is the array containing each mm squared TPS, B:B is the array containing function 'rand()', and BN is a randomly selected TPS from A:A, unique to N. 10 numbers could then be selected from the N:N array to provide 10 random TPS from the original A:A array to replicate random sampling of the tumour, increasing to 20, 30, 40 or 50 squares as required. DCB arrays as illustrated in Fig 6.1.5 used Boolean TRUE/FALSE operators in relation to TPS cut-offs (e.g. =IF(AND(N>=1,N<50),TRUE,FALSE) when ascertaining if N is in the 1-49% category) with sum total of TRUE/FALSE returns used to generate percentage accuracies of sampling for each case. Kaplan-Meier Log-Rank survival analysis was used to compare PD-L1 heterogeneity expression groups for differences in OS.

6.2 Results

6.2.0 PD-L1 intra-tumoural heterogeneity – descriptives

A visual review of PD-L1 expression in all 113 cases and their matched nodes revealed variable patterns of heterogeneity with differences of PD-L1 seen between cm² areas (Fig 6.0.1), mm² areas or even single cells (Fig 6.2.0). Expression of PD-L1 varied from uniform expression to wide-spread scattered PD-L1 expression or focal/multi-focal expression. (Fig 6.2.2 A-D) In addition, transition from positive to negative cells can be abrupt or gradual (Fig 6.2.2 E-F) and tumours can express a mixture of these features (Fig 6.2.2 G-H). In both primary and metastatic deposits the invading edge, that is, the area of tumour that is infiltrating normal tissue, can be shown to have increased PD-L1 expression over the rest of the tumour (Fig 6.2.3), but in other cases expression had no apparent relationship to anatomy. Tumours could have a mixture of both strong and weak expression for PD-L1 (Fig 6.2.0, Fig 6.2.1) and to have a mix of both focal and scattered patterns of PD-L1 expression. (Fig 6.2.1) A qualitative summary of each cases' most prominent pattern of heterogeneity (if present) is shown in Table 6.2.6.

6.2.1 Primary NSCLC - Small-scale heterogeneity and digital core biopsies

There was sufficient quantity and quality (≥ 1 cm² of continuous viable tumour cells) for assessment by the 'squares method' in 72 cases of primary NSCLC. These were found to have a large range of COV (from 0 to 623) with a mean COV of 140, illustrating that cases ranged from no heterogeneity of PD-L1 expression (i.e. 0% or 100% TPS) to extremely heterogeneous expression. There was similar findings if splitting tumours into morphology for both ADCs (n=42, COV = 120) and SCCs (n=30, mean COV = 170). Six cases (8%) had a 1cm² area mean TPS different to its whole section TPS such that it would change clinical category. The average difference between whole section TPS and a 1cm² area TPS was 1 (p=0.195), with eight cases (11%) have a TPS change of ≥10%.Example of PD-L1 expression heterogeneity between 1mm² areas in shown in Fig 6.2.0.

All 72 specimens had a single DCB and all possible 2 DCBs combinations simulated. Cases with less than a 99% correct classification rate at 2 DCBs would be tested for 3 DCBs, increasing to 4 and 5 DCBs if the ≥99% classification rate cannot be achieved. A total of 437,445 simulated core biopsies in 72 specimens was thus performed. All 72 cases underwent DFNA sampling for the equivalent tissue of 10, 20, and 30 1mm² areas, with cases not achieving ≥99% concordance having 40 or 50 1mm² areas. A total of 288 DFNA simulations was thus performed. These are summarised in table 6.2.0.

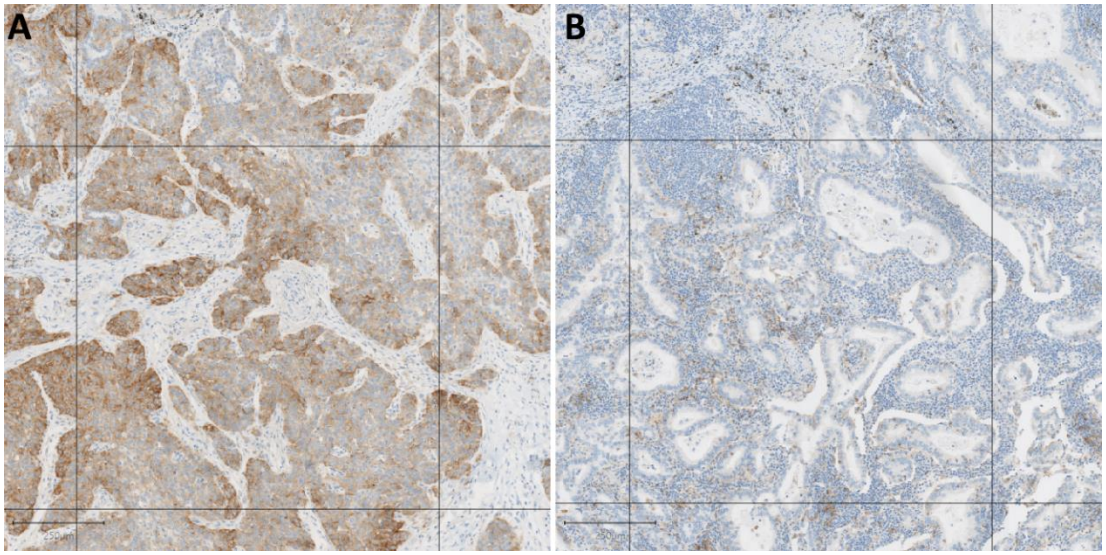


Fig 6.2.0 Example of small-scale heterogeneity: NSCLC section stained for PD-L1 (SP63). **A** – 1mm² area with a TPS of 95% **B** – 1mm² area with a TPS of <1%. **A** and **B** are from the same 1cm² area of primary NSCLC.

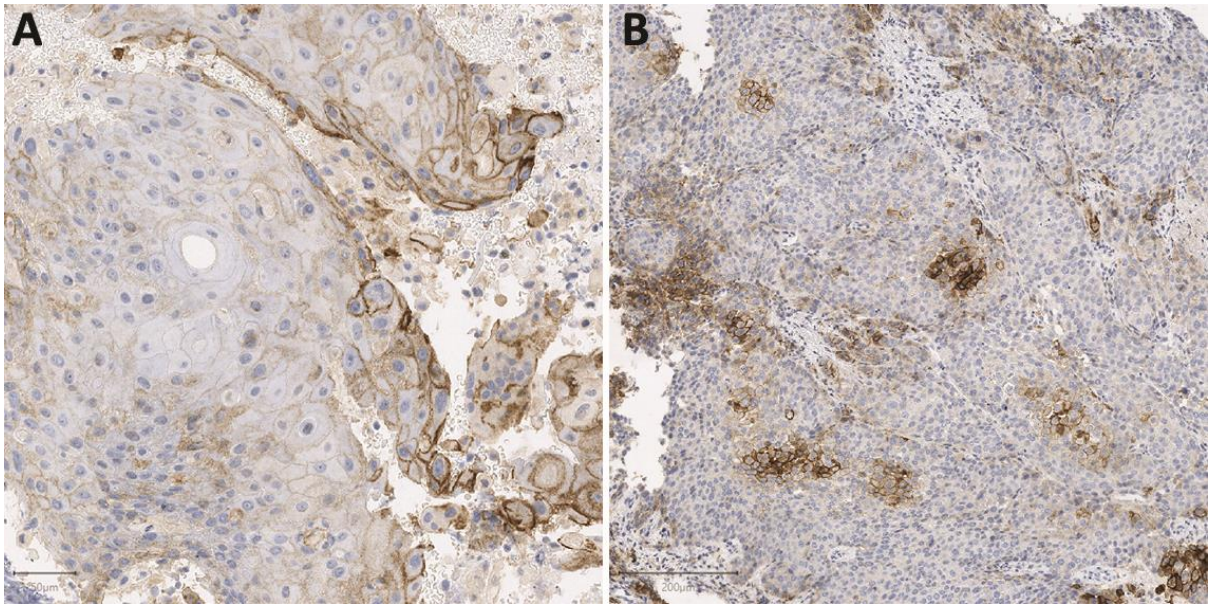


Fig 6.2.1 Patterns of PD-L1 (SP263) expression. A – Weak to strong staining. B – Multi-focal and scattered single cells of PD-L1 positive tumour cells.

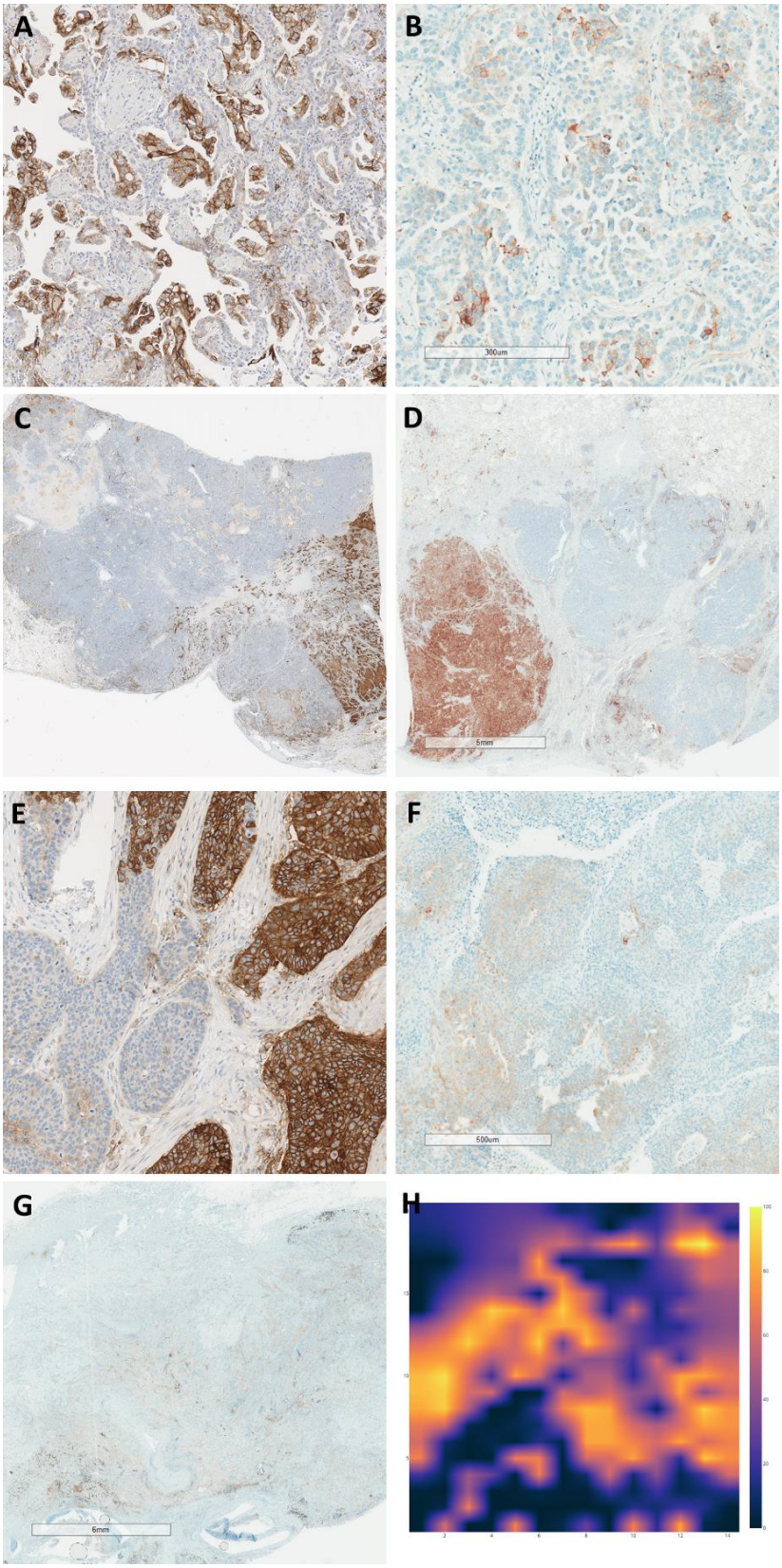


Fig 6.2.2 Patterns of PD-L1 (SP263) expression heterogeneity. **A, B** –small scale heterogeneity with scattered expression. **C, D** – large scale heterogeneity with areas of focal expression. **E** – Abrupt change from positive to negative cells for PD-L1. **F** – Gradual change from positive to negative cells for PD-L1. **G, H** – slide and heat map of widespread scattered heterogeneity with a mixture of heterogeneity features.

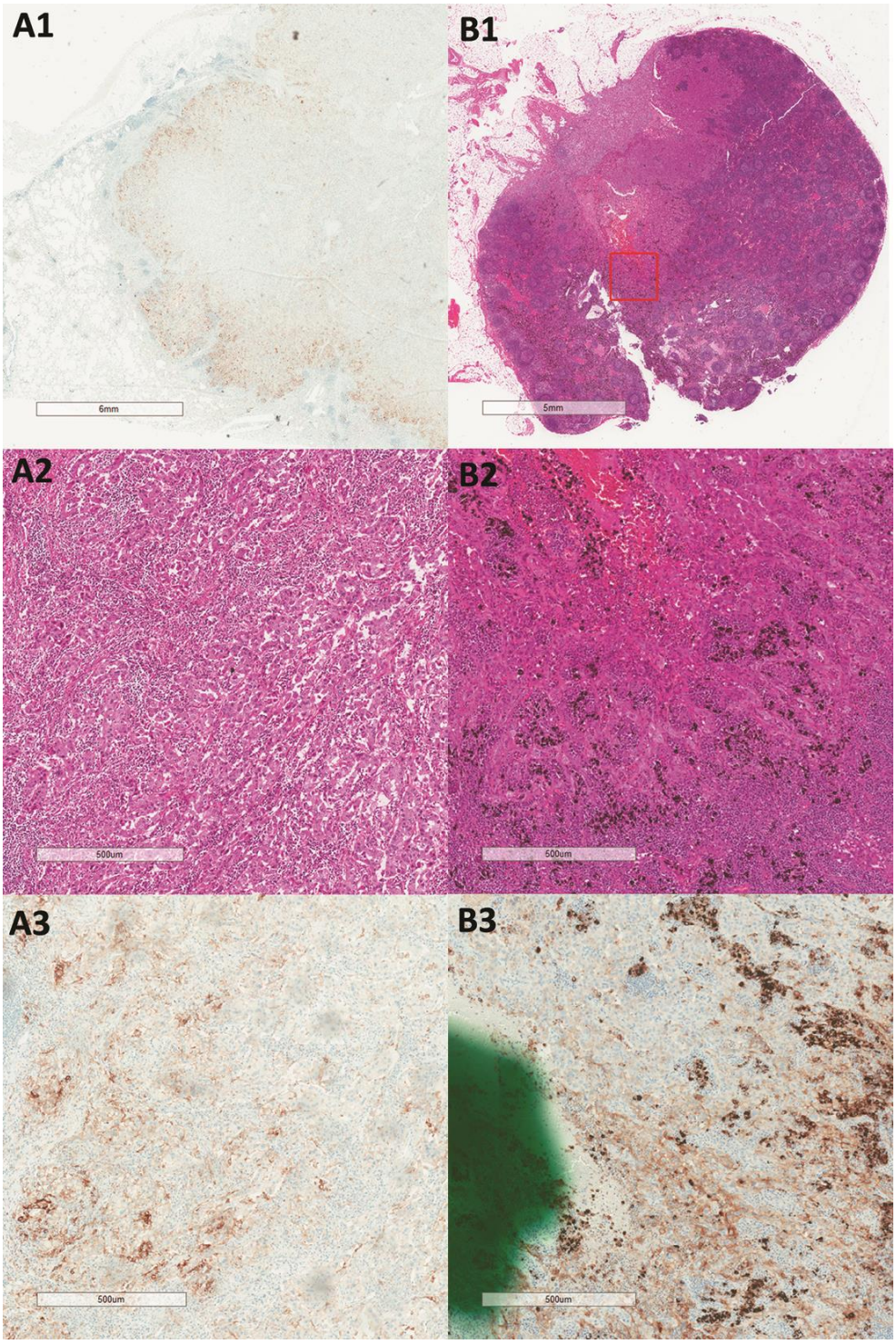


Fig 6.2.3 Infiltrating edge of tumour preferentially expressing PD-L1 (SP263). **A1** – PD-L1 expression in primary NSCLC at low power with increased expression at the border between tumour and non-malignant tissue. **A2, A3** – H&E and PD-L1 of infiltrating border of primary NSCLC at high power. **B1** – H&E low power view of lymph node with metastatic deposit of NSCLC. **B1, B2** – H&E and PD-L1 expression at infiltrating border of NSCLC in a lymph node at high power.

Case	TPS	Clinical Category	Square TPS	DCB simulations misclassification rate (%)					DFNA Simulations misclassification rate (%)					COV
				1	2	3	4	5	1	2	3	4	5	
1	5	1-49%	5	60	45	28	22	10	1	0	0			408
2	0	<1%	<1	5	0.1				0	0	0			180
3	1	1-49%	5	18	7	2	0.1		17	4	3	0		231
4	0	<1%	0	0	0				0	0	0			0
5	5	1-49%	5	15	6	0.1			13	6	0			185
6	95	≥50%	100	0	0				0	0	0			10
7	30	1-49%	20	0	0				0	0	0			69
8	20	1-49%	20	0	0				0	0	0			92
9	0	<1%	<1	0	0				0	0	0			459
10	0	<1%	<1	0	0				0	0	0			358
11	30	1-49%	30	1	6	0.1			1	0	0			86
12	1	1-49%	1	15	2	0.1			3	0	0			142
13	1	1-49%	1	55	35	19	14	6	63	52	48	36	30	208
14	70	≥50%	70	10	2	0.1			2	0	0			41
15	0	<1%	<1	0	0				0	0	0			198
16	80	≥50%	70	0	0				0	0	0			30
18	0	<1%	<1	15	4	0.9			8	2	0			341
19	5	1-49%	5	20	14	5	2	0.01	10	4	1	0		138
20	80	≥50%	85	0	0				0	0	0			15
22	5	1-49%	5	0	0				0	0	0			163
24	10	1-49%	1	23	12	4	2	0.01	21	8	3	0		231
25	0	<1%	<1	0	0				0	0	0			248
26	50	≥50%	50	25	12	4	4	0.6	42	38	33	28	15	72
28	1	1-49%	1	25	21	7	8	8	24	6	9	4	0	201
31	10	1-49%	20	8	0.1				3	0	0			183
32	0	<1%	0	0	0				0	0	0			236
33	0	<1%	0	0	0				0	0	0			403
36	50	≥50%	55	25	28	25	22	18	32	28	36	12	0	54
39	10	1-49%	10	0	0				0	0	0			231
40	85	≥50%	90	0	0				0	0	0			13
41	40	1-49%	50	15	6	0.1			10	12	3	0		94
44	30	1-49%	55	5	4	0.2			6	2	0			91
46	5	1-49%	5	0	0				0	0	0			48
48	85	≥50%	80	0	0				0	0	0			30
49	80	≥50%	80	0	0				0	0	0			21
51	80	≥50%	80	0	0				0	0	0			36
52	20	1-49%	30	5	0.1				2	0	0			154
53	10	1-49%	2	27	19	6	5	5	12	2	0			191
55	60	≥50%	55	35	33	35	29	14	36	30	30	24	20	68
56	90	≥50%	95	0	0				0	0	0			10
59	10	1-49%	1	45	19	5	6	1	24	10	3	0		282
60	50	≥50%	50	35	35	31	38	30	40	28	24	16	15	68
61	100	≥50%	100	0	0				0	0	0			8
64	35	1-49%	35	30	8	2	0.2		29	10	3	0		137
69	25	1-49%	25	0	0				10	2	3	0		154
75	40	1-49%	40	30	26	17	27	15	22	26	1	0		90
80	40	1-49%	50	35	30	21	27	18	23	36	12	8	8	86
82	50	≥50%	50	25	19	15	17	11	18	10	9	4	0	51
85	10	1-49%	10	0	0				0	0	0			69
86	90	1-49%	10	5	1	0.1			0	0	0			176

87	5	1-49%	5	0	0				0	0	0			162
89	0	<1%	<1	5	0.1				6	10	0			623
90	0	<1%	0	0	0				0	0	0			0
91	85	≥50%	80	0	0				0	0	0			29
92	0	<1%	0	0	0				0	0	0			0
93	<1	<1%	1	30	36	24	29	19	27	36	18	16	15	303
94	<1	<1%	<1	40	32	31	27	26	52	54	48	48	45	127
95	<1	<1%	<1	35	30	24	24	15	42	42	36	36	40	196
96	1	1-49%	1	0	0				3	0	0			176
100	0	<1%	0	0	0				0	0	0			169
101	100	≥50%	100	0	0				0	0	0			0
102	90	≥50%	90	0	0				0	0	0			0
103	0	<1%	0	0	0				0	0	0			243
105	0	<1%	0	0	0				0	0	0			0
106	100	≥50%	100	0	0				0	0	0			3
107	0	<1%	<1	0	0				0	0	0			268
108	30	1-49%	50	35	13	14	20	14	0	0	0			79
109	40	1-49%	70	25	13	4	3	0.2	0	0	0			63
110	5	1-49%	5	5	0.1				0	0	0			135
111	5	1-49%	5	15	3	0.1			8	0	0			262
112	75	≥50%	75	8	0.1				1	0	0			31
113	20	1-49%	20	5	0				0	0	0			103

Table 6.2.0 Small-scale heterogeneity of PD-L1 expression in 72 NSCLCs and the accuracy of sampling by DCBs and DFNA for PD-L1 expression. TPSs for DCBs and DFNA are compared to TPSs from the section that they were sampled from, and the rate of misclassification by multiple DCBs or DFNA is recorded as a percentage of the total number of simulations performed using negative, weak or strong positive clinical categories (<1%, 1-49%, ≥50% TPS). COV is a measure of heterogeneity.

Correct Classification rate	**DCB1**	**DCB2**	**DCB3**	**DCB4**	**DCB5**
Cases %	49	58	72	75	82
Simulations %	89	93	95	95	97
	DFNA1	**DFNA2**	**DFNA3**	**DFNA4**	**DFNA5**
Cases %	54	67	74	85	89
Simulations %	92	94	96	97	97

Table 6.2.1 Summary of the accuracy of DCBs and DFNAs of NSCLC for scoring PD-L1. Misclassification rate is given as a percentage of simulations that resulted in the incorrect classification of a tumour by PD-L1 expression, and the percentage of cases affected by this using negative, weak or strong positive clinical categories (<1%, 1-49%, ≥50% TPS).

The use of a single DCB resulted in a correct classification of a tumour (<1%, 1-49%, ≥50% TPS) in 89% of all simulations across the entire cohort. The sum total of all misclassifications (11%) occurred in 37 cases (51%), with 35 cases (49%) correctly categorised by a single DCB ≥99% of the time. This improved to a correct classification rate in 95% of all simulations, resulting in 51 cases (71%) correctly categorised by all of their respective three DCB combinations ≥99% of the time. By five DCBs, only 3% of all simulations were misclassifying tumours, but this still affected 13 cases (18%) for whom even five DCBs failed to correctly classify the tumour ≥99% of the time, with several cases being misclassified 20-30% of the time.

The use of DFNA equivalent to one core biopsy resulted in the correct classification of a tumour in 92% of all simulations, with 39 cases (54%) correctly categorised by DFNA1 ≥99% of the time. This improved with increasing DFNA quantities to a correct classification rate in 97% of all simulations when using DFNA equivalent to five (DFNA5) core biopsies, with 64 cases (89%) correctly classified 1 ≥99% of the time. Eight cases (11%) failed to be correctly categorised by DFNA5 ≥99% of the time, of which seven cases were the same as those that failed to be correctly categorised by 5 DCBs, with several cases being misclassified 40-45% of the time.

The use of DFNA to the equivalent tissue of 1-5 core biopsies found a similar number of incorrect classifications, with a modest non-significant reduction by this method compared to DCBs (8% vs 11% p=0.252 for 1 DCB/DFNA, 6% vs 7% p=0.990 for 2DCBs/DFNA, 4% vs 5% p=0.683 for 3 DCBs/DFNA, 3% vs 5% p=0.238 for 4 DCBs/DFNA, 3% vs 3% p=0.980 for 5 DCBs/DFNA). (Table 6.2.1).

6.2.2 Primary NSCLC - Medium-scale heterogeneity

A total of 27 cases had sufficient quantity and quality (≥ 2 cm^2 of viable tumour cells) for assessment by the 'squares method' in 2 independent squares on a single slide, with eight of these cases having sufficient tissue (≥ 3 cm^2 of viable tumour cells) for three independent squares on a single slide, with an overall mean TPS for each square generated. The average TPS difference between 1cm^2 areas was seven (p=0.368), with six cases (22%) changing clinical groups (<1%, 1-49%, ≥50% TPS) between 1cm^2 areas. Six cases (22%) had a TPS change of ≥10% between 1cm^2 areas, with ten cases (36%) having a significant difference in COV between 1cm^2 areas. Data are shown in Table 6.2.2.

Case	Whole Section TPS	Square 1 Mean	Group	COV	Square 2/3 Mean	COV	COV change sig.	Group Change?	TPS change
2	<1	<1	<1%	168	<1	380		N	0
3	1	5	1-49%	163	5	290		N	0
4	0	0	<1%	0	0	0		N	0
5	5	5	<1%	329	5	187		Y	2
6	95	100	≥50%	0	100	0		N	0
7	30	20	1-49%	138	25	70	Y p<0.001	N	6
					62	52		Y	44/37
8	20	20	1-49%	245	19	118	Y p=0.05	N	13
9	0	<1	<1%	260	<1	572		N	0
					<1	398		N	0
10	0	<1	<1%	505	<1	360		N	0
					<1	419		N	0
11	30	30	1-49%	82	30	111		N	0
19	5	5	<1%	221	2	137		Y	1
					1	176		Y	1/0
22	5	5	1-49%	177	9	164		N	4
24	10	1	1-49%	137	12	135		N	1
31	10	20	1-49%	220	15	177		N	6
					15	187		N	4/2
36	50	55	≥50%	31	75	54	Y p<0.001	N	18
39	10	10	1-49%	119	5	400	Y p=0.018	N	16
49	90	80	≥50%	3	99	28	Y p<0.001	N	21
					85	16		N	15/6
51	80	80	≥50%	28	75	38	Y p=0.003	N	7
55	60	55	1-49%	92	70	50	Y p<0.001	Y	26
					50	65		Y	10/16
56	90	95	≥50%	12	95	9	Y p<0.001	N	3
61	100	100	≥50%	1	95	13	Y p<0.001	N	3
62	2	2	1-49%	139	5	224		N	2
					2	196		N	0/2
65	1	1	1-49%	144	1	183		Y	1
72	0	<1	<1%	139	1	180		N	0
83	0	<1	<1%	172	<1	200		N	0
85	10	10	1-49%	158	15	178		N	6
104	60	<1	<1%	168	<1	380		N	0

Table 6.2.2 Medium-scale heterogeneity of PD-L1 expression in 27 NSCLCs. 2 to 3 1cm² areas for each case are compared for TPS change with 8 cases (30%) having differences in TPS sufficient to change clinical categories of negative, weak or strong positive (<1%, 1-49%, ≥50% TPS). COV as a measure of heterogeneity is also compared between 1cm² areas, with 10 cases (37%) found to have a significant difference.

6.2.3 Primary NSCLC - Large-scale heterogeneity

A total of 61 cases had sufficient tissue in two blocks of primary NSCLC for a whole section of both blocks to be assessed for PD-L1 expression. The average TPS change between blocks was 4 (p=0.910), with four cases (7%) having sufficient difference in TPS to change clinical categories. 12 cases (20%) had a TPS change of ≥10% between blocks. Results are shown in Table 6.2.3.

Case	Block 1 TPS	Block 2 TPS	Difference	Clinical change
2	0	0	0	N
4	0	0	0	N
9	0	0	0	N
10	0	0	0	N
15	0	0	0	N
17	0	0	0	N
18	0	0	0	N
21	0	0	0	N
23	0	0	0	N
25	0	0	0	N
30	0	0	0	N
33	0	0	0	N
34	0	0	0	N
35	0	0	0	N
37	0	0	0	N
38	0	0	0	N
43	0	0	0	N
58	0	0	0	N
3	1	1	0	N
13	1	1	0	N
27	1	1	0	N
28	1	1	0	N
45	1	1	0	N
57	1	1	0	N
65	1	1	0	N
66	1	1	0	N
29	2	2	0	N
62	2	0	2	Y
1	5	20	15	N
5	5	1	4	N
19	5	1	4	N
22	5	1	4	N
42	5	1	4	N
46	5	5	0	N
24	10	1	9	N
31	10	20	10	N
39	10	10	0	N
50	10	5	5	N

59	10	5	5	N
85	10	10	0	N
8	20	25	5	N
52	20	30	10	N
7	30	50	20	Y
11	30	70	40	Y
64	35	25	10	N
26	50	40	10	Y
36	50	55	5	N
44	50	60	10	N
60	50	40	10	N
55	60	80	20	N
16	80	70	10	N
47	80	85	5	N
51	80	75	5	N
40	85	80	5	N
48	85	90	5	N
49	90	80	10	N
56	90	90	0	N
63	90	90	0	N
67	90	90	0	N
86	90	95	5	N
6	95	90	5	N

Table 6.2.3 Large-scale heterogeneity of PD-L1 expression in 61 NSCLCs. A whole section from 2 blocks of primary NSCLC is compared for TPS in each case, with 4 cases (7%) having differences in TPS sufficient to change clinical categories of negative, weak or strong positive (<1%, 1-49%, ≥50% TPS)

6.2.4 Lymph node (secondary) metastatic deposits – intra-tumoural heterogeneity

Using a modified version of the 'squares method', small-scale heterogeneity within lymph node metastatic deposits was examined. A total of 30 cases were selected with sufficient quantities of tissue (>0.5cm² of continuous viable tumour cells). LN metastases demonstrated similar patterns of variation in heterogeneity to primary tumour, (Fig 6.2.4 and Fig 6.2.1), however the COV range was smaller than primary NSCLC (0-398) with a significantly smaller mean COV (84 vs 140 $p<0.0001$) (Table 6.2.4).

Metastatic deposits are on average far smaller than primary tumours, so DCBs and DFNAs were only performed to the equivalent of two core biopsies/20 1mm² squares for the LN samples. The use of a single DCB resulted in the correct classification of a tumour (<1%, 1-49%, ≥50% TPS) in 88% of all simulations across the entire cohort. The sum total of all misclassifications (12%) occurred in 30 cases (33%), with 40 cases (67%) correctly categorised by a single DCB ≥99% of the time.

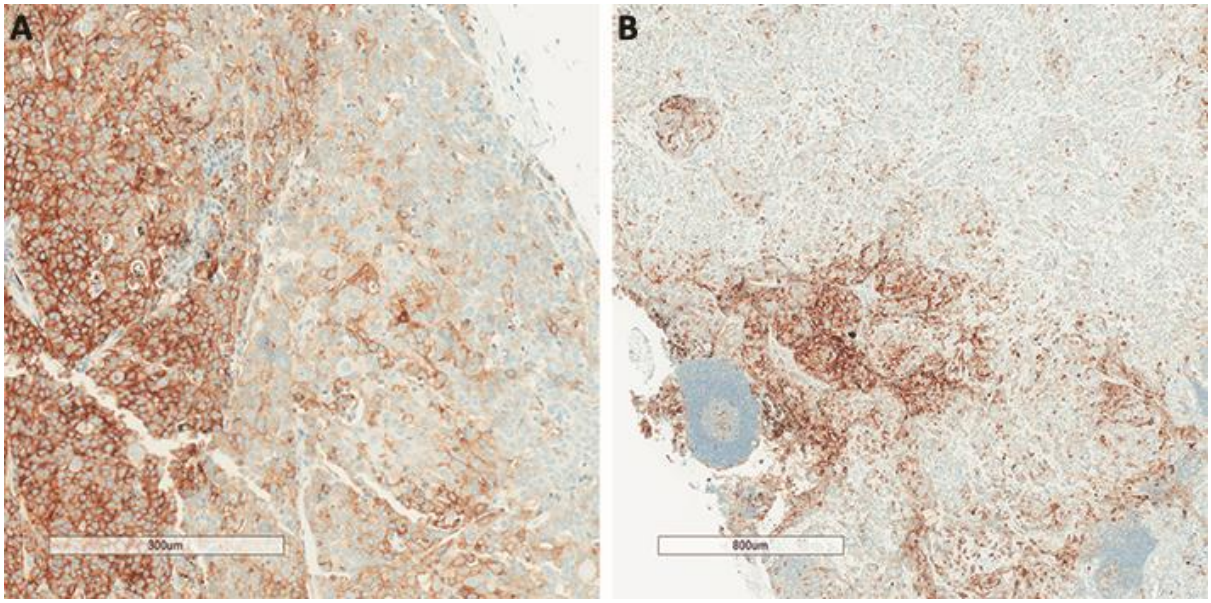


Fig 6.2.4 PD-L1 (SP263) heterogeneity in NSCLC metastatic deposits in lymph nodes. **A** – A gradual change from positive to negative cells with various intensities of staining. **B** – Focal staining of PD-L1 with occasional scattered positive tumour cells.

Case	Whole TPS	Category	DCB 1	DCB 2	DFNA 1	DFNA 2	COV
7	30	1-49%	9	13	1	0	101
11	70	≥50%	0	0	0	0	23
14	45	1-49%	44	45	44	32	67
19	90	≥50	0	0	0	0	6
22	<1%	<1%	13	0	0	0	126
24	80	≥50%	0	0	0	0	21
29	90	≥50%	0	0	0	0	3
32	25	1-49%	33	13	1	0	133
35	5	1-49%	0	0	2	0	35
36	95	≥50%	0	0	0	0	13
38	50	≥50%	47	37	24	16	45
39	10	1-49%	0	0	0	0	114
40	10	1-49%	0	0	0	0	85
41	100	≥50%	0	0	0	0	0
42	0	<1%	0	0	0	0	0
46	5	1-49%	41	21	19	0	289
48	<1%	<1%	0	0	0	0	300
52	90	≥50%	0	0	0	0	6
54	<1%	<1%	44	22	24	0	79
55	<1%	<1%	41	48	33	34	197
58	55	≥50	0	0	9	4	44
60	1	1-49%	57	51	42	20	205
61	80	≥50%	0	0	0	0	17
64	90	≥50%	0	0	0	0	16

65	70	≥50%	16	5	4	0	34
66	95	≥50%	0	0	0	0	5
67	<1%	<1	0	0	0	0	397
69	5	1-49%	0	0	0	0	54
71	15	1-49%	0	0	0	0	112
72	0	<1%	0	0	0	0	0

Table 6.2.4 Intra-tumoural heterogeneity of PD-L1 expression in 30 lymph nodes containing metastatic NSCLC and the accuracy of sampling by DCBs and DFNA for PD-L1. TPSs for DCBs and DFNAs are compared to TPSs from the section that they were sampled from, and the rate of misclassification by multiple DCBs or DFNA is recorded as a percentage of the total number of simulations performed using negative, weak or strong positive clinical categories (<1%, 1-49%, ≥50% TPS). COV is a measure of heterogeneity.

This improved to a correct classification rate in 91% of all simulations, resulting in 42 cases (70%) correctly categorised by all of their respective two DCB combinations ≥99% of the time. Nine cases (30%) failed to be correctly categorised by two DCBs ≥99% of the time with several cases being misclassified 40-50% of the time. The use of DFNA equivalent to one core biopsy resulted in the correct classification of a tumour in 93% of all simulations, with 38 cases (63%) correctly categorised by DFNA 1 ≥99% of the time. This improved to a correct classification rate in 96% of all simulations when using DFNA equivalent to two core biopsies, with 50 cases (83%) correctly classified ≥99% of the time. Five cases (17%) failed to be correctly categorised by DFNA 2 ≥99% of the time, with several cases being misclassified 20-30% of the time. Results are summarised in Table 6.2.5.

	DCB1	**DCB2**
Cases %	67	70
Simulations %	88	91
	DFNA1	**DFNA2**
Cases %	63	83
Simulations %	93	96

Table 6.2.5 Summary of the accuracy of DCBs and DFNAs of metastatic NSCLC for scoring PD-L1. Misclassification rate is given as a percentage of cases and simulations that resulted in the incorrect classification of a tumour for PD-L1 expression using negative, weak or strong positive clinical categories (<1%, 1-49%, ≥50% TPS) by 1 or 2 DCBs, or DFNAs equivalent to 10mm² or 20mm² area.

focal expression, for which repeated sampling does not improve the chances of correct classification. Example is shown in Fig 6.2.5. The use of DFNA to the equivalent tissue of 1-2 core biopsies found a significantly reduced number of incorrect misclassifications compared to use of one or two DCBs respectively (7% vs 12% p=0.025 for 1 DCB/DFNA, 4% vs 9% p=0.008 for 2DCBs/DFNA).

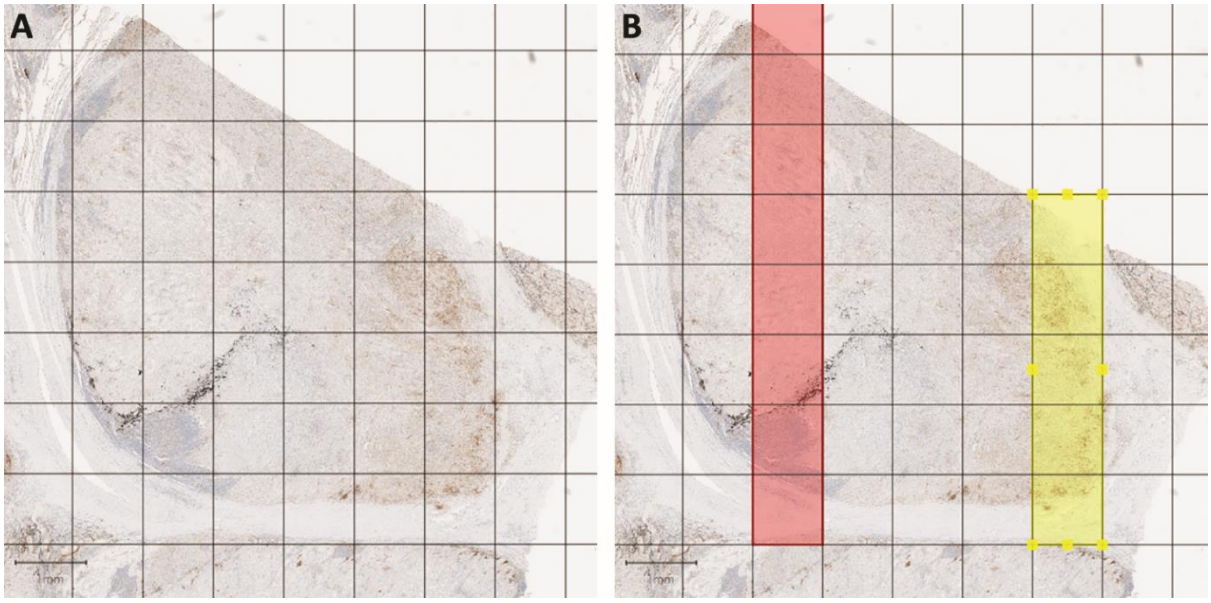


Fig 6.2.5 Challenges of focal PD-L1 (SP263) expression in metastatic NSCLC in a lymph node. **A –** Overview of the PD-L1 expression in the metastatic deposit (TPS 30%). **B** – Example of potential DCBs that may underscore (Red - <1% TPS) or over score (Yellow - ≥50% TPS) compared to whole metastasis.

6.2.5 Inter-tumoural heterogeneity – Primary vs metastases

All 107 cases had PD-L1 expression assessed in primary tumour and matched nodal metastatic deposits. In cases with multiple metastases, all metastatic tissue was included to generate an overall TPS score. In 50 cases (47%) there was no difference in TPS. Of the remaining 57 cases (53%) with a difference in TPS, 38 cases (36%) had a change of ≥10% TPS and 7 cases (6.5%) a change of ≥50% TPS. In all available tested tissue, three cases had no PD-L1 expression in the primary tumour, but were PD-L1 positive in their metastases, and 13 cases went from PD-L1 positive primary to entirely negative in their metastases, examples shown in Fig 6.2.6. 22 cases (21%) were placed in different clinical categories (<1%, 1-49%, ≥50% TPS) depending on their primary or metastatic tissue, with six cases going to a higher, and 16 cases to a lower category if scoring metastatic tissue. There is no significant difference in the overall mean TPS between primary and secondary tumours (mean 28 vs 27 p=0.403) but significantly more metastases were placed in the <1% clinical group (47 vs 37, p<0.0001), and fewer in the 1-49% group (20 vs 37) and ≥50% group (31 vs 33) in the lymph node metastases. Data is shown in Table 6.2.6.

Case	1ry TPS	2ry TPS	TPS change	1ry Clinical Group	2ry Clinical Group	Clinical change?	1ry PD-L1 Pattern	2ry PD-L1 Pattern
2	0	0	0	<1%	<1%	No	Negative	Negative
4	0	0	0	<1%	<1%	No	Negative	Negative
9	0	0	0	<1%	<1%	No	Negative	Negative
10	0	0	0	<1%	<1%	No	Negative	Negative
15	0	0	0	<1%	<1%	No	Negative	Negative
17	0	0	0	<1%	<1%	No	Negative	Negative
18	0	0	0	<1%	<1%	No	Negative	Negative
21	0	0	0	<1%	<1%	No	Negative	Negative
23	0	0	0	<1%	<1%	No	Negative	Negative
25	0	0	0	<1%	<1%	No	Negative	Negative
30	0	0	0	<1%	<1%	No	Negative	Negative
32	0	0	0	<1%	<1%	No	Negative	Negative
33	0	0	0	<1%	<1%	No	Negative	Negative
34	0	0	0	<1%	<1%	No	Negative	Negative
35	0	0	0	<1%	<1%	No	Negative	Negative
37	0	95	95	<1%	≥50%	Yes	Negative	Scattered
38	0	0	0	<1%	<1%	No	Negative	Negative
43	0	0	0	<1%	<1%	No	Negative	Negative
58	0	0	0	<1%	<1%	No	Negative	Negative
71	0	0	0	<1%	<1%	No	Negative	Negative
72	0	0	0	<1%	<1%	No	Negative	Negative
73	0	0	0	<1%	<1%	No	Negative	Negative
74	0	0	0	<1%	<1%	No	Negative	Negative
77	0	0	0	<1%	<1%	No	Negative	Negative
78	0	0	0	<1%	<1%	No	Negative	Negative
83	0	0	0	<1%	<1%	No	Negative	Negative
84	0	0	0	<1%	<1%	No	Negative	Negative
88	0	0	0	<1%	<1%	No	Negative	Negative
89	0	0	0	<1%	<1%	No	Negative	Negative
90	0	0	0	<1%	<1%	No	Negative	Negative
92	0	0	0	<1%	<1%	No	Negative	Negative
94	0	0	0	<1%	<1%	No	Negative	Negative
100	0	0	0	<1%	<1%	No	Negative	Negative
103	0	0	0	<1%	<1%	No	Negative	Negative
105	0	1	1	<1%	1-49%	Yes	Negative	Scattered
107	0	1	1	<1%	1-49%	Yes	Negative	Scattered
3	1	0	-1	1-49%	<1%	Yes	Scattered	Negative
12	1	1	0	1-49%	1-49%	No	Scattered	Scattered
13	1	1	0	1-49%	1-49%	No	Focal	Focal
27	1	1	0	1-49%	1-49%	No	Scattered	Scattered
28	1	2	1	1-49%	1-49%	No	Scattered	Focal
45	1	0	-1	1-49%	<1%	Yes	Focal	Negative
57	1	1	0	1-49%	1-49%	No	Scattered	Scattered
65	1	0	-1	1-49%	<1%	Yes	Scattered	Negative
66	1	1	0	1-49%	1-49%	No	Scattered	Scattered
79	1	5	4	1-49%	1-49%	No	Scattered	Scattered
81	1	1	0	1-49%	1-49%	No	Scattered	Scattered

93	1	1	0	1-49%	1-49%	No	Focal	Focal
95	1	5	4	1-49%	1-49%	No	Scattered	Scattered
96	1	1	0	1-49%	1-49%	No	Scattered	Scattered
29	2	1	-1	1-49%	1-49%	No	Scattered	Scattered
62	2	0	-2	1-49%	<1%	Yes	Scattered	Negative
1	5	0	-5	1-49%	<1%	Yes	Scattered	Negative
5	5	5	0	1-49%	1-49%	No	Scattered	Focal
19	5	1	-4	1-49%	1-49%	No	Scattered	Scattered
22	5	0	-5	1-49%	<1%	Yes	Focal	Negative
42	5	0	-5	1-49%	<1%	Yes	Focal	Negative
46	5	5	0	1-49%	1-49%	No	Scattered	Scattered
87	5	15	10	1-49%	1-49%	No	Focal	Focal
24	10	0	-10	1-49%	<1%	Yes	Focal	Negative
31	10	90	80	1-49%	≥50%	Yes	Focal	Uniform
39	10	20	10	1-49%	1-49%	No	Scattered	Scattered
50	10	5	-5	1-49%	1-49%	No	Scattered	Scattered
59	10	0	-10	1-49%	<1%	Yes	Focal	Negative
85	10	45	35	1-49%	1-49%	No	Focal	Focal
53	15	0	-15	1-49%	<1%	Yes	Scattered	Negative
8	20	20	0	1-49%	1-49%	No	Scattered	Scattered
52	20	1	-19	1-49%	1-49%	No	Focal	Focal
69	25	5	-20	1-49%	1-49%	No	Focal	Scattered
7	30	30	0	1-49%	1-49%	No	Focal	Focal
11	30	70	40	1-49%	≥50%	Yes	Focal	Focal
64	35	80	45	1-49%	≥50%	Yes	Focal	Focal
41	40	0	-40	≥50%	<1%	Yes	Focal	Negative
75	40	10	-30	1-49%	1-49%	No	Focal	Scattered
80	40	30	-10	1-49%	1-49%	No	Focal	Focal
26	50	50	0	≥50%	≥50%	No	Focal	Focal
36	50	20	-30	≥50%	1-49%	Yes	Uniform	Scattered
44	50	70	20	≥50%	≥50%	No	Focal	Uniform
60	50	90	40	≥50%	≥50%	No	Focal	Uniform
82	50	60	10	≥50%	≥50%	No	Scattered	Scattered
55	60	80	20	≥50%	≥50%	No	Uniform	Uniform
104	60	60	0	≥50%	≥50%	No	Focal	Uniform
14	70	95	25	≥50%	≥50%	No	Uniform	Uniform
16	80	60	-20	≥50%	≥50%	No	Uniform	Uniform
20	80	80	0	≥50%	≥50%	No	Uniform	Uniform
47	80	100	20	≥50%	≥50%	No	Uniform	Uniform
51	80	5	-75	≥50%	1-49%	Yes	Uniform	Scattered
99	80	90	10	≥50%	≥50%	No	Uniform	Uniform
40	85	100	15	≥50%	≥50%	No	Uniform	Uniform
48	85	0	-85	≥50%	<1%	Yes	Uniform	Uniform
91	85	95	10	≥50%	≥50%	No	Uniform	Uniform
49	90	100	10	≥50%	≥50%	No	Uniform	Uniform
54	90	100	10	≥50%	≥50%	No	Uniform	Uniform
56	90	100	10	≥50%	≥50%	No	Uniform	Uniform
63	90	70	-20	≥50%	≥50%	No	Uniform	Uniform
67	90	100	10	≥50%	≥50%	No	Uniform	Uniform
86	90	60	-30	≥50%	≥50%	No	Uniform	Uniform

98	90	40	-50	≥50%	1-49%	Yes	Uniform	Focal
102	90	100	10	≥50%	≥50%	No	Uniform	Uniform
6	95	0	-95	≥50%	<1%	Yes	Uniform	Negative
68	95	100	5	≥50%	≥50%	No	Uniform	Uniform
70	95	60	-35	≥50%	≥50%	No	Uniform	Uniform
76	95	80	-15	≥50%	≥50%	No	Uniform	Uniform
97	95	95	0	≥50%	≥50%	No	Uniform	Uniform
61	100	90	-10	≥50%	≥50%	No	Uniform	Uniform
101	100	100	0	≥50%	≥50%	No	Uniform	Uniform
106	100	100	0	≥50%	≥50%	No	Uniform	Uniform

Table 6.2.6 Inter-tumoural heterogeneity of PD-L1 expression in 107 NSCLCs showing PD-L1 TPSs from primary tumour and matched metastases in regional lymph nodes. 22 cases (21%) had sufficient inter-tumoural variation of PD-L1 expression sufficient to classify primary and metastatic tissue into different clinical categories of negative, weak or positive (<1%, 1-49%, ≥50% TPS). Predominant patterns of PD-L1 expression heterogeneity is summarised for each sample. In addition to loss or gain of TPS, 5 cases (5%) demonstrated different predominance of focal or scattered staining between matched tumour samples.

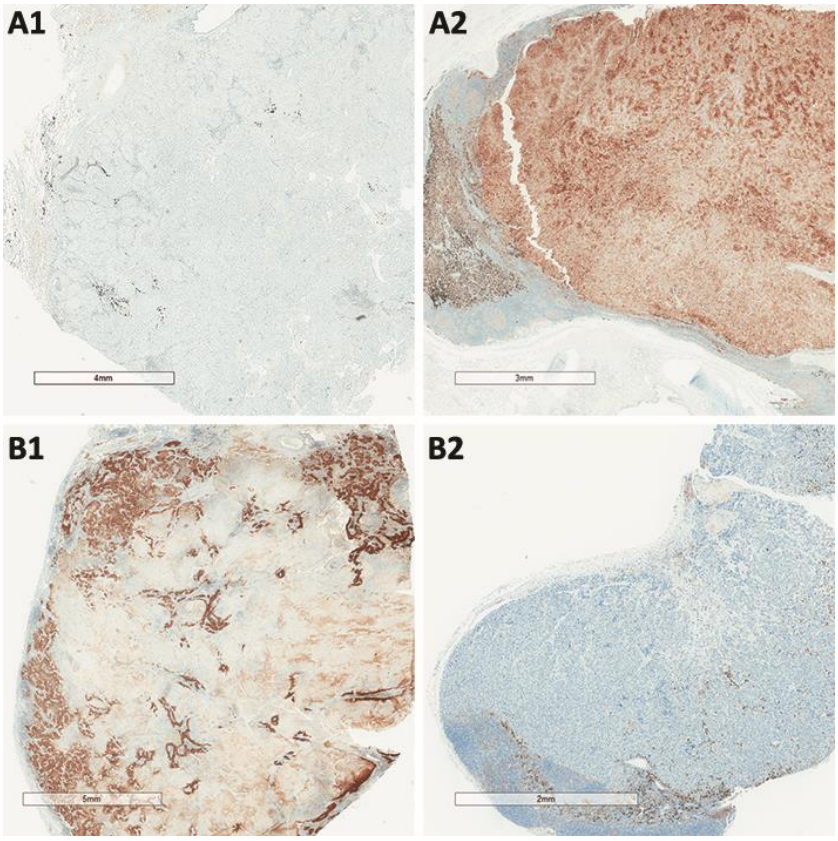


Fig 6.2.6 Examples of dramatic inter-tumoural heterogeneity of PD-L1 (SP263) expression between primary NSCLC and matched metastatic deposits. **A1 –** Primary NSCLC (TPS 0%) with matched (**A2**) metastases (TPS 95%). **B1** - Primary NSCLC (TPS 95%) with matched (**B2**) metastases (TPS 0%).

6.2.6 Inter-tumoural heterogeneity – Metastases vs metastases

35 cases had multiple lymph nodes with metastatic deposits suitable for testing PD-L1, with a total of 85 lymph nodes tested (range 2-5 per case). Of these, 29 cases (83%) had no difference in TPS between metastatic deposits. Of the 6 cases with variation between metastatic deposits, four (11%) had TPS changes sufficient to change clinical groups. Two cases had small foci of PD-L1 expression (<1%). Of the remaining four cases, TPS changes ranged from 1% to 95%, with one case having five lymph nodes with different TPS score for each metastasis. Heterogeneity was variable: for example, two lymph nodes from the same station having vastly differing TPSs (Fig 6.2.7) and significant intra-tumoural heterogeneity across a lymph node metastases (Fig 6.2.8). Data is summarised in Table 6.2.7.

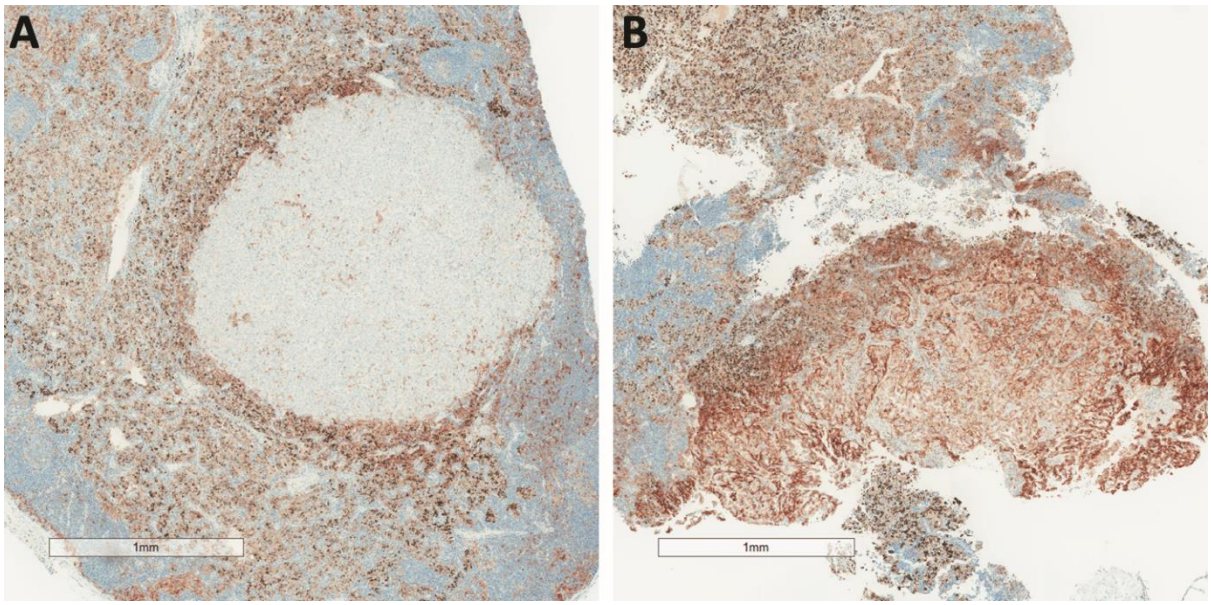


Fig 6.2.7 Examples of dramatic inter-tumoural heterogeneity of PD-L1 (SP263) expression between matched metastatic NSCLC deposits in regional lymph nodes. **A –** Lymph node metastases (TPS 5%) **B** – Matched lymph node metastases (TPS 100%)

6.2.7 Heterogeneity and pathological/clinical features

All 107 cases were assessed for their patterns of heterogeneity. PD-L1 expression is minimally heterogeneous at the extremes of TPSs, either being negative (37 cases) or uniformly positive (all cases ≥70% TPS, 27 cases). Of the remaining 43 cases, intra-tumoural heterogeneity was either classified as scattered (example Fig 6.2.2 A,B) or focal/multifocal (Fig 6.2.2 C,D). Cases with both elements were grouped based on the predominant pattern. As such, 21 cases were deemed to have their PD-L1 expression in the primary tumour as predominantly focal, and 22 cases as having scattered PD-L1 expression. Focal expression had a higher average TPS (26 vs 6 $p<0.0001$) but there

was no significant difference in clinical categorisation (16 vs 21 1-49% TPS, 5 vs 1 ≥50% TPS, p=0.068). There was no difference between focal or scattered heterogeneity for morphology (p=0.287), T stage (p=0.681), age (63 vs 68 p=0.630), pleural status (p=0.427), necrosis (p=0.454), gender (p=0.284) or smoking status (p=0.770). Survival analysis by Kaplan-Meier Log-Rank testing found no significant difference in outcome between focal and scattered patterns of PD-L1 expression (45 vs 42 months OS p=0.780).

Case	Average 2ry TPS	N1	N2	Nodes tested	TPS changes (Values)
6	0	1	1	2	N
24	0	1	1	2	N
32	0	2	1	2	N
35	0	3	4	2	N
38	0	3	1	2	N
42	0	2	1	2	N
43	0	1	1	2	N
62	0	4	0	4	N
65	0	1	1	2	N
78	0	5	1	2	N
83	0	3	1	4	N
88	0	4	2	2	N
94	0	2	1	2	Y – (1%/0%)
100	0	1	1	2	Y – (<1%/0%)
103	0	1	1	2	N
29	1	1	1	2	N
66	1	2	1	2	N
107	1	1	1	2	Y – (<1%/0%)
28	2	1	1	2	N
46	5	2	0	2	N
50	5	2	1	2	N
51	5	1	3	4	Y – (5%/5%/5%/100%)
69	5	2	1	2	N
36	20	4	1	5	N
39	20	3	1	2	N
80	30	2	1	2	N
98	40	2	2	5	Y – (70%/20%/50%/30%/95%)
85	45	2	1	2	Y – (50%/15%)
26	50	3	0	3	N
16	60	2	1	2	N
11	70	1	1	2	N
31	90	3	2	5	N
37	95	2	0	2	N
101	100	1	1	2	N
106	100	1	1	2	N

Table 6.2.7 Inter-tumoural heterogeneity of PD-L1 expression in 35 NSCLCs showing PD-L1 TPSs between matched metastases in regional lymph nodes. 4 cases (11%) had sufficient inter-tumoural variation of PD-L1 expression sufficient to classify primary and metastatic tissue into different clinical categories of negative, weak or positive (<1%, 1-49%, ≥50% TPS).

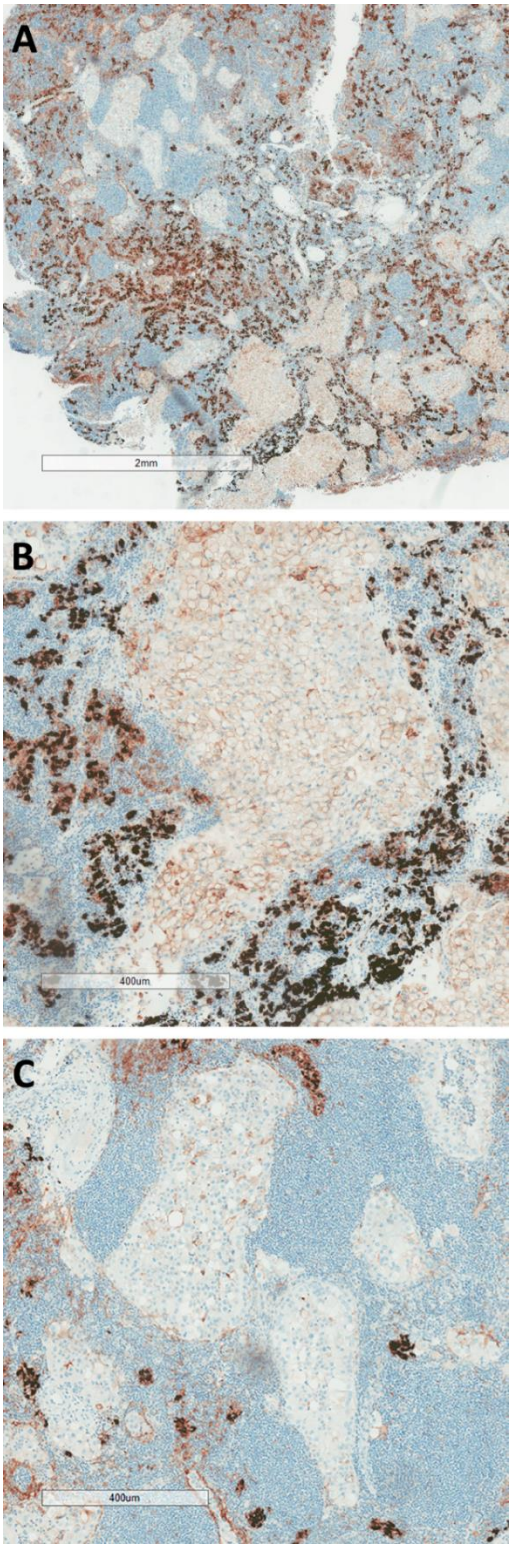


Fig 6.2.8 Examples of dramatic intra-tumoural heterogeneity of PD-L1 (SP263) expression in metastatic NSCLC within a single lymph node. **A –** Low powered view of a single lymph node with multiple metastatic NSCLC deposits **B** – High power view of metastases (TPS 80%) **C –** High power view of metastases (TPS 1%)

6.3 Discussion

6.3.0 Extensive sampling and PD-L1 heterogeneity

PD-L1 expression by IHC quantitatively interpreted to provide a TPS remains the only clinically-validated method of predicting response to PD-1/PD-L1 IMs in NSCLC, with greater levels of PD-L1 expression broadly associated with improved responses.[47, 48, 254, 255] However, only the tissue sampled can be tested for PD-L1, and often these are small biopsies or cytology specimens that contain only a very small proportion of the total tumour burden. Intra-tumoural heterogeneity therefore presents itself as a problem in that small samples may return a TPS that is lower or higher than the true overall TPS of the tumour. Various other studies have studied the impact of intra-tumoural PD-L1 expression and how this might impact concordance between specimens. Studies comparing TMAs to matched larger specimens suggest that these very small samples have a high risk of providing inconsistent results, ranging in discordance between 2.6% to 66% of cases.[376, 378, 380, 382, 384, 386, 570, 572] Papers using small biopsies compared to surgical specimens also found a variable amount of concordance between the specimens, with discordance ranging from 8.6% of cases to 48%, with most studies generally finding small biopsies could misclassify patients.[388, 390-392, 568] However, studies that used whole tissue sections from surgical specimens to analyse intra-tumoural heterogeneity found a much lower rate of discordance with the authors generally concluding that these larger pieces of tumour, if available, would in most cases provide sufficient tissue to overcome intra-tumoural heterogeneity of PD-L1 expression, with one study suggesting that variation within a block was generally greater than between blocks (91% vs 9% field of view variance).[379, 393] The overall conclusion from these papers is that more tissue provides a greater chance of achieving a TPS truly representative of the whole tumour. This study agrees with these findings: intra-tumoural heterogeneity is a considerable problem with significant variation seen within 1cm² areas for some tumours, and no variation at all seen for others. However, when using as much as a whole section from a surgical specimen, the difference between two blocks is typically small, and rarely enough to change clinical categories, with typically more variation occurring between 1cm² areas within the same section than seen between blocks; suggesting a whole section from a surgical specimen is indeed sufficient to overcome intra-tumoural heterogeneity in the majority of cases. However, as this is not plausible in any patient pre-surgery (or not undergoing surgery), the question of how representative biopsies are is highly pertinent. In keeping with other studies, the more core biopsies taken, generally the more accurate the TPS is. However, this is not universally true, and a small number of cases saw minimal, or no improvement despite the use of more simulated samples. These cases were generally associated with very focal expression of PD-L1. The presence of very focal PD-L1 expression, therefore, appears to be more problematic than scattered PD-L1 expression

heterogeneity. For example, if a relatively small T1 tumour with a maximum diameter of 20mm can be assumed to be ~20mm³ in total dimension, and if a generous core biopsy is equivalent to 2x10x1mm, then this tumour has the theoretical equivalent of 400 potential core biopsies. The difference between one or two small biopsies seems inconsequential against many hundreds, if not thousands of potential sites, but if the TPS is a relatively modest 10% and PD-L1 expression is equally scattered across the tumour, there is actually a high chance of sampling an area with at least some PD-L1 expression. Although in this theoretical scenario any given core biopsy might not have a TPS of precisely 10%, an expression level of at least 1% TPS, and therefore remaining in the same clinical category, remains more likely to occur than not. However, if PD-L1 expression is very focally expressed, the majority of core biopsies would return a TPS of 0%, and therefore incorrectly classify this theoretical tumour. In addition the chances of hitting any given tumour area is partially determined by anatomical location: if the tumour is bronchial and central, the most likely route of access is via EBUS, and therefore sampling of the tumour is more likely to occur at the region closest to the bronchus. If the expression is equally scattered, this may make little difference, but if the focal area of PD-L1 expression is at an area distant to the bronchus, then a core biopsy may fail to sample the area of PD-L1 positive tumour cells.

Despite this, the results from this study suggest a dramatic improvement in successful classification of tumours for PD-L1 if increasing from just one to two biopsies, with the majority of cases correctly classified with 3 biopsies, after which diminishing returns are achieved by the use of four and five biopsies. A previous study has suggested at least four biopsies are required for accurate PD-L1 expression analysis, though their approach was TMA construction using much smaller 0.6mm diameter cores.[386] Random sampling via DFNA performed modestly better than the equivalent quantity of sampling via DCBs, which likely reflects the increased chance of hitting PD-L1 expressing areas even in focally positive areas, but again struggled with specific cases, and in both instances the areas of biggest difficulty were, unsurprisingly, in cases very close to the clinical cut-offs of ≥1% or ≥50%, with misclassifications of tumours into higher or lower groups both occurring, a not uncommon problem seen in other studies.[382, 386, 388, 391, 392]

These results are mirrored when considering the intra-tumoural heterogeneity of the lymph node metastases. However, overall variation was significantly less in metastatic tissue than primary tissue, and, most likely as a reflection of the smaller size of metastatic deposits, less tissue was required to accurately score PD-L1. Both DCBs and DFNAs were superior at correctly classifying metastatic tumour PD-L1 compared to the equivalent in primary NSCLC specimens, and DFNAs was significantly better than DCBs within metastatic tissue. As the rationale behind DFNAs was to simulate the more random sampling that cytology specimens achieve, this may suggest that cytology samples acquired

from lymph node metastases might be more accurate than core biopsies of the same specimens. As discussed in Chapter 3 (Pre-analytics), the use of cytology specimens in terms of fixation and other pre-analytical factors should not been seen as a barrier to their use, and can be considered equivalent to histology samples,[302, 307-309, 311-314, 471, 472, 573, 574] and this data suggests there may be further benefits currently underappreciated.

6.3.1 Inter-tumoural heterogeneity: hinderer and helper

Metastatic deposits in lymph nodes therefore provide an attractive target for ascertaining PD-L1 expression in NSCLC cases, but the issue of inter-tumoural variation between primary tumour and metastatic sites, including both nodal disease and distant metastases has been seen in a number of other studies, ranging from 9% to 40% discordance, suggesting that they cannot be considered equivalent to primary tumours.[306, 355, 396, 400, 401, 403, 542, 570] This study echoes these findings in that over half of cases have some discrepancy in TPS between primary and metastatic tissue, with 22% of cases having differences sufficient to place the primary and metastatic samples in different clinical categories. In addition to being a significant proportion of cases potentially affected by this, it was interesting to note that PD-L1 positive primary tumours could become entirely negative in their metastases, and that the inverse was also true. Both could be explained by focal PD-L1 expression simply being missed by sampling, but in several cases PD-L1 expression went from nearly 100% to 0% and vice versa, with several cases demonstrating profound differences of PD-L1 expression between metastatic deposits, in one case, even on the same slide, thus suggesting this is a genuine observation in at least some instances. Immune escape of NSCLC is thought to require, in addition to PD-L1 expression, specific conditions within the TME, such as the proximity of CD8+ cytotoxic T-cell lymphocytes and a non-suppressive immune environment.[75, 545, 575] It is perhaps not surprising therefore that PD-L1 expression varies between a primary NSCLC and its nodal metastases; the microenvironment in the lung is very different from that in a lymph node.[576, 577]

Apparent *de novo* expression of PD-L1 in metastatic cells from an apparently negative primary tumour is harder to explain, and may more easily be dismissed as an error of sampling, though the acquisition of survival traits and presence of differing immune markers in metastatic cells has been described before[577-580] and certain forms of PD-L1, such as soluble or highly glycosylated forms of PD-L1, which are not detectable by IHC, may play a clinically relevant role currently not fully characterised, and result in occult PD-L1 expression not identifiable by standard IHC assays.[360, 366, 581]

Regardless of the underlying mechanism, primary tissue and metastatic tissue cannot be uniformly considered to be equivalent when testing for PD-L1 expression. In addition, whilst the phenomenon of variable PD-L1 expression between metastatic deposits was relatively uncommon in this cohort,

micro-environmental differences between evolving NSCLC metastatic deposits has been previously described,[580] which could potentially account for some of the inter-tumoural heterogeneity observed between metastatic sites. As such, any single site, no matter how thoroughly sampled, should not be considered representative of a patients total tumour burden when deciding IM therapy, and this study would strongly suggest thorough sampling of multiple sites in patients with metastatic disease to ensure an accurate reflection of PD-L1 expression can be established.

However, rather than considering the lack of PD-L1 expression concordance between primary tumour and metastases as a setback, it should perhaps be seen as an opportunity to re-evaluate the rationale for testing certain specimens and how their result should guide treatment. Neoadjuvant treatment of NSCLC by IMs is being assessed in current clinical trials[55, 56] and extensive sampling of the primary tumour pre-surgery would seem prudent . Metastasis, however, is a reflection of evolution of the tumour, and it would seem reasonable to assume that the most advanced and potentially successful component of a disseminated tumour would be the most informative in terms of targeting for biopsy.[545, 579, 582, 583] If the primary tumour is negative for PD-L1, but the metastases are positive, the latter would intuitively provide a better rationale for making a decision to treat with PD-1/PD-L1 IMs. As such, in patients with metastatic disease, it might be considered a reasonable approach to preferentially target metastases to ascertain PD-L1 expression, particularly as even cytology specimens from lymph nodes are demonstrably adequate for molecular profiling of actionable targets[308, 438-440] and this study would suggest they are superior at overcoming intra-tumoural heterogeneity issues. In addition, multiple samples from different nodes would be advisable as inter-tumoural heterogeneity between metastatic deposits is also observed, albeit in a minority of cases.

6.3.2 Novel approaches to quantifying heterogeneity

Whilst the use of digital images to simulate core biopsies and cytology sampling has advantages, specifically in that many thousands of replicates could be performed without the loss of additional tissue, a number of limitations must be considered. Some limitations to this approach mimic the limitations in the reality of sampling and are not easily overcome: a FFPE section is a 2D representation of a 3D structure and only samples a relatively small part of the total tumour. Biopsies of tumours near important structures (e.g. the pericardium) or of tumours limited by access due to anatomical location will not typically present the sampling clinician with the option to thoroughly sample every area of the tumour or every tumour deposit.[584-586] In addition, the synthesis of TPSs taken from different mm^2 areas results in a mathematical average, which may put a biopsy in a different category to the true TPS by only a very small margin; for example if the true TPS is 50%,

and the biopsy average is 49%, these will be placed in different categories, whereas in reality a pathologist may be more likely to round this up. As a result, this novel digital simulation approach is probably unfair on a number of cases, but I have made no attempt to manually round-up any of the figures, on the basis that this would not be good practice, and also highlights the difficult nature of cases that are around clinical cut-offs. In addition, DCBs and DFNAs have been simulated from a limited area of the tumour (1cm² area), which in a number of cases has an average TPS different to the whole section TPS, which may over or underestimate the accuracy of DCBs and DFNAs in these instances.

This study focuses purely on intra- and inter-tumoural heterogeneity, and does not address the impact of temporal or iatrogenic heterogeneity. Whilst this is a limitation, it is also a strength: other studies that have attempted to quantify PD-L1 expression heterogeneity include studies that include metastatic samples that are acquired weeks or months subsequent to the primary tissue or have had treatment in between paired sample collections.[303, 312, 313, 396, 397, 400] The advantage of this cohort is that they received no neoadjuvant therapy, and all lymph nodes metastases were taken at the time of surgical resection, meaning the scale of intra- and inter-tumoural heterogeneity could be studied without the impact of temporal or iatrogenic heterogeneity. Whilst this study focuses purely on NSCLC, PD-L1 expression heterogeneity in also seen in other tumour types, with intra-tumoural, inter-tumoural or spatial heterogeneity seen in breast cancer,[406, 407] malignant mesothelioma,[408] HNCC[409] CRC,[410] lymphoma[411] and various tumours that have metastasised specifically to the lung.[412, 413] Whilst the data is not directly transferable, the implications of this data in regards to adequate and widespread sampling would seem a reasonable precaution to apply to any tumour type when considering PD-L1 expression.

Finally, whilst an attempt to classify PD-L1 expression heterogeneity by qualitative features was made, there is no clear pattern or apparent underlying reason for PD-L1 heterogeneity. Grouping by heterogeneity type was associated with no other pathological or any clinical traits, including by the separation into ADC or SCC morphologies. PD-L1 heterogeneity is unlikely, however, to be a randomly driven phenomenon; merely that its presence represents underlying aspects of the relationship of tumour cells and the immune response within the TME that is not fully understood. It is important with PD-L1 IHC, as with any single biomarker, to realise the limitations of what a single target can provide in terms of biological and clinical data. Whilst the optimisation of PD-L1 IHC can be achieved to ensure consistent results that minimise the impact of heterogeneity, additional biomarkers within the context of the TME are required to both understand the underlying reasons for heterogeneous expression, and to improve the predictive power of PD-L1 IHC. These will be explored in the following chapter.

Chapter 7 – The tumour microenvironment

7.0 Introduction

In the previous chapter PD-L1 expression heterogeneity was explored in depth, but despite being able to quantify intra- and inter-tumoural heterogeneity and suggest approaches that would minimise the clinical impact when sampling tumours, the underlying mechanism of PD-L1 expression heterogeneity remains elusive. Further interrogation of the TME in an attempt to understand this heterogeneity will be the focus of this chapter.

7.0.1 PD-L1 and the TME

The main mechanism by which PD-1/PD-L1 IMs are proposed to work is via disruption of the PD-L1-PD-1 axis; PD-L1 expressed on tumour cells are able to bind to their receptor, PD-1, expressed on cytotoxic T-cells and exert an immunosuppressive effect on these T-cells, resulting in the inhibition of their cytotoxic function and decrease in cytokine production, eventually culminating in cell death.[60, 587, 588] It has been demonstrated that tumour cells can be poor antigen presenters, and so the recruitment of professional APCs to express PD-L1 are necessary to suppress cytotoxic T-cell function in a similar fashion and induce an immunosuppressive TME.[60, 589] By both direct inhibition of cytotoxic T-cells and the generation of a more immunosuppressive TME, the tumour cells are able to achieve immune escape. Targeting this interaction with PD-1/PD-L1 IMs will revert the TME to more active, anti-tumour immune response, and provide a better outcome for patients.[60, 588]

PD-L1 expression within the TME is therefore likely to have a survival benefit for the tumour cells that express it, particularly in the context of an immune response targeted against them. This reactive expression of PD-L1 is regulated by many factors within the TME including IFN-γ, TNF-α, ILs, exosomes and other immune-related cytokines and messengers.[60, 194, 210, 590] Not all PD-L1 expression is reactive, however, with some expression driven purely by chance as a result of genomic alterations within the tumour cells; so-called constitutive expression.[564] NSCLCs have some of the highest levels of mutations relative to most other tumour types, and thus have an increased rate of expressing proteins, antigens and neo-antigens by chance (Fig 7.0.0)[591, 592] Constitutive expression of PD-L1 may therefore occur even in the context of an immune desert, and as there is no active immune response targeted against the tumour cells, PD-L1 may be present, but it does not necessarily convey a survival benefit to the tumour cells.[593] Furthermore, there are multiple possible pathways of immune escape: in addition to the myriad ways in which immunosuppressive TMEs can be altered by the cancerous cells, other direct immune checkpoints, such as IDO1, VISTA, TIM3 and others may be present.[75]

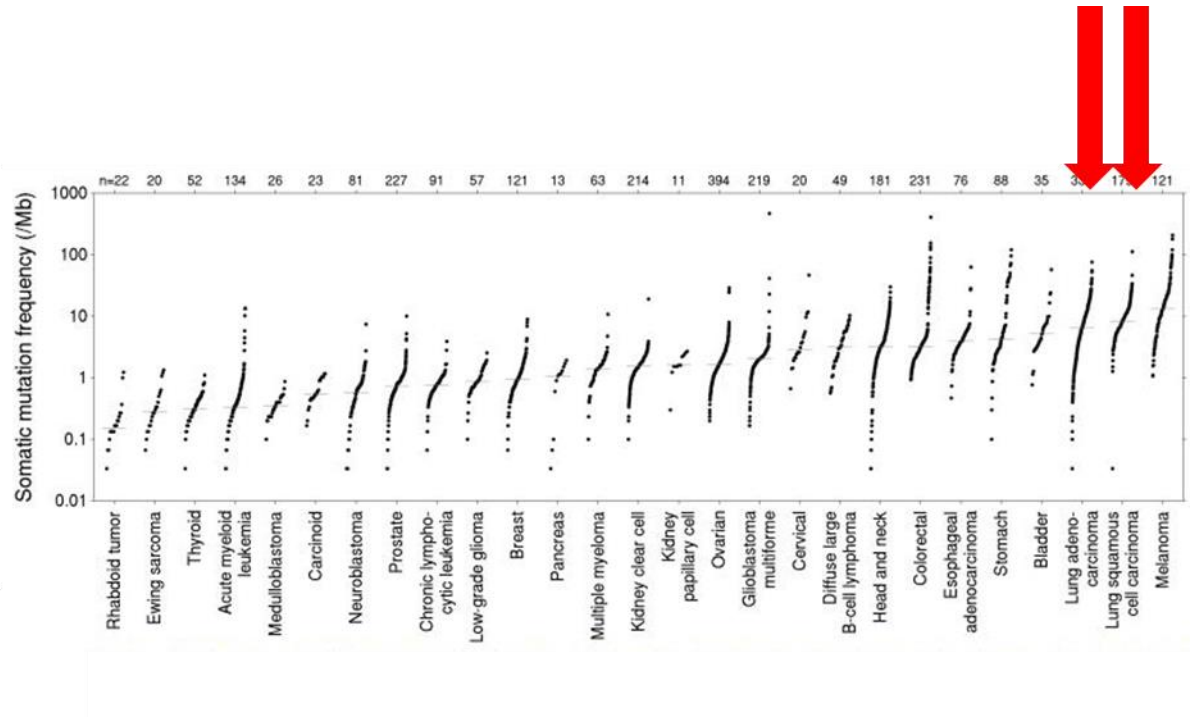


Fig 7.0.0. Somatic mutation frequencies observed in exomes of 27 cancer types. Lung ADCs and lung SCCs (red arrows) have two of the highest rates of mutations. Modified from Lawrence *et al*. 2013.[592]

PD-L1 expression alone is therefore unlikely to be a perfect predictor of response to PD-1/PD-L1 IM therapy, and indeed its presence has been associated with superior response in only some of the major clinical trials.[46, 47, 49] Therefore the presence, type and quantity of an immune response targeted against the cancerous cells within the TME is required to more fully understand the role of PD-L1 expressed by tumour cells.

7.0.2 Types of immune TME

Multiple different attempts to classify the TME based on immune activity have been made previously, with efforts including gene-signature based approaches,[594-596] and the presence of TILs within the TME.[597-599] Defining the types of TME remains a challenge, however, with multiple different definitions depending on the approach used. For example, Immunoscore, used in colorectal cancers, provides a sliding scale of prognostic information depending on two T-cell markers and their location within the tumour.[597] Other approaches such as X-cell splits TME types based on clusters of gene expression profiles instead.[594] TMEs can also be split based on the spatial context of TILs alone; immune desert (no TILs), immune excluded (TILs are present but cannot infiltrate the tumour) and immune infiltrated (TILs infiltrating the tumour epithelium), which can all be demonstrated with simple H&E (Fig 7.1.0). Further sub classifications of TMEs exist, such as separating immune

infiltrated TMEs by the presence or absence of tertiary lymphoid structures.[600] TMEs can also be classified as 'immune hot', usually defined as the presence of cytotoxic T-cells as the major infiltrating TIL, with 'immune cold' tumours described as either having a predominance of immunosuppressive Tregs or a lack of cytotoxic T-cells infiltrating the tumour.[600, 601] However, these categories are not perfect, as they oversimplify a complex relationship, and in reality there are many overlaps and sub classifications that have yet to be elucidated,[600, 602] and indeed many tumours have a mix of the patterns shown in Fig 7.0.1, with some tumours defying easy categorisation. (Fig 7.0.2). Given the different patterns of PD-L1 expression heterogeneity and the difficulty in classifying it, it seems reasonable to assume that the heterogeneity of multiple immune factors within the TME will also defy simplistic classification.

7.0.3 T-cells within the TME

Many immune cells are important within the TME, with fibroblasts, macrophages, MDSCs and others all playing roles in determining the nature of the immune response. However, PD-L1 expression as a means of immune escape is primarily involved with the interaction of T-cells and tumour cells, so a brief overview of the main T-cell types will be given here. The chief mature effector T-cells include cytotoxic, memory and helper T-cells. CD4+ve Th cells play a variety of roles, including the assistance of cytotoxic T-cells, whilst memory T-cells largely remain circulating in order to produce long term immunity to specific antigenic signals. CD8+ve cytotoxic T-cells have been associated with improved outcome in a variety of cancers, and the impact of these cells is determined by both their tumour infiltration and their differentiation.[603, 604] Naïve CD8+ve T-cells infiltrate the TME and differentiate into effector T-cells. Cytotoxic T-cell differentiation is dependent on the tumour antigen-MHC formation, and influenced by cytokines and co-stimulatory factors from APCs, as well as other sources (such as metabolic, epigenetic, transcription factors and so forth). CD8+ve cytotoxic T-cells mostly act to produce cytokines such as IL-2, IL-12, Granzyme B and IFN-γ which promote cytotoxic functions via TNF-related apoptosis ligands, ROS and perforin pathways.[605-607] Trafficking and localisation of CD8+ve T-cells is important, and requires the matching of chemokines secreted by both tumour and non-tumour cells within the TME with the receptors on the T cells.[608, 609] For example CXCR3 on the surface of CD8+ve T-cells can bind to CXCL9, 10 and 11 which are highly secreted by many solid tumours, with corresponding low levels of these chemokines resulting in low levels of CD8+ve T-cells within the TME.[609, 610] Vasculature also plays an important role in T-cell trafficking; aberrant vasculature and poor angiogenesis, a feature often seen in TMEs, has been shown to result in fewer CD8+ve T-cells infiltrating the tumour.[611]

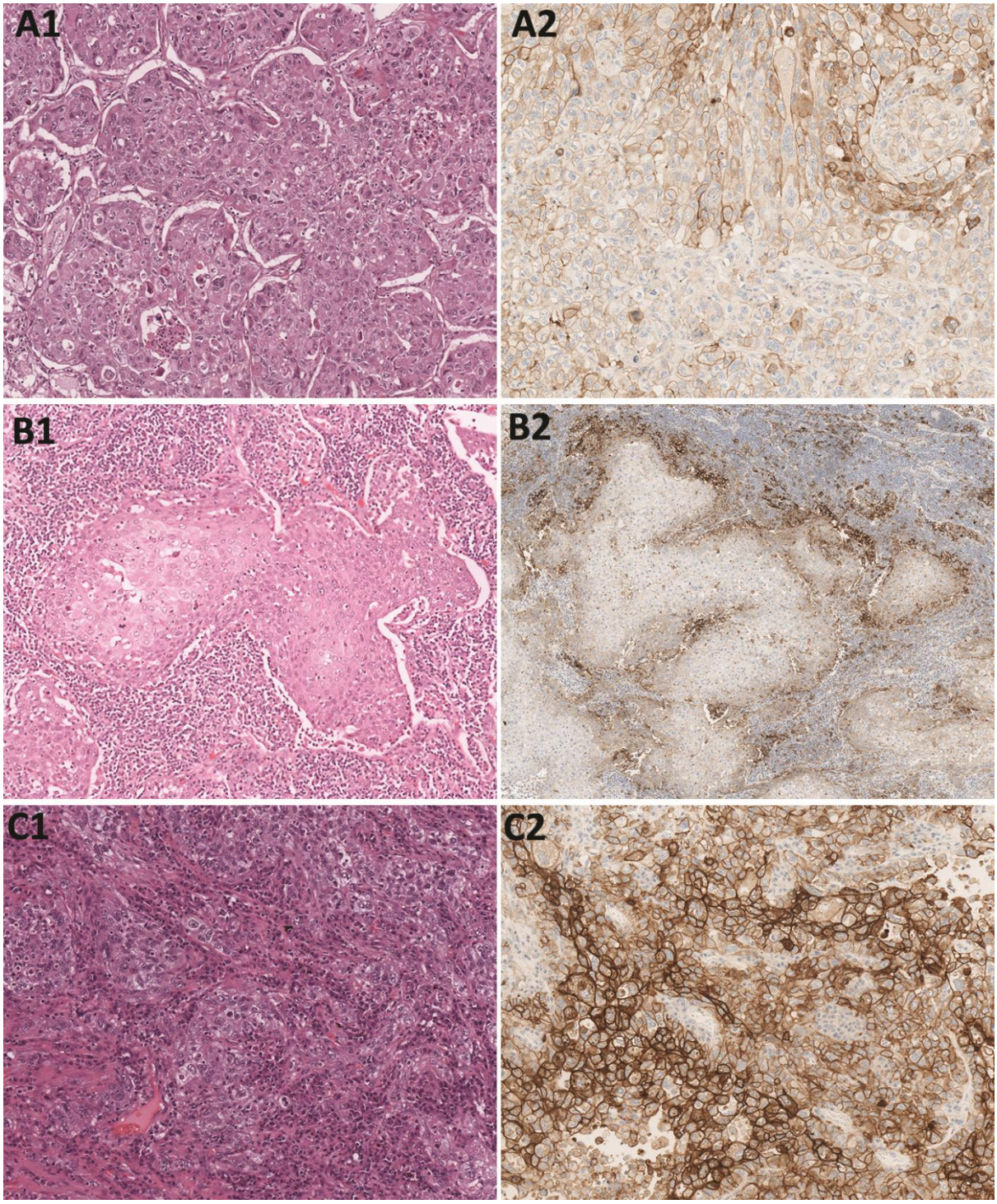


Fig 7.0.1. Different TMEs by the pattern of TILs by H&E with corresponding PD-L1 expression (SP263) for each tissue section. **A** – Immune desert. **B** – Immune excluded. **C** – Immune infiltrated.

Once T-cells have been successfully trafficked, they need to be activated. This is a three step process: interaction between TCR and antigen-MHC complex, the delivery of co-regulatory signals, and stimulation from extracellular cytokines. Once activated they are differentiated into an early CD8+ve state, and then finally differentiated into terminal effector CD8+ve T-cells. The second stage of activation, delivery of co-regulatory signals, is the key step for PD-L1, as it acts as a co-inhibitory signal. PD-1, expressed on T-cells and upregulated in response to TCR activation, binds to PD-L1 expressed on tumour cells and APCs, which downregulates effector cytokines produced by both Th and CD8+ve T-cells, and results in increased Treg function, as well as directly inhibiting the CD8+ve T-cells.[60, 600, 609] T-cells can become dysfunctional, and exhausted CD8+ve T-cells may express high levels of co-inhibitory receptors, including PD-1, (as well as others such as CTLA4, LAG3, TIM3) and lose their ability to produce immunostimulatory cytokines.[225, 612] Various markers are lost at various stages, (e.g. IL-2 is lost at an early stage, TNF-α at a mid-stage with IFN-γ and Granzyme B lost at a late stage).[609, 613] Thus, for example, decreasing levels of Granzyme B in the presence of high PD-1 would imply exhausted cytotoxic T-cells.

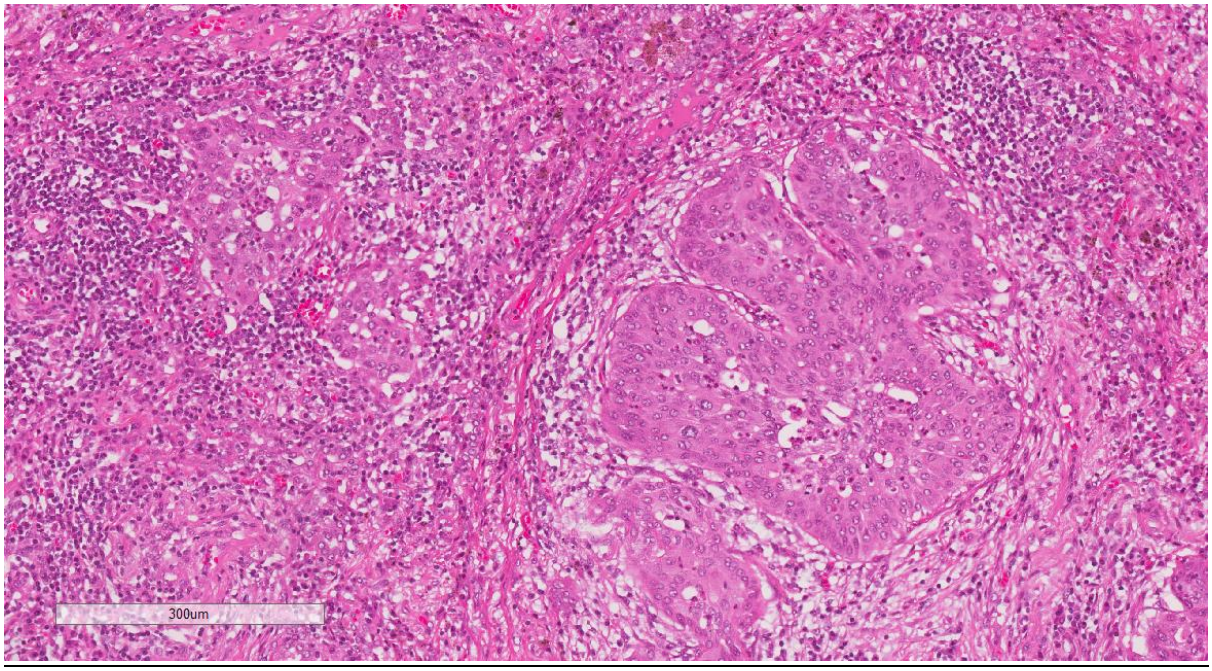


Fig 7.0.2. H&E of a squamous cell carcinoma of lung showing a difficult to classify TME. On the right is an island of tumour cells relatively free of TILs, but on the left tumour cells are admixed with TILs. The overall prevalence of an infiltrating pattern of TILs was dominant throughout this case, so was ultimately considered to be an immune infiltrating pattern of TME.

Tregs inhibit APCs, impair the interaction between APCs and T-cells, and supress the function of NK cells, and have a generally immunosuppressive pro-tumoural effect within the TME.[614, 615] Tregs are typically produced from multiple sources including the migration from lymphoid or blood supply, the differentiation of naïve T-cells driven by APC suppression and conversion of effector T-cells into Tregs by TGF-β. FoxP3 is a marker of Tregs and Treg activity, and high levels of FoxP3 is associated with generally poorer outcomes for patients and higher risks of tumour recurrence in most solid tumours, although the converse is true for colorectal cancers specifically.[616, 617] Interestingly, if FoxP3 expression is limited to Tregs within the stroma, there is a similar outcome to tumours with low overall FoxP3 levels, suggesting that infiltration of FoxP3 Tregs is just as important as it is for cytotoxic T-cells.[618]

A combination of these T-cell and immune markers may therefore determine the presence of cytotoxic T-cells, the probability of their being regulated by the PD-1/PD-L1 axis, and a measure of how exhausted/dysfunctional they may be, as well as a measure of immunosuppressive features within the TME more generally.

7.0.4 Defining the TME

In attempting to decide which markers are appropriate in ascertaining the immune features within the TME, I had to balance availability, cost, tissue use, specificity, sensitivity and simplicity to ensure a clear and precise outcome. To that end the following immune markers were chosen: CD3, CD8, PD-1, FoxP3, Granzyme B, CD68 and PD-L1. CD3 is a general marker of mature T-cells (both resting and activated T lymphocytes), but does not stain for other immune cells such as B-cells, macrophages or myeloid cells.[619] The presence of CD3 can thus be used to determine an overall level of TILs, both cytotoxic and Th and others, and the precise location of them can help ascertain which parts of the tumour are most heavily inflamed (e.g. epithelium vs non-epithelium). CD8 is a marker of cytotoxic T-cells (both naïve and fully differentiated, as well as weakly found in NK-cells) but not of T-helper cells.[620] It is therefore well established as a marker of cytotoxic lymphocyte activity. FoxP3 is a marker of Treg cells and therefore a surrogate marker of an immunosuppressive TME.[621] Granzyme B is a cytotoxic serum protease protein that participates in inducing apoptosis of target cells by cytotoxic T-cells and NK cells, and has been shown to be a marker of mature cytotoxic T-cell differentiation, activation and, inversely, exhaustion.[620, 622] CD68 is expressed on a variety of cells but functions as a marker of macrophages, and is not expressed by epithelial tumour cells (carcinomas).[620] CD68 is therefore used in this context to differentiate between tumour cells and macrophages, both of which can express PD-L1. Finally, PD-L1 and PD-1 must of course be looked for, as the chief components of the immune escape pathway regulated by the PD-1/PD-L1 axis. The

relationship of these markers in the context of each other is critical to understanding the role of PD-L1 within the TME; in terms of the quantity of each cell type, the location of each cell type, and the spatial relationship of each cell type.

7.0.5 Multiplex technologies

PD-L1 expression by IHC and H&E can begin to group tumours into simple TMEs based on the pattern of TILs with possible clinical relevance (e.g. PD-L1 expression in a TME devoid of TILS is probably constitutive PD-L1 expression and likely won't respond to IMs) but is problematic for a number of reasons: it can be hard to categorise tumours, especially when heterogeneity of the tumour may result in different TIL patterns, there may be insufficient tissue in a small biopsy to accurately assess the pattern of TILs, and specific samples, such as FNAs of LN metastases, are impossible to categorise by this approach, due to an inability to distinguish resident LN lymphocytes from tumour associated lymphocytes. In addition, the presence of TILs by H&E alone cannot tell the nature of the lymphocytes (if they are cytotoxic, Ths, Tregs and so forth) or the functional status of the lymphocytes (e.g. if they are exhausted or anergic). Using IHC to categorise the immune features of the lymphocytes is a simple and powerful tool, but with monoplex IHC, tumour heterogeneity can be such that when multiple sections for multiple markers are used, wholly different cell populations may be observed when several sections apart. An eloquent solution to this problem is the use of multiplex assays, which combine multiple markers of immune activity within a single slide, and are expressed as chromophores or fluorophores.[339, 623] Not only does this save tissue and prevent cells being too far distant, this also allows for the specific nature of individual cells to be quantified in relationship to others; a single lymphocyte can be ascertained as being CD8+ve, and therefore cytotoxic, but can also be stained for PD-1, the combined presence of which would suggest a cell prone to inhibition by PD-L1, and therefore a TME likely to respond to PD-1/PD-L1 IMs. These technologies are unlikely to be integrated into the routine clinical setting of a pathology laboratory in the foreseeable future due to the expense, technical difficulty in achieving viable assay results and increased time associated with their use, but do provide an attractive research tool to further study the relationship of PD-L1 with immune cell markers.[339, 623, 624]

The main aim of this chapter is to attempt to define TMEs based on the presence of immune markers, including PD-L1, within the TME in an attempt to elucidate associations with intra-tumoural and inter-tumoural PD-L1 expression heterogeneity, and to find prognostic data based on these categories. Other objectives include the assessment of different platforms, ranging from simple H&E to multiplex immuno-techniques, in both their ability to determine these classifications of TMEs, and their ease of use from the perspective of a routine clinical pathological assessment.

7.1 Methods

7.1.0 Selection of cohort

All patients from the LLP cohort were reviewed as per the main methods section to ensure suitable quantity and qualities of tissue, as well as to confirm a diagnosis of NSCLC. H&E sections prepared at the time of original pathological assessment, alongside the FFPE blocks, original reports and data from the LLP were used to collect pathological and clinical details for each case. A total of 92 patients were selected for use in this project, all of which were previously used in the heterogeneity studies in Chapter 6. As a result each case already had H&E and PD-L1 IHC sections prepared, as well as a primary tumour PD-L1 TPS, a secondary (metastatic) tumour PD-L1 TPS, and a sub-cohort which were used in the 'squares method' that had data on intra-tumoural heterogeneity. Each patient was assigned an arbitrary case number so that all blocks and sections could be pseudo-anonymised. A summary of these cases are shown in Table 7.1.1.

		Image analysis		-PD-L1 scores by different methods-						-Triplex Immune - Markers in tumour core					
Case num	Histology	TC area	IM area	PD-L1 1ry	PDL1 2ry	PDL1+ TE	PD-L1 WS	PD-L1 TC	Triplex PD-L1	CD8	FoxP3	CD3	PD-1	CD68	Gran. B
1	ADC	171	4	0	0	W	1	0	W	H	H	H	H	L	L
2	SCC	269	0	1	0	N	1	1	W	L	L	L	L	L	L
3	SCC	73	2	100	0	S	90	100	S	H	H	L	L	L	L
4	ADC	176	7	50	30	S	55	55	S	L	L	L	L	H	H
5	ADC	175	14	25	20	W	11	8	W	L	L	L	L	L	L
6	ADC	128	7	30	70	W	18	16	W	H	L	H	H	L	L
7	ADC	97	7	0	1	W	0	0	W	L	L	L	L	L	L
8	SCC	62	14	70	95	S	56	65	S	H	H	H	H	H	H
9	ADC	57	2	0	0	N	1	0	N	L	H	L	L	L	L
10	ADC	63	13	70	60	S	65	76	S	H	H	H	H	L	H
11	SCC	105	16	0	0	N	11	1	N	L	L	L	L	H	H
12	ADC	90	17	85	80	S	83	91	S	H	H	H	H	L	L
13	ADC	307	0	0	0	N	0	0	N	L	L	L	L	L	H
14	SCC	124	12	50	50	S	42	43	S	L	L	L	L	L	L
15	ADC	114	5	0	1	W	0	0	N	L	L	L	L	L	L
16	SCC	144	8	1	2	N	2	1	N	L	H	L	L	H	L
17	ADC	93	15	1	0	W	4	2	W	L	L	L	L	H	L
18	ADC	189	14	5	90	W	10	11	W	H	H	H	H	H	L
19	SCC	80	10	0	0	N	2	0	N	H	H	L	H	H	H
20	ADC	111	0	0	0	N	0	0	N	L	L	L	L	H	L
21	ADC	177	0	0	0	N	0	0	N	H	L	H	H	L	H
22	ADC	50	11	50	20	S	49	47	W	H	H	H	H	H	L
23	ADC	106	7	0	95	W	0	0	W	H	H	L	H	L	L

24	ADC	298	0	0	0	N	0	0	N	L	L	L	L	L	L
25	ADC	178	20	10	20	S	9	7	S	L	L	L	L	L	L
26	SCC	69	3	80	100	S	34	35	S	H	H	H	H	H	H
27	SCC	121	11	1	0	W	3	2	W	L	L	L	L	L	L
28	ADC	330	0	0	0	N	0	0	N	L	L	L	L	H	H
29	SCC	203	22	30	70	S	10	10	W	L	L	L	L	H	H
30	SCC	106	5	1	0	W	4	2	W	L	L	L	L	L	L
31	ADC	274	12	85	100	S	93	93	S	H	H	H	H	H	H
32	SCC	108	8	50	0	S	37	43	S	L	L	H	L	H	L
33	ADC	92	6	100	100	S	94	100	S	H	H	H	L	H	H
34	ADC	306	0	15	5	W	7	7	W	H	H	H	L	H	H
35	ADC	171	21	80	5	S	83	84	S	H	H	H	H	H	H
36	SCC	68	15	10	1	W	14	3	S	H	H	H	H	L	H
37	SCC	51	5	90	100	S	99	100	S	H	H	H	H	H	H
38	SCC	216	0	50	80	S	39	39	S	H	H	L	H	L	H
39	SCC	112	16	90	100	S	84	88	S	H	H	H	H	H	H
40	SCC	126	13	0	0	N	2	0	N	L	L	L	L	H	L
41	ADC	98	25	1	0	N	3	1	N	H	L	L	L	L	L
42	ADC	47	5	60	90	W	56	62	W	H	L	H	L	H	L
43	SCC	42	15	5	0	N	18	0	N	L	L	L	L	L	L
44	SCC	209	19	90	70	S	89	90	S	H	H	H	H	L	H
45	SCC	175	11	25	80	S	39	38	S	H	H	H	H	L	H
46	ADC	192	14	0	0	W	0	0	W	L	L	L	H	L	L
47	ADC	214	0	0	1	W	0	0	N	L	L	L	L	L	L
48	ADC	203	14	90	100	S	94	94	S	H	H	H	H	H	H
49	SCC	131	17	95	60	S	82	83	S	H	H	H	L	L	H
50	ADC	50	10	5	0	W	10	6	W	L	H	L	L	H	L
51	ADC	250	0	0	0	N	2	2	N	H	L	H	H	H	H
52	ADC	236	11	0	0	N	0	0	N	L	L	L	L	H	L
53	ADC	179	6	40	10	S	21	20	S	H	L	H	H	L	H
54	ADC	166	16	95	80	S	96	96	S	H	H	H	H	H	L
55	ADC	310	10	0	0	N	0	0	N	L	L	L	L	L	L
56	SCC	59	16	1	5	W	13	2	W	H	H	H	H	H	H
57	ADC	240	20	40	30	W	35	36	W	H	H	H	H	L	H
58	ADC	121	4	0	1	W	0	0	W	H	H	H	H	H	L
59	SCC	73	27	80	60	W	88	89	W	L	H	L	H	L	L
60	SCC	130	18	0	0	N	0	0	N	L	L	L	L	H	L
61	ADC	163	12	10	45	S	12	11	W	L	L	L	L	L	H
62	ADC	147	16	90	60	S	78	81	S	H	H	H	H	H	H
63	SCC	71	23	5	15	W	6	6	W	H	L	H	H	H	H
64	ADC	265	0	0	0	N	0	0	N	L	L	L	L	L	L
65	ADC	110	13	0	0	N	0	0	W	L	H	H	H	L	L
66	ADC	161	11	0	0	N	0	0	N	L	H	H	H	H	L
67	SCC	130	7	85	95	S	77	78	W	H	H	L	H	H	H
68	ADC	312	20	1	0	N	2	2	N	L	L	L	L	L	L
69	ADC	344	0	0	1	N	0	0	N	L	L	L	L	L	H
70	ADC	149	13	0	0	N	0	0	N	L	L	L	L	L	L
71	SCC	37	14	1	5	N	1	0	N	H	H	H	H	H	L
72	ADC	148	15	1	1	N	4	0	N	H	H	H	L	H	L
73	SCC	186	21	95	95	S	95	95	S	H	H	H	H	H	H

74	ADC	120	12	90	40	S	79	78	S	H	L	H	H	H	H
75	ADC	293	8	80	90	S	87	87	S	L	L	L	H	L	H
76	SCC	209	23	0	0	W	2	1	N	H	H	H	H	H	H
77	ADC	11	8	100	100	S	100	100	S	H	H	H	H	H	H
78	ADC	346	5	90	100	S	99	99	S	L	L	L	L	L	H
79	ADC	91	31	0	0	N	3	2	N	L	H	H	L	L	H
80	SCC	61	18	60	60	S	76	77	S	H	H	H	H	H	L
81	SCC	132	18	0	1	N	0	0	N	L	L	H	H	H	H
82	ADC	122	9	100	100	S	98	100	S	H	H	H	H	H	H
83	SCC	132	13	0	1	N	1	0	N	L	H	L	L	H	L
84	ADC	200	18	10	30	W	3	3	W	L	L	L	L	L	H
85	ADC	110	15	35	70	W	34	34	S	H	H	H	H	L	L
86	SCC	199	13	30	50	W	7	7	W	L	L	L	L	L	H
87	ADC	170	2	2	1	W	2	2	W	L	L	L	L	L	L
88	ADC	148	28	1	20	W	7	6	W	L	L	L	L	H	H
89	ADC	288	0	0	0	N	0	0	N	L	L	H	H	L	H
90	SCC	130	15	30	1	W	16	13	W	L	H	L	L	H	H
91	SCC	198	24	80	100	S	75	75	S	H	H	H	H	H	H
92	ADC	208	0	20	90	W	12	12	W	H	L	H	H	L	L

Table 7.1.0. 92 cases of NSCLC assessed by multiplex IHC with immune markers as high (H) or low (L) by median. PD-L1 expression by manual interpretation, algorithm assessment of monoplex IHC and assessment of triplex IHC are also shown as TPS or Negative (N) weak (W) or strong (S) expressers.

7.1.1 TME categorisation by H&E

H&E and monoplex PD-L1 sections were scanned via Aperio CS2 Scanscope slide scanner at 20x and viewed via Aperio ImageScope or the open source QuPath (v0.1.2) software package. A whole section slide of H&E for each case was reviewed to give a subjective interpretation of the immune activity in regards to TILs. Each case had the predominant TIL expression pattern recorded as either immune infiltrated, immune excluded, or an immune desert. This was based on the presence of significant quantities of TILs within the tumour epithelium (infiltrated), significant quantities of TILS present in either the stroma of the tumour and/or around the invading margin of the tumour (excluded) or featured no significant TIL presence (immune desert). Examples are shown in Fig 7.0.1.

7.1.2 Multiplex immunohistochemistry

All 92 cases were prepared for multiplex IHC by having a block of primary tumour cut to produce five serial sections, each of 4µm thickness, placed on a separate glass slides as detailed in the main methods section. A total of 460 tissue sections were thus prepared for this study. After tissue was warmed overnight in a 60°C oven, sections were stored in a 4°C fridge for no longer than two days before being prepared for further work. One section from each case were stored locally as a

precautionary backup of tissue, with four sections from each case shipped to Mosaic Laboratories under cool and dry conditions. (Mosaic Laboratories LLC, 12 Spectrum Pointe Drive, Lake Forest, CA 92630, USA). Mosaic Laboratories provide a range of analytical services including a range of multiplex IHC protocols and have been cited in a large number of publications, including multiple Phase III clinical trials.[625] Sections sent to Mosaic had the following stains performed:

- Slide 1 – Triplex IHC for PD-L1-CD68-CD3
- Slide 2 – Triplex IHC for FoxP3-PD-1-CD8
- Slide 3 – Monoplex IHC for Granzyme B;
- Slide 4 – H&E.

In so doing each case would have a total of 8 markers of the TME produced. Finally, PD-L1 is studied by both SP263 (monoplex) and SP142 (triplex). Example shown in Fig 7.1.0. Slides were then scanned by Mosaic to produce digital images for each section, and the images sent to Definiens (Definiens AG, Bernhard-Wicki-Straße 5, 80636, München, Germany). In addition, PD-L1 (SP263) already performed in this study were scanned and sent to Definiens directly. Definiens were a company specialising in digital image analysis and machine learning to improve analysis of digitally acquired images (they were fully taken over by AstraZeneca during this project in 2019). Definiens' approach was to segment and annotate regions of the tumour into distinct areas by a mixture of morphology and IHC to generate the following for each case: the 'tumour core' (TC), which was further divided into 'tumour epithelium' (TE) and 'tumour non-epithelium' (NE), to divide the tumours into regions of carcinoma tumour cells and stroma. In addition, the 'invasive margin' (IM) was also determined. Image analysis of the sections was then performed to provide measures of each component in mm^2. Finally, the total number of cells in each component positive for immune markers was provided in cells/mm^2 to provide a total of 33 metrics. These are shown in Table 7.1.1.

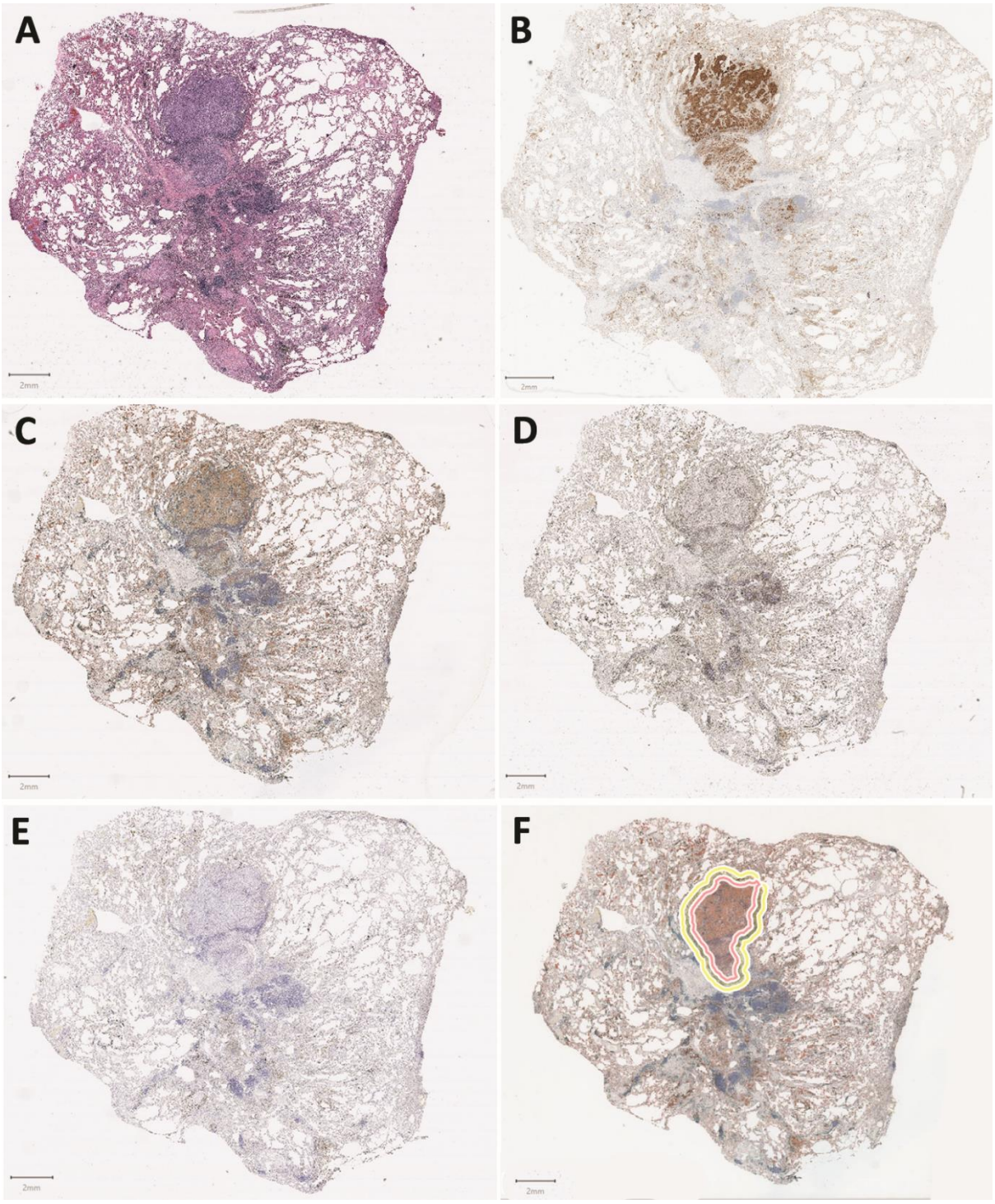


Fig 7.1.0 Low power view of NSCLC tumour with panel of stains performed for multiplex IHC. **A** – H&E; **B** –PD-L1 IHC (SP263); **C** – Triplex PD-L1, CD68, CD3; **D** – Triplex FoxP3, PD1, CD8; **E** – Granzyme B; **F** – Annotated regions of TC (RED) and IM (Yellow).

Area (mm²)			IHC markers (cells/mm²)						
Triplex 1	**Triplex 2**	**Gran B**	**PD-L1**	**PD-1**	**CD3**	**CD8**	**FoxP3**	**CD68**	**Gran B**
TC	TC	TC	TE	TC	TC	TC	TC	TC	TC
TE	TE	IM		TE	TE	TE	TE	TE	IM
NE	NE			NE	NE	NE	NE	NE	
IM	IM			IM	IM	IM	IM	IM	

Table 7.1.1. The 33 metrics provided by the multiplex IHC images and subsequent image analysis.
IHC, immunohistochemistry; TC, tumour core; TE, tumour epithelium; NE, non-epithelial tumour component; IM invasive margin.

Scanned images were accessible via an online portal (Definiens Insights Portal), as well as a report detailing the above metrics for each case. The glass slides were also shipped back to Liverpool where I could review each slide manually as required.

7.1.3 Multiplex Immunofluorescence (mIF)

20 cases were prepared for mIF by having a block of primary tumour cut to produce two serial sections, each of 4µm thickness, placed on a separate glass slides as detailed in the main methods section. A total of 40 tissue sections were prepared, which were dried overnight in a 60°C oven, and then stored in a 4°C fridge for no longer than two days before being prepared for further work. Both sections were shipped to Ultivue under cool and dry conditions, (Ultivue Inc, 763D Concord Avenue, Cambridge, MA, 02138 USA) whom provide a variety of multiplex immuno based techniques. One section for each case was stained with multiplex DAPI-CD8-CD68-PDL1-Pan-CK to provide a measure of PD-L1 expression and immune activity. DAPI (4',6-diamidino-2-phenylindole) is a DNA stain that fluoresces blue and binds to regions of A-T base pair dsDNA, and thus acts as a marker of all cells with nuclei. Cell types are differentiated through the use of CD8 (a cytotoxic T-cell marker) CD68 (a macrophage marker) and a pan-CK which will stain both normal and tumour epithelium, with PD-L1 expression also detected. Example shown in Fig 7.1.1. Slides were then scanned by Ultivue to produce digital images for each section, and the images sent to Definiens. Definiens again applied their image analysis approach to divide each marker into areas of the TC, TE, NE and IM, and in addition to raw values of cell densities per mm², also provided a percentage value of all PD-L1 positive cells, and percentage value of PD-L1 positive and CD68 positive cells to provide 21 metrics for each case. These are shown in Table 7.1.2.

	IF markers (cells/mm²)				
Area (mm²)	**PD-L1**	**CD8**	**CD68**	**PD-L1 %**	**PD-L1/CD68 %**
TC	TC	TC	TC	TC	TC
TE	TE	TE	TE		TE
NE	NE	NE	NE		NE
IM	IM	IM	IM	IM	IM

Table 7.1.2. The 21 metrics provided by the mIF images and subsequent image analysis. IF, immunofluorescence; TC, tumour core; TE, tumour epithelium; NE, non-epithelial tumour component; IM invasive margin.

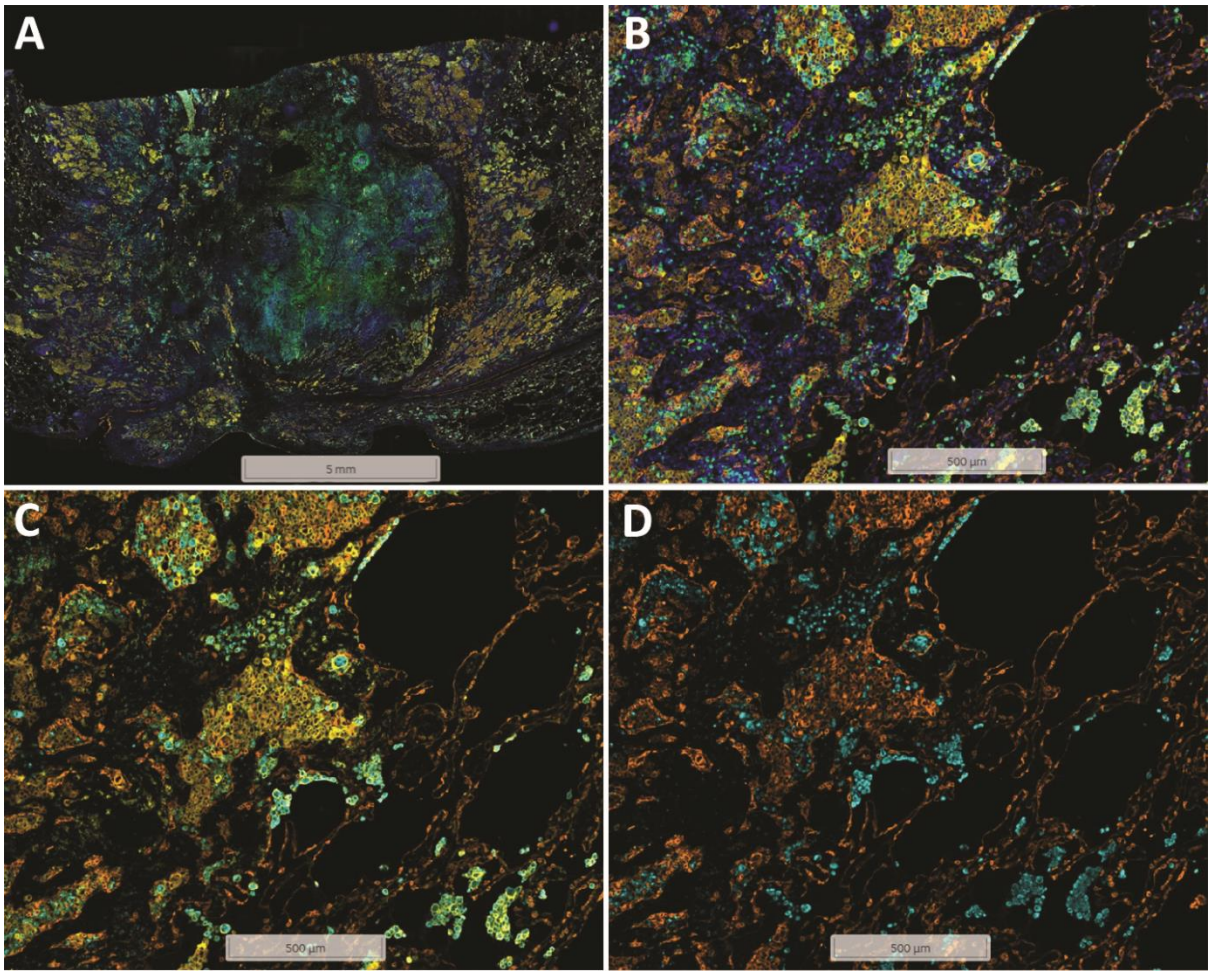


Fig 7.1.1. mIF example **A** - mIF showing DAPI (blue), pan-CK (orange) PD-L1 (yellow), CD8 (green) and CD68 (turquoise). **B** – Higher power view of PD-L1 positive cluster of cells with all IF. **C** – PD-L1 positive cluster of cells with CD8 and DAPI channel removed. **D** – PD-L1 positive cluster of cells with only CD68 and pan-CK channels observable. The identification of the cells as pan-CK positive is now easily achievable.

7.1.4 Validation of image analysis and multiplex IHC approach

As far as possible the same cases and blocks have been used for this project as with previous projects using the LLP cohort. This gives the advantage that many of the cases will already be well characterised for PD-L1 expression. Although Definiens provided explanations of the image analysis and cell density findings within this cohort, and I had access to the images after the event, much of the deep learning and algorithms used to provide the cell densities for markers were something of a 'black-box' approach. In an attempt to validate the accuracy of the image analysis approach, the monoplex IHC for PD-L1 using the SP263 clone was also sent to Definiens for their interpretation using their image analysis. Two PD-L1 TPS sets of data were provided from Definiens: a whole section PD-L1 TPS, and a tumour only ('tumour core') PD-L1 TPS. These could then be compared to the true TPS taken as the whole section manual interpretation scored as per Chapter 4. In addition, whilst the advantages of multiplex IHC make it an attractive tool, difficulties surrounding the interpretation of multiple chromophores within a single section render them difficult to allow for manual interpretation of any specific marker. Example shown in Fig 7.1.2. As a measure of validity of the multiplex approach, the triplex scored PD-L1 cases were also used to generate cell densities of PD-L1 positive, CD68 negative and CD3 negative populations (CD68 excluding macrophages and CD3 excluding TILs, both of which can express PD-L1 (Fig 7.1.3) and therefore increase false positive TPS rates) which could be compared to PD-L1 TPS from the monoplex sections both as raw data (cell densities/mm^2) and by converting the raw data to TPSs by using TC area values. (See statistics section).

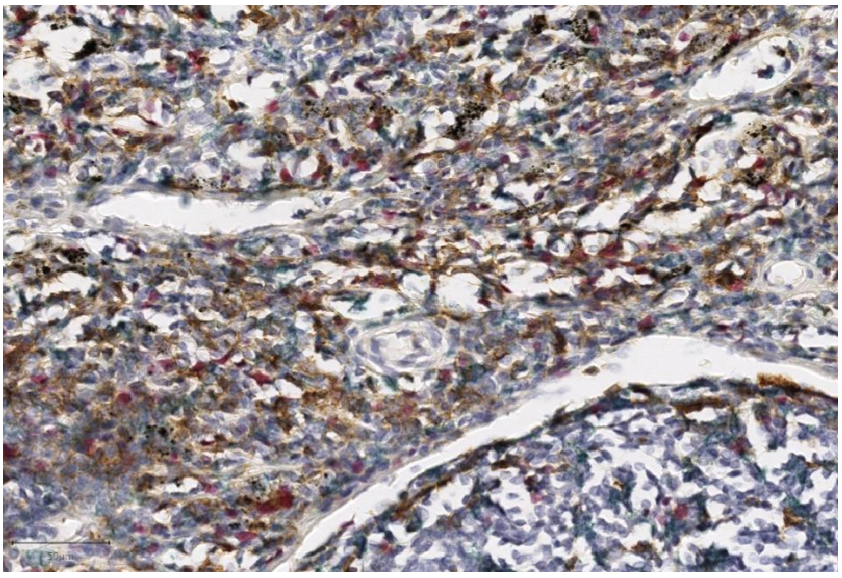


Fig 7.1.2. Multiplex IHC with 3 immune cell markers: Haematoxylin nuclear stain in blue, CD8 in red, FoxP3 in aquamarine and PD-1 in orange. The nature of cells that are positive for multiple markers is hard to discern by eye.

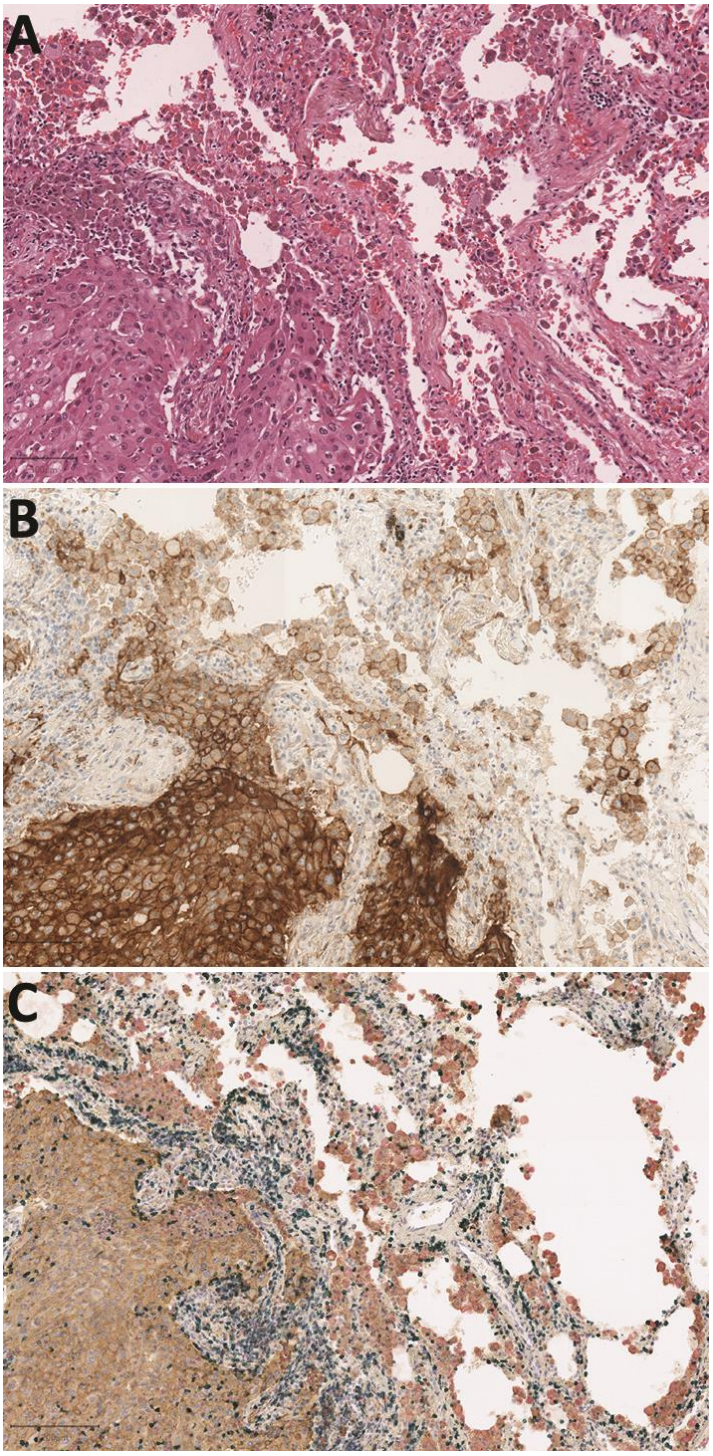


Fig 7.1.3. Tumour cells and macrophages positive for PD-L1 **A** - H&E of tumour cells and macrophages at IM. **B** - PD-L1 IHC (SP263) is positive for both tumour cells and macrophages. **C** – Multiplex IHC shows PD-L1 (brown) for both cell populations, with CD68 staining (red) limited to macrophages.

7.1.5 Association of TME features with PD-L1 expression heterogeneity

Intra-tumoural heterogeneity was defined by the COV within a 1cm² area of primary tumour of PD-L1 expression as per Chapter 4, with higher COV indicating more intra-tumoural heterogeneity. Inter-tumoural heterogeneity was determined by a comparison of primary PD-L1 TPS and metastatic tissue PD-L1 TPS as per Chapter 4, and recorded both as absolute TPS value differences, and as a change in clinical categorisation using PD-L1 negative, weak or strong positive categories (<1%, 1-49% ≥50% TPS). Associations with clinical and pathological features including TME data from the multiplexed tissue analysis was determined in relation to both types of PD-L1 heterogeneity by linear regression for each variable. In addition, multivariate linear regression using all variables was performed for both types of heterogeneity, with a backwards step linear regression model applied to determine all significant factors associated with a change in heterogeneity. A final model was thus determined that describes the significant features associated with both intra- and inter-tumoural PD-L1 expression heterogeneity.

7.1.6 Classification of TMEs

Classification of TMEs was performed by use of the immune cell markers (split into high and low expressers as per median value for each metric in Table 7.1.0) and PD-L1 expression by several methods. First, the patterns of TILs within the TME were examined to classify patients into immune infiltrated, excluded or desert, based on the properties from the multiplex stains. Secondly, tumours were classified by the type of immune activity present within the TME as cytotoxic, immuno-suppressive, mixed-immune (elements of both) or Thelper (Non-Treg) predominant. Thirdly, tumours were classified by their PD-L1 expression (reactive versus constitutive). Finally, the various marker data and classifications devised were combined to devise an approach that classifies tumours into groups based on the likelihood of response to PD-1/PD-L1 IM therapy. These classification tools are shown in Fig 7.1.4-7.1.7. For the final classification, PD-L1 expression is split into clinical categories of negative, weak and strong positive (<1%, 1-49%, ≥50% TPS). As this represents a categorical division of an essentially continuous variable, weak expressers could potentially follow either route. In a bid to average out the difference, tumours with weak PD-L1 expression and high CD8+ve TE are considered to be partial responders to treatment. (Fig 7.1.7)

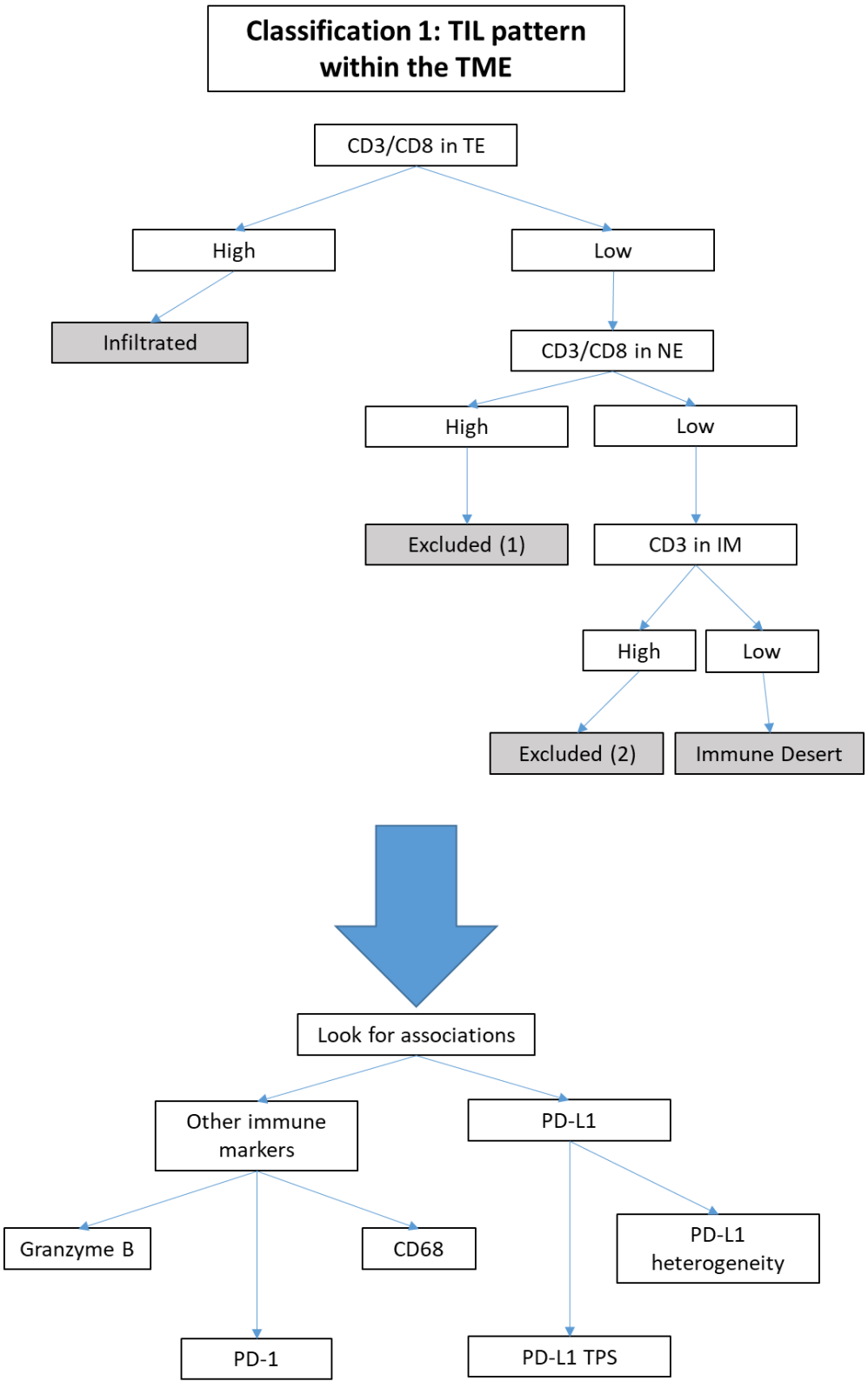


Fig 7.1.4. Flow chart detailing approach to dividing NSCLC carcinomas into different types of immune TME based on expression and distribution of TIL types within the TME. Excluded (1) is stromal, Excluded (2) is IM

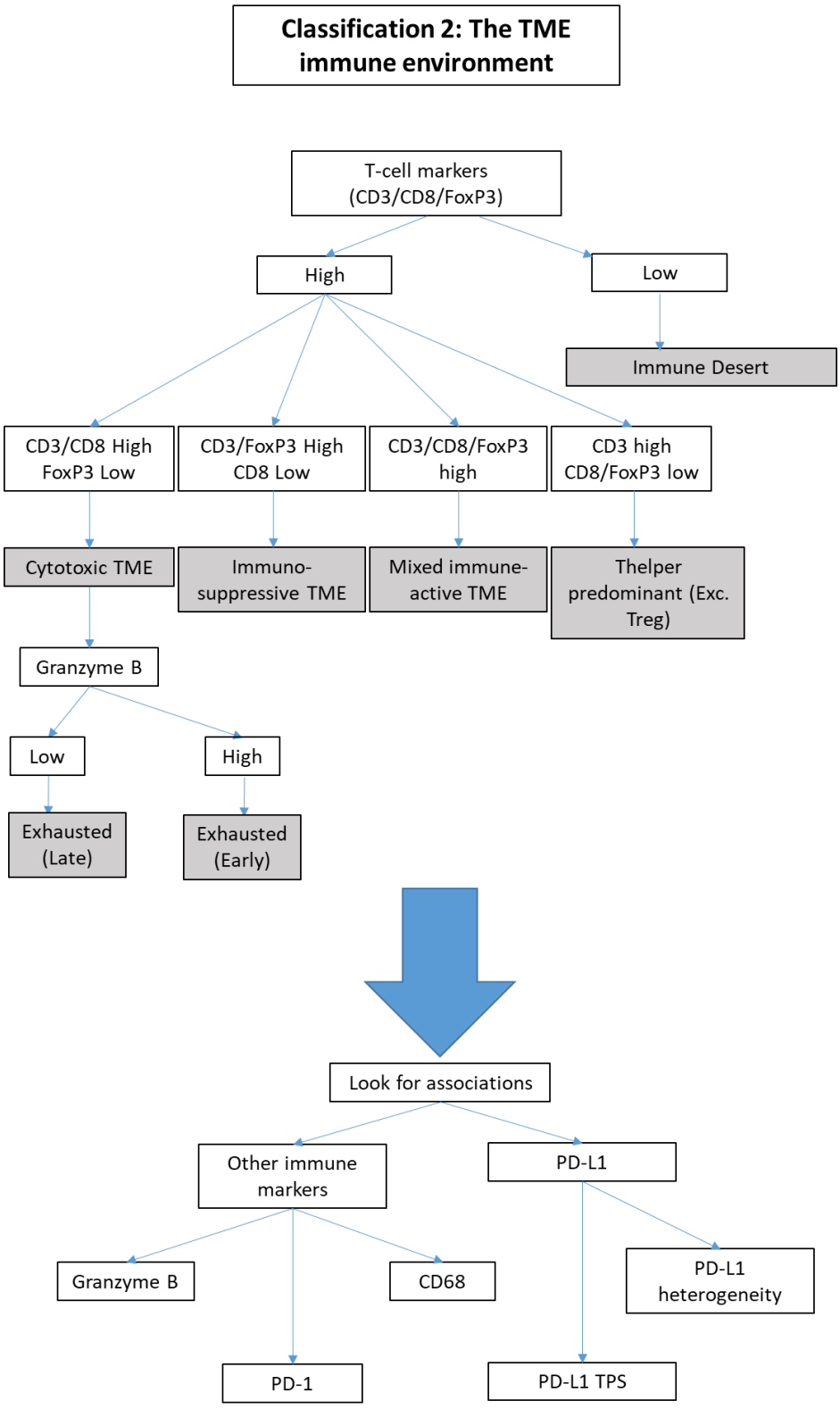


Fig 7.1.5 Flow chart detailing approach to dividing NSCLC carcinomas into different types of immune TME based on expression and distribution of TIL types within the TME, with a further division using Granzyme B.

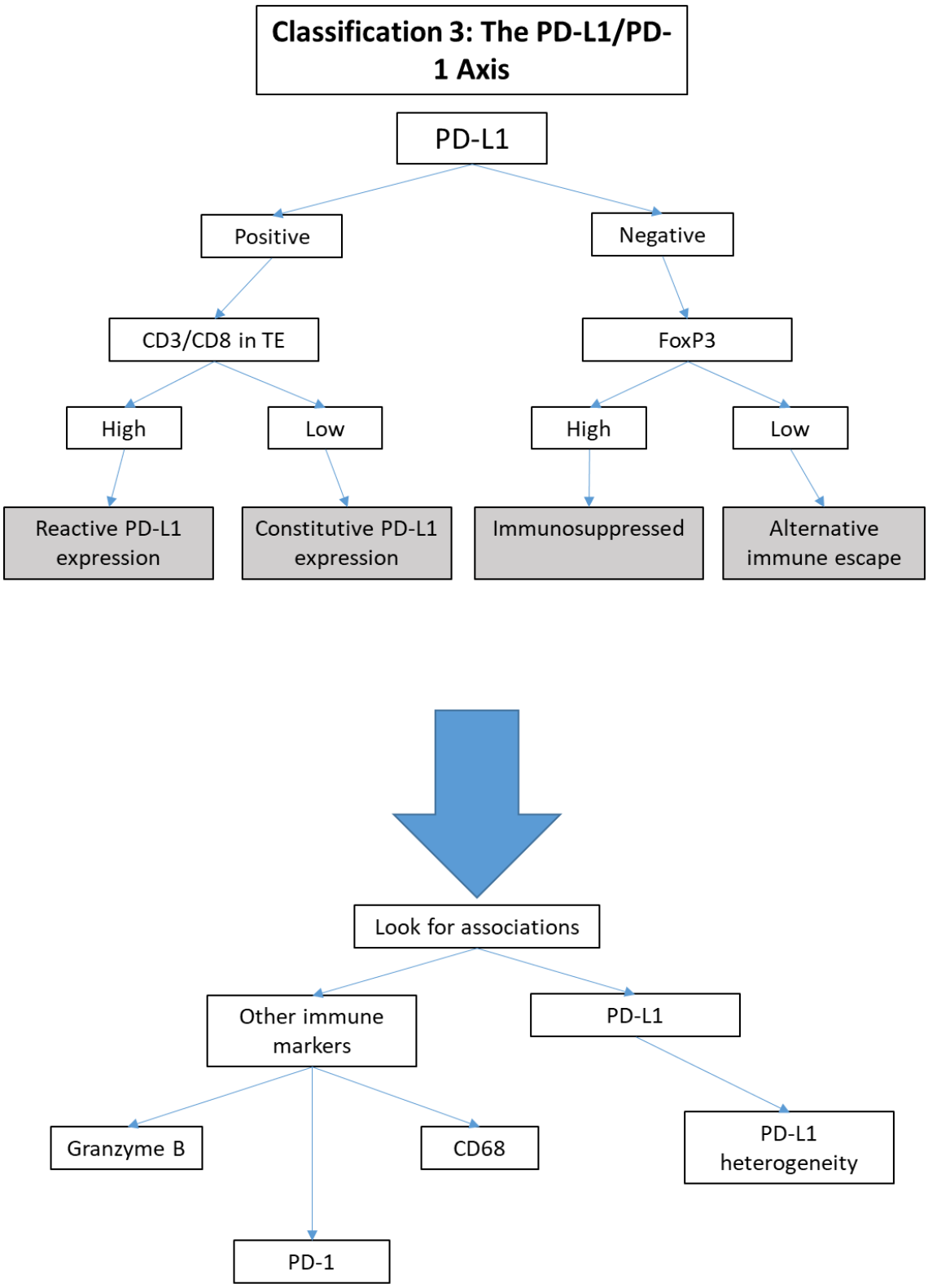


Fig 7.1.6. Flow chart detailing approach to dividing NSCLC carcinomas by types of PD-L1 expression and potential cause for lack of PD-L1 expression based on expression and distribution of TIL types within the TME

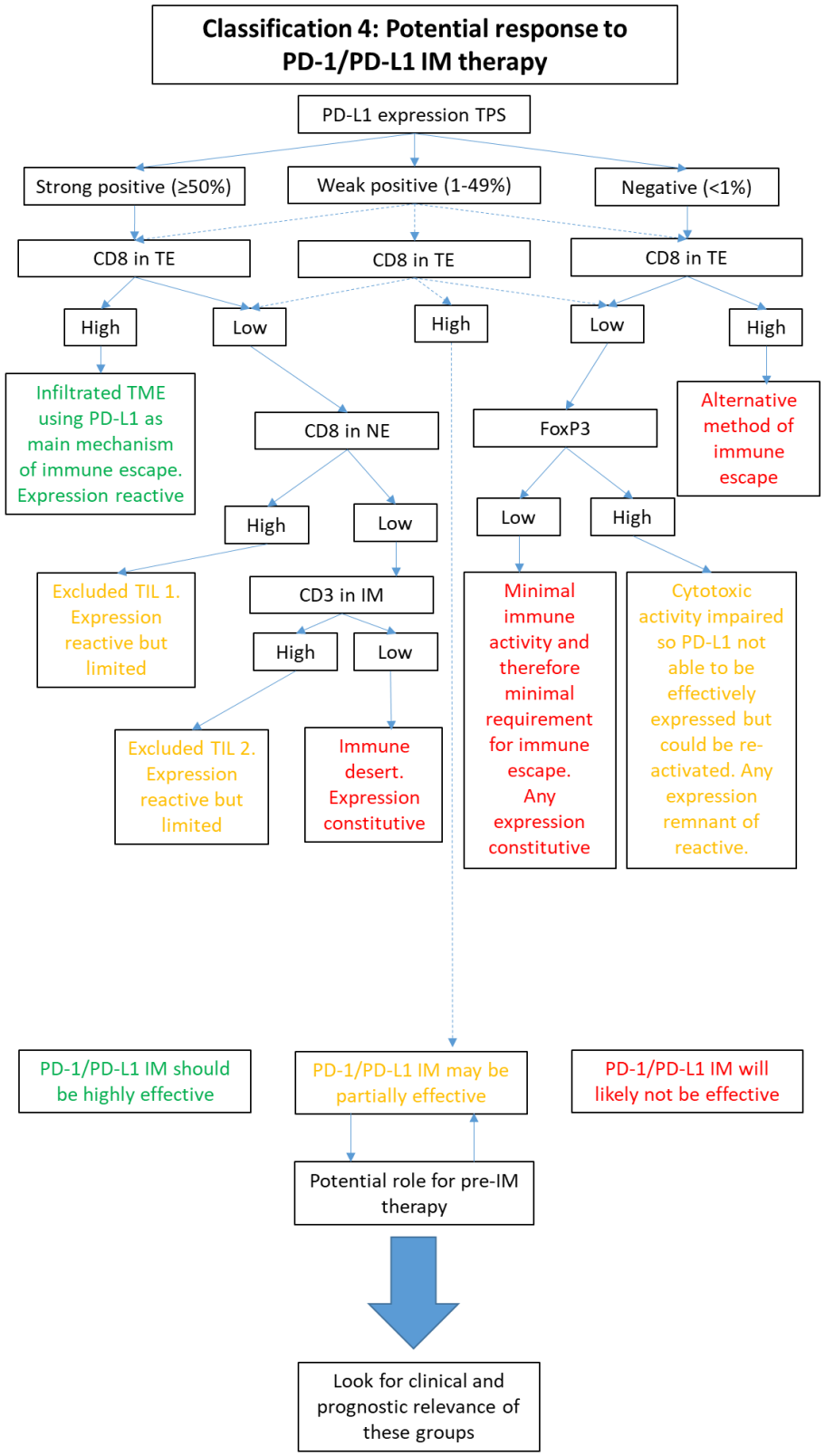


Fig 7.1.7. Flow chart detailing approach to dividing NSCLC carcinomas into predicted response to PD-1/PD-L1 IMs based on PD-L1 expression and the expression and distribution of TIL types within the TME.

7.1.7 Validation of Multiplex immunofluorescence

mIF provides an additional metric to study the validity of multiplex immune approaches for analysing the TME. A PD-L1 TPS generated from the mIF was compared to the manual TPS for all 20 cases. Digital image analysis by Definiens included a simulated rendering of PD-L1 IHC monoplex based on the mIF data, which could be compared to the actual PD-L1 IHC stains for qualitative assessment, example shown in Fig 7.1.8. In addition, 11 of the cases underwent both multiplex IHC and mIF, although fewer markers were used for mIF; (FoxP3, CD3 and Granzyme B are absent). The markers explored in both assays (CD8, PD-L1 and CD68) were compared for densities within the TC (all 3 markers) and TE (CD8 and PD-L1)

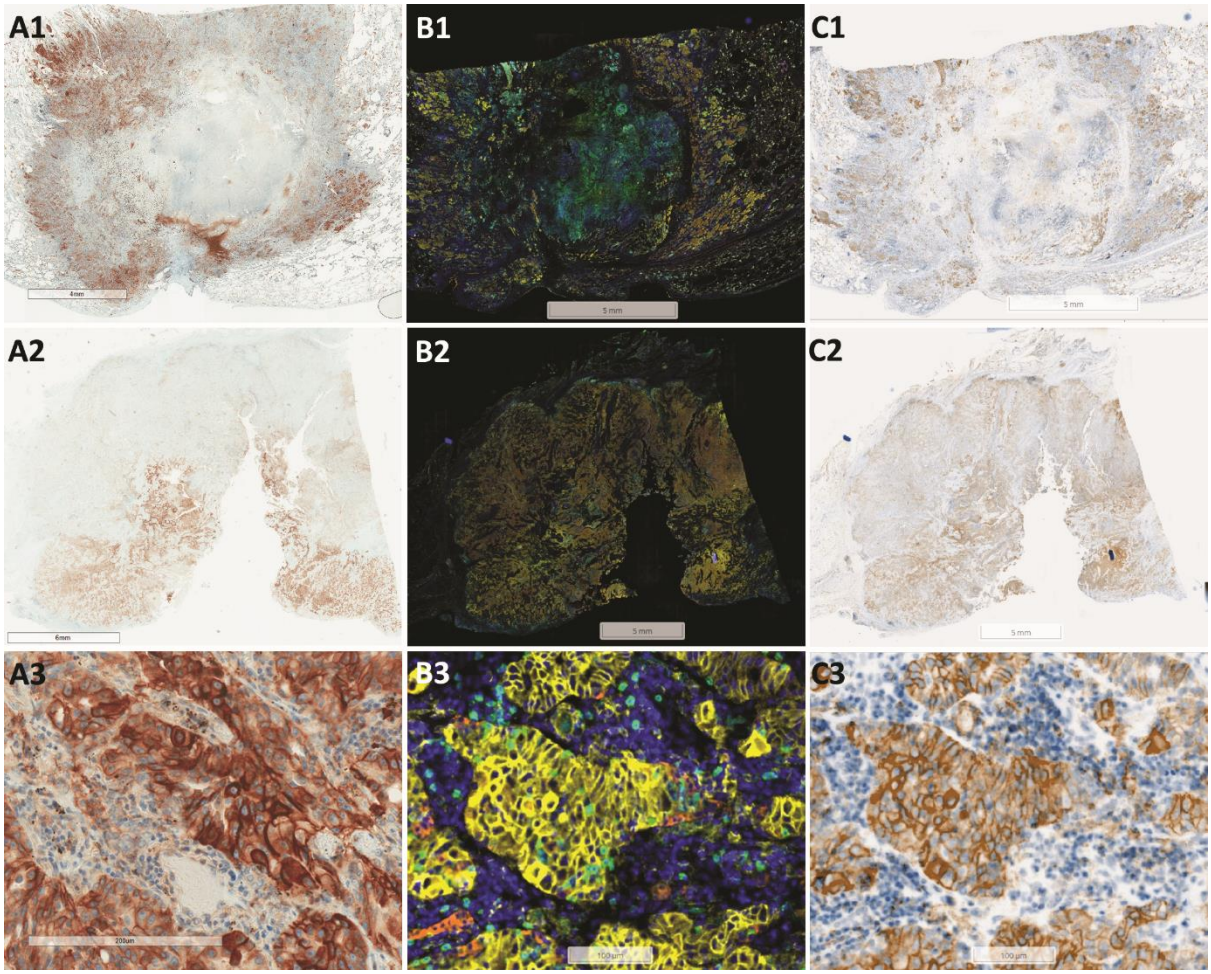


Fig 7.1.8. Matched images for multiplex IHC and miF. **A1-3** IHC stain for PD-L1 (SP263). **B1-3** mIF showing DAPI (blue), pan-CK (orange) PD-L1 (yellow), CD8 (green) and CD68 (turquoise). **C1-C3** – Pseudo-IHC for PD-L1 from the mIF images. The quality and shade of PD-L1 expression is different, but generally in agreement.

7.1.8 Statistics

Comparison between image analysis for monoplex IHC TPS, triplex IHC TPS and mIF compared to manual TPS was performed using ICC for raw TPSs and Cohen's kappa for TPS scores divided into clinical categories of negative, weak and strong positive (<1%, 1-49%, ≥50% TPS). Multiplex IHC and mIF data for PD-L1 was taken as a cell density of PD-L1 positive, CD68/CD3 negative cells per mm^2 area. This was compared to manual TPS scores by two methods: firstly, raw cell densities were divided into clinical categories using the 33rd and 66th percentile, secondly cell densities were converted to a surrogate TPS using the uppermost triplex score within this cohort to define maximum TPS (IE 100%) after outliers were excluded using the inter-quartile method. TPS could then be compared using ICC and in both instances clinical categories compared using Cohen's Kappa. Classification of immune markers from multiplex data was performed used the median value to define cases as 'high' or 'low' for each variable. Spearman's correlation and univariate linear regression was performed to ascertain the relationship between variables. Univariate linear regression was performed in SPSS for each individual independent factor, with subsequent multivariate linear regression using a backwards elimination process to determine significant factors for the dependent variable (PD-L1 expression intra and inter-tumoural heterogeneity). R^2 values, β co-efficient values and p values are acquired from this method. Fitted models were compared using SPSS generalised linear models with all included variables used as main effects within the model to generate a Log Likelihood and Akaike's information Criterion (AIC) as additional measures of the 'goodness of fit' of the model, with changes of ≥4 in AIC taken as important. Prognostic data is calculated using Kaplan-Meier log-rank testing for predicting survival. All significances are taken as $p<0.05$.

7.2 Results

7.2.0 Multiplex IHC Data

All 92 cases were successfully stained for all triplex and monoplex IHC to give 33 metrics as defined in the methods section above. Raw data tables for each case and variable is shown in Appendix 1. 78 cases (85%) had a measurable IM, with the remaining 14 cases therefore unable to have any of the immune markers measured in the IM. All other data was complete for all cases.

7.2.1 Comparison of image analysis PD-L1 TPS and manual TPS

All 92 cases had a manual PD-L1 (SP263) TPS compared to the image analysis TPS score of matched tissue for both the whole section (WS) and within the tumour core (TC) only. There was no significant difference in average TPS between the scoring methods (Manual 31 vs WS 30 vs TC 30

p=0.252) with a high ICC for both WS and TC compared to manual scoring. Compared to manual TPS scores, WS misclassified 16 cases (17%) and TC misclassified 11 cases (12%) with overall OPA, PPV and NPAs generally improved for TC than for WS scoring. Data is summarised in Table 7.2.0.

	OPA	PPV	NPV	ICC	Significance	Kappa	Significance
WS	83	89	85	0.986	p<0.0001	0.888	p<0.0001
TC	88	87	90	0.988	p<0.0001	0.909	p<0.0001

Table 7.2.0. Correlations and predictive values of image analysis of IHC for assessing PD-L1 expression compared to manual scoring of PD-L1 IHC (SP263).

Triplex IHC PD-L1/CD68/CD3 sections were analysed to give a cell density/mm² of PD-L1 (SP142) positive-CD68/CD3 negative cells that were used as a surrogate marker of a PD-L1 TPS. These were compared to manual PD-L1 (SP263) TPSs. Raw data for cell densities/mm² ranged from 0-5100, with a strong relationship to manual TPS scores (Correl. 0.888 (p<0.001) R^2 0.722 (p<0.001)). Categorisation of triplex raw data in clinical categories using 33rd and 66th percentiles showed a reasonably good OPA of 79% (Kappa 0.69, p<0.0001). Converting the triplex data to TPS values showed a good consistency between triplex and manual scoring (ICC 0.801 (p<0.0001)) with no significant difference in average TPS (31 vs 27 p=0.646). However, triplex PD-L1 TPS underscored PD-L1 in several cases, relative to monoplex scoring; an example is shown in Fig 7.2.0. Grouping into clinical categories was generally more accurate when using raw triplex data than for converted TPSs. Results summarised in Table 7.2.1

	OPA	PPV	NPV	ICC	Significance	Kappa	Significance
Tri	79	89	90	n/a	n/a	0.69	p<0.0001
Tri TPS	68	89	79	0.801	p<0.0001	0.572	p<0.0001

Table 7.2.1. Correlations and predictive values of image analysis of multiplex IHC for assessing PD-L1 expression using Triplex cell densities (Tri) and Triplex converted to percentage (Tri TPS) compared to manual scoring of monoplex PD-L1 IHC (SP263).

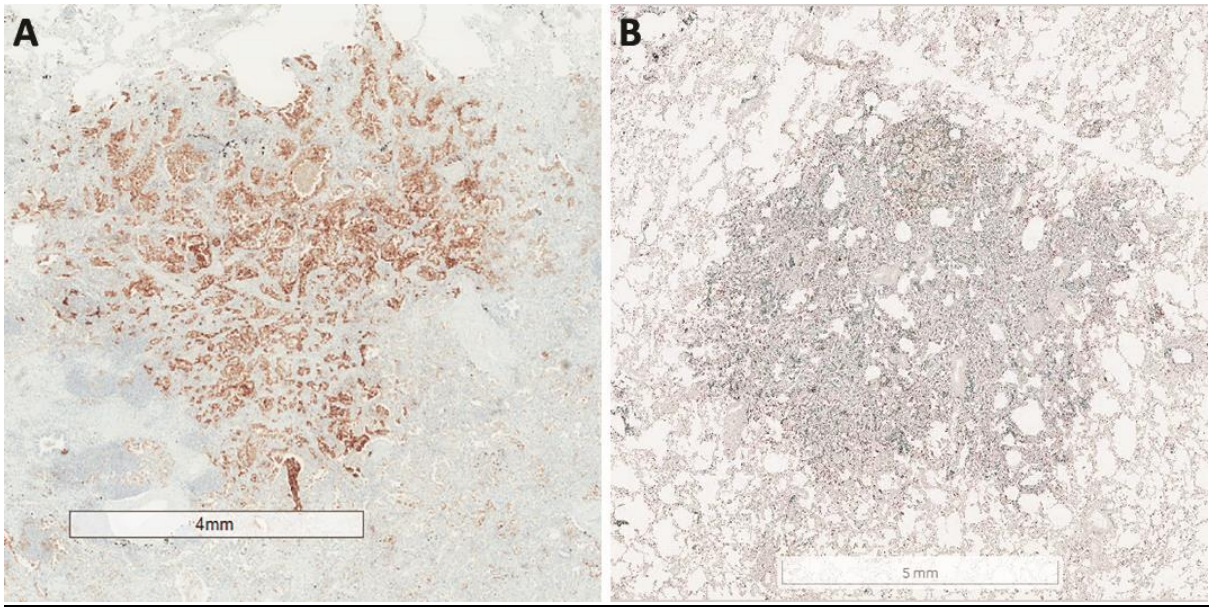


Fig 7.2.0. Matched sections of NSCLC tumour stained for PD-L1 by monoplex and triplex assays. **A –** Monoplex PD-L1 (SP263) shows majority of tumour cells are staining positive for PD-L1 **B** – Triplex PD-L1 (SP142) shows few tumour cells staining for PD-L1.

7.2.3 Associations of immune markers with PD-L1 expression heterogeneity using multivariate analysis modelling

Four immune variables within the TME were associated with significant but generally modest changes in PD-L1 expression inter-tumoural heterogeneity by univariate linear regression: increasing CD8 and CD3 were both associated with increasing inter-tumoural heterogeneity (CD8: correl. 0.336 p=0.001, R^2=0.102, p=0.002; CD3 correl. 0.293 p=0.022, R^2=0.062 p=0.016) as was categorical placement into an infiltrating pattern of TILs by TME (R^2 0.058, p=0.021) and the ratio of CD8TE:NE (a measure of infiltrating TILs) (correl. 0.384 p<0.0001, R^2 0.058 p=0.021) Being female was also associated with increased inter-tumoural heterogeneity (R2 0.041, p=0.041). Increasing PD-1 and CD68 levels both trended towards decreasing inter-tumoural heterogeneity (R^2 0.026 p=0.122 and R^2 0.035 p=0.074 respectively). A multivariate analysis model using all variables (Appendix 2) and inter-tumoural heterogeneity of PD-L1 expression was then performed to confirm associations and control for confounding factors, which provided a baseline AIC of 848 and LLP of -394. Linear regression with backwards elimination for least significant variables resulted in a 20 step process resulting in four significant variables remaining: CD8 and CD3, which were both associated with significantly increased inter-tumoural variation (b 0.027 p=0.001; b 0.012 p=0.024) and PD-1 and CD68 which were both associated with significantly decreased inter-tumoural variation (b -0.065 p=0.003; b -.031 p=0.02). Gender and measures of infiltration did not retain their significance in multivariate analysis.

Use of the four significant variables in multivariate analysis resulted in an improved fit of the model: AIC 818 and LL -403 (an AIC decrease of 30).

This process was then repeated to ascertain the association of each variable with PD-L1 expression intra-tumoural heterogeneity. By univariate linear regression five variables were significantly associated with a modest decrease in intra-tumoural heterogeneity: increasing CD8, CD3, FoxP3 and CD68 levels (CD8 correl. -0.451 p=0.003, R^2=0.137, p=0.017; CD3 correl. -0.519 p=0.001, R^2=0.215 p=0.002; FoxP3 correl. -0.411 p =0.008, R^2 0.108 p=0.036, CD68 correl. -0.477 p=0.003, R^2 0.180 p=0.006), and an increased ratio of CD8TE:NE (correl. -0.391 p=0.011, R^2 0.122 p=0.014). Increasing Granzyme B levels significantly correlated with decreasing intra-tumoural heterogeneity (correl. -0.331 p=0.034) but not as an independent variable in linear regression (R^2 0.078 p=0.077). Multivariate analysis was then performed to look for associations with PD-L1 expression intra-tumoural heterogeneity and provided a baseline AIC of 450 and LL of -195. Linear regression with backwards elimination for least significant variables resulted in a 12 step process resulting in a single remaining significant variable: CD8 expression, which was associated with a significant decrease in intra-tumoural heterogeneity (b -0.84 p=0.023). Use of a single variable, however, did not improve the goodness of fit of the model (AIC 487, LLP-241).

7.2.4 Categorisation of TILs by H&E

The 92 selected cases were reviewed for the predominant pattern of TILs within a whole tumour section H&E slide and categorised as infiltrated, excluded or immune desert. The PD-L1 TPS for each category was also recorded. Immune desert TIL TMEs had a significantly lower average PD-L1 TPS than both infiltrated (6 vs 36 p<0.012) and excluded (6 vs 38 p<0.007). Results are summarised in Table 7.2.2.

TME	Cases	PD-L1	Desert (%)	Excluded (%)	Infiltrated (%)
Desert	17	**Negative**	9 (53)	9 (36)	14 (28)
Excluded	25	**Weak**	7 (41)	5 (20)	15 (30)
Infiltrated	50	**Strong**	1 (6)	11 (44)	21 (42)

Table 7.2.2. PD-L1 expression in 92 NSCLCs divided into clinical categories for different immune TMEs by H&E.

Immune desert TMEs had significantly less inter-tumoural PD-L1 expression heterogeneity compared to excluded (3 vs 14 p=0.025) and infiltrated TMEs (3 vs 17 p=0.002) although this did not translate to a significant change in cases placed into different clinical group categories. Immune desert TMEs had significantly less CD3, CD8, FoxP3, PD-1 and Granzyme B than both infiltrated (CD3 665 vs 1398

p<0.001, CD8 265 vs 651 p<0.001, FoxP3 107 vs 239 p=0.001, PD-1 56 vs 144 p<0.0001, Granzyme B 0.01 vs 0.11 p=0.001) and excluded TMEs (CD3 665 vs 1385 p=0.001, CD8 265 vs 549 p=0.001, FoxP3 107 vs 179 p=0.024, PD-1 56 vs 174 p<0.001, Granzyme B 0.01 vs 0.06 p=0.032), and significantly less CD68 than infiltrated TMEs (223 vs 424 p=0.007). Results summarised in Table 7.2.3.

	Desert	Excluded	Infiltrated	Sig
COV (Intra)	122	109	86	N (p=0.576)
TPS change (Inter)	3	14*	17*	Y (p=0.05)
Group change	7	10	13	N (p=0.336)
CD3	665	1385*	1398*	Y (p<0.0001)
CD8	265	549*	651*	Y (p<0.0001)
FoxP3	107	179*	231*	Y (p=0.002)
PD-1	56	174*	144*	Y (p<0.0001)
CD68	223	315*	424*	Y (p=0.008)
Granzyme B	0.01	0.06*	0.11*	Y (p=0.001)

Table 7.2.3. Comparison of different immune TMEs (desert, excluded and immune) by H&E for PD-L1 expression in 92 NSCLCs for intra-tumoural and inter-tumoural heterogeneity and immune cell markers. * denotes significantly higher value.

TME categories were also compared for prognostic information by KM survival analysis, with a non-significant trend in worse predicted outcome for immune desert vs excluded and infiltrated TMEs (43 vs 54 vs 59 months OS, χ^2 value 1.677, p=0.432).

7.2.5 Classification of TMEs by multiplex IHC TIL data

All 92 cases were classified into TIL status as per Fig 7.1.4. Cases were thus divided into immune infiltrating, excluded 1 (stromal), excluded 2 (invasive margin) or immune desert. The PD-L1 TPS for each category was also recorded. There was an overall 81% concordance rate for categorisation of TIL expression by multiplex IHC with categorisation by H&E assessment. Infiltrated TIL TMEs had a significantly higher PD-L1 TPS than both immune desert (53 vs 9 p<0.0001) and excluded (53 vs 10 p<0.0001). These results are summarised in Table 7.2.4.

TME	Cases	H&E	PD-L1	Desert (%)	Exclude 1 (%)	Exclude 2 (%)	Infiltrated (%)
Desert	25	17	Negative	13 (52)	11 (55)	2 (40)	6 (13)
Exclude 1	15	25*	Weak	10 (40)	7 (35)	3 (60)	14 (30)
Exclude 2	5		Strong	2 (8)	2 (10)	0 (0)	27 (57)
Infiltrated	47	50					

Table 7.2.4. PD-L1 expression in 92 NSCLCs divided into clinical categories for different immune TMEs by multiplex IHC. * = includes both types of excluded

A significant increase in intra-tumoural PD-L1 expression heterogeneity is seen between excluded and infiltrated TMEs (209 vs 73 p=0.023) and a non-significant trend of increased intra-tumoural heterogeneity between desert and infiltrated TMEs (152 vs 73 p=0.056). There is a significant increase in inter-tumoural PD-L1 expression heterogeneity between infiltrated and excluded TMEs (22 vs 7 p<0.0001) and infiltrated and immune desert TMEs (22 vs 5 p<0.0001), although this did not translate to more cases changing clinical categories. Immune desert TMEs had less PD-1 than both infiltrated (63 vs 180 p<0.001) and excluded TMEs (63 vs 123 p=0.08) and less CD68 than infiltrated TMEs (232 vs 311 p=0.002). Infiltrated TMEs had significantly more Granzyme B than both excluded (0.14 vs 0.03 p=0.002) and immune desert TMEs (0.14 vs 0.01 p<0.001) and significantly more FoxP3 that immune desert TMEs (251 vs 105 p<0.001) and trended towards higher levels than excluded TMEs (251 vs 171 p=0.067. These results are summarised in Table 7.2.5

	Desert	Excluded**	Infiltrated	Sig
COV (Intra)	152	209	73	Y (p=0.005)
TPS Change (Inter)	5	7	22*	Y (p<0.0001)
Group change	6	14	10	N (p=0.586)
FoxP3	105	171	251*	Y (p<0.0001)
PD-1	63	123*	180*	Y (p<0.0001)
CD68	232	311	444*	Y (p=0.003)
Granzyme B	0.01	0.03	0.14*	Y (p<0.0001)

Table 7.2.5. Comparison of different immune TMEs (desert, excluded, infiltrated) by multiplex IHC in 92 NSCLCs for intra-tumoural and inter-tumoural PD-L1 expression heterogeneity and immune cell markers. *denotes significantly higher value. ** = includes both types of excluded.

TME categories were also compared for prognostic information by KM survival analysis: Infiltrated TIL TMEs trended non-significantly towards longer predicted OS vs excluded and desert (57 vs 37 vs 35 months, χ^2 value 3.743, p=0.154).

7.2.6 Classification of TME immune status by multiplex IHC immune markers

All 92 cases were classified into TME immune status as per Fig 7.1.5. The PD-L1 TPS for each category was also recorded. Mixed immune status had a significantly higher average TPS than immune desert (54 vs 7 p<0.0001) and immuno-suppressive TMEs (54 vs 4 p=0.012). Cytotoxic TMEs trended towards a higher average TPS vs immuno-suppressive (32 vs 4 p=0.072). Results summarised in Table 7.2.6.

TME	Cases	PD-L1: Negative (%)	Weak (%)	Strong (%)
Desert	19	11 (58)	7 (37)	1 (5)
Mixed	39	6 (15)	10 (26)	23 (59)
Cytotoxic	18	5 (28)	6 (33)	7 (39)
Immuno-suppressive	10	2 (20)	8 (80)	0
Thelper	3	2 (67)	1 (33)	0

Table 7.2.6. PD-L1 expression in 92 NSCLCs divided into different TMEs based on immune marker expression and distribution by multiplex IHC.

There was no significant difference in intra-tumoural PD-L1 expression heterogeneity between TME types, although immune desert and immuno-suppressed TMEs trended non-significantly towards more intra-tumoural heterogeneity compared to mixed immune TMEs (158 vs 85 p=0.062, 140 vs 85 p=0.068 respectively). There was a significant increase in inter-tumoural PD-L1 expression heterogeneity of mixed immune TMEs compared to immune deserts (54 vs 7 p=0.006) and Th only TMEs (54 vs 0 p=0.043) and trended towards more inter-tumoural heterogeneity compared to immuno-suppressive TMEs (54 vs 4 p=0.086). Cytotoxic TMEs also had a significant increase in inter-tumoural PD-L1 expression heterogeneity versus immune desert TMEs (32 vs 7 p=0.014) and Th only TMEs (32 vs 0 p=0.045) and trended towards higher inter-tumoural heterogeneity vs immuno-suppressive TME (32 vs 4 p=0.101).

Immune deserts had significantly fewer PD-1 positive cells than immuno-stimulatory (45 vs 108 p=0.037) and mixed immune TMEs (45 vs 207 p<0.001), and mixed immune TMEs had significantly more PD-1 positive cells than cytotoxic TMEs (207 vs 108 p=0.036). Mixed immune TMEs had significantly more CD68 positive cells than immune desert (490 vs 218 p<0.0001) and cytotoxic TMEs (490 vs 273 p=0.012). Mixed immune TMEs has significantly more Granzyme B (0.15 vs 0.01 p=0.001) than immune desert and immuno-suppressive TMEs (0.15 vs 0.03 p=0.021). Sub-group analysis of cytotoxic TMEs divided by the presence of high vs low Granzyme B found no significant difference in intra-tumoural heterogeneity (COV 97 vs 108 p=0.748) or inter-tumoural heterogeneity (TPS change 15 vs 25 p=0.408). Results summarised in Table 7.2.7.

	Desert	Mixed	Cytotoxic	Immuno-suppressive	Thelper	Sig Difference
COV (Intra)	158	85	104	140	n/a	N (p=0.08)
TPS change (Inter)	7	20*	20*	4	0	Y (p=0.011)
Group Change	7	8	8	5	1	N (p=0.253)
PD-1	45	207*	108*	99	79	Y (p<0.0001)
CD68	218	490*	273	278	307	Y (p<0.0001)
Granzyme B	0.01	0.15*	0.03	0.03	0.02	Y (p<0.0001)

Table 7.2.7. Comparison of different immune TMEs by multiplex IHC in 92 NSCLCs for intra-tumoural and inter-tumoural PD-L1 expression heterogeneity and immune cell markers. * denotes significantly higher value.

Mixed immune and cytotoxic TMEs trended towards longer predicted OS than immune deserts, immuno-suppressive and Thelper only TMEs (62 vs 52 vs 37 vs 34 vs 22 months OS respectively, χ^2 value 6.303, p=0.178). Sub-group analysis of cytotoxic TMEs divided by the presence of high versus low Granzyme B found a significant difference in predicted OS in favour of low Granzyme B (79 vs 26 months, p=0.009). Analysis of the entire cohort comparing CD8+ve TE high, Granzyme B low TMEs versus all other cases showed an improved predicted OS in favour of the former, (77 vs 42 months p=0.006), but this significance is lost by further stratifying the groups for high PD-1 expression (73 vs 45 months p=0.081).

7.2.7 Classification of PD-L1 expression by multiplex IHC

All 92 cases were classified into PD-L1 expression groups as per Fig 7.1.6 and shown in Table 7.2.8. There was no significant intra-tumoural heterogeneity difference between reactive and constitutive expression or between PD-L1 negative categories, although intra-tumoural heterogeneity was significantly greater for reactive PD-L1 expression versus immunosuppressed categories (COV 260 vs 77 p=0.042). Inter-tumoural heterogeneity was significantly higher for reactive PD-L1 expression than constitutive expression (23 vs 9 TPS difference p=0.018). PD-1, CD68 and Granzyme B were all significantly higher for reactive PD-L1 expression than constitutive expression (185 vs 54, p<0.0001; 436 vs 244 p=0.05; 0.13 vs 0.01 p=0.001 respectively). Results are shown in Table 7.2.9.

TME	Cases	Negative (%)	Weak (%)	Strong (%)
Reactive	47	0	17 (36)	30 (64)
Constitutive	15	0	14 (93)	1 (7)
Immunosuppressed	15	15 (100)	0	0
Alternate	15	15 (100)	0	0

Table 7.2.8. PD-L1 expression in 92 NSCLCs divided into clinical categories based on expression of PD-L1 (reactive or constitutive) or possible reasons for being PD-L1 negative (immunosuppressed TME or alternative immune-escape mechanism deployed) by multiplex IHC.

	Reactive	Constitutive	Immunosuppressed	Alternate	Sig Difference
COV	77	118	242	260*	p=0.006
TPS change	23*	9*	<1	7	p<0.0001
PD-1	185*	54	55	145*	p<0.0001
CD68	436*	244	248	332	p=0.015
Granzyme B	0.13*	0.01	0.01	0.04	p<0.0001

Table 7.2.9 Comparison of different immune TMEs by multiplex IHC for intra-tumoural and inter-tumoural PD-L1 expression heterogeneity and immune cell markers. * denotes significantly higher value.

There was a non-significant trend towards improved predicted OS for reactive expression than constitutive, immunosuppressed and alternate immune escape categories (52 vs 37 vs 43 vs 38 months respectively, χ^2 value 2.944, p =0.400)

7.2.8 Classification by potential response to PD-1/PD-L1 IMs by multiplex IHC

As per Fig 7.1.7, all 92 cases were divided into probable response to PD-1/PD-L1 IM therapy categories. PD-L1 strong expressers (≥50% TPS) and PD-L1 negative (<1% TPS) tumours followed a single route, whereas PD-L1 weak expressers (1-49% TPS) could potentially follow either the strong positive or negative classification routes; however there was only a single case difference in final categories if all weak positive cases followed either. This is summarised in Table 7.2.10 with final groups shown in Table 7.2.11.

PDL1	Strong	Weak	PDL1	Negative	Weak
Total	31	31	**Total**	30	31
CD8 TE High	27	13	**CD8 TE High**	6	13
CD8 NE High	2	4	**FoxP3 High**	10	8
CD3 IM high	0	3	**FoxP3 Low**	14	10
CD8 /3 Low	2	11			

Table 7.2.10. NSCLC tumours divided by PD-L1 expression and immune cell markers by multiplex IHC into categories of potential response to PD-1/PD-L1 IMs (Green – good, Orange – partial, Red – no response).

	N	%
Good response	27	29
Partial response	32-33	35-36
No response	32-33	35-36

Table 7.2.11. Final categories for potential response to PD-1/PD-L1 IMs. PD-L1 weak positive (1-49%) could potentially follow positive or negative route, but only a single case change is noted in final categorisation.

Intra-tumoural heterogeneity is significantly reduced in the good response category compared to both partial and no response groups (30 vs 160 p=0.001, 30 vs 141 p<0.0001 respectively). Inter-tumoural heterogeneity is significantly reduced in the no response category compared to both good and partial response groups (5 vs 23 p<0.0001, 5 vs 17 p=0.037 respectively). The no response group had significantly lower PD-1 expression levels than both good and partial response groups (71 vs 205 p<0.001, 71 vs 145 p<0.001 respectively). The good response group had significantly higher CD68 and Granzyme B levels than both partial (CD68 546 vs 295 p=0.004; Granzyme B 0.19 vs 0.05 p<0.001) and no response groups (CD68 546 vs 264 p=0.002, 0.19 vs 0.02 p<0.001). These results are shown in table 7.2.12

	Good response	Partial response	No response	Significance
COV (Intra)	30	160*	141*	Y (p<0.0001)
TPS change (Inter)	23*	17*	5	Y (p<0.0001)
Group Change	6	12	11	N (p=0.436)
PD-1	205*	145*	71	Y (p<0.0001)
CD68	546*	295	263	Y (p<0.0001)
Granzyme B	0.19*	0.05	0.02	Y (p<0.0001)

Table 7.2.12. Comparison of different groups of NSCLCs categorised by potential response to PD-1/PD-L1 IMs for intra-tumoural and inter-tumoural PD-L1 expression heterogeneity and immune cell markers by multiplex IHC. * denotes significantly higher value.

There was a non-significant trend toward improved survival for the good respond group compared to partial and no response groups (55 vs 43 vs 40 months respectively, χ^2 value 2.384 p=0.304). Sub group analysis by Granzyme B found no difference in predicted OS in the good responders (high vs low 41 vs 68 months p=0.351), partial responders (high vs low 37 vs 47 months p=0.450) or no responders group (high vs low 22 vs 47 months p=0.085). Fig 7.2.1..

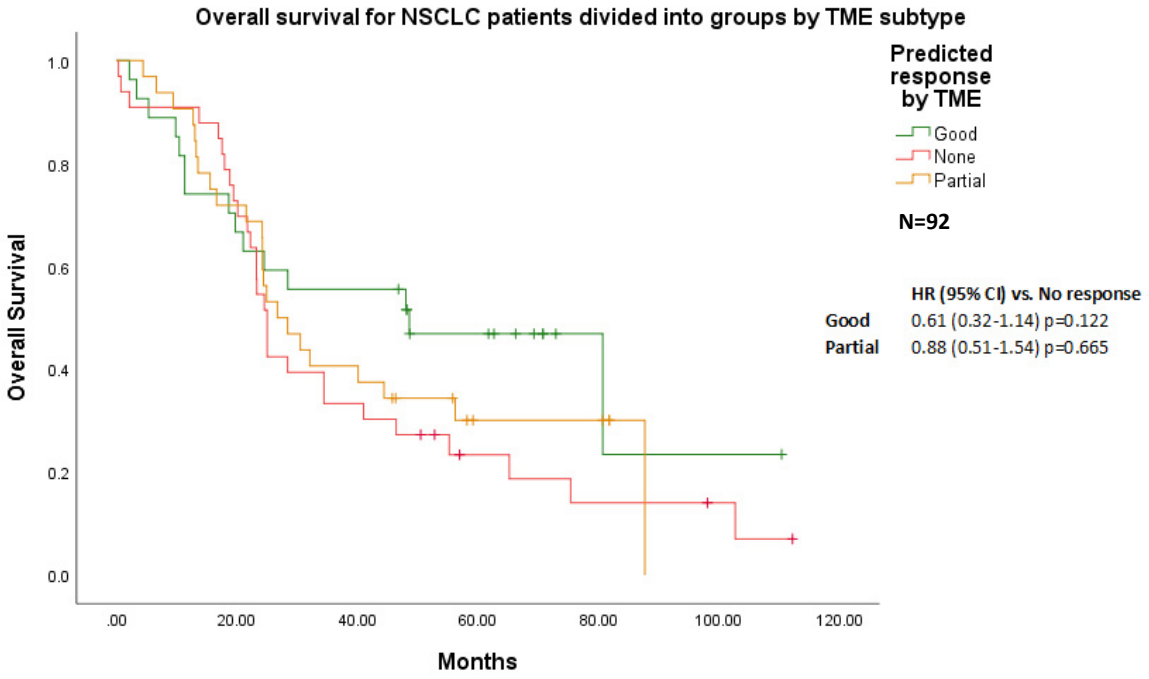


Fig 7.2.1. Kaplan-Meier (Log-Rank) survival curve for NSCLC patients divided by predicted response to PD-1/PD-L1 IMs based on the immune characteristics of their tumour TME. No significant differences are seen, but improved response to IMs is associated with a trend towards improved OS. (55.1 vs 42.9 vs 39.7 months, p=0.304.)

7.2.9 Predicted survival by TME features

PD-L1 alone trended towards a higher OS when using a ≥50% TPS dichotomous division (41 vs 55 months (p=0.99)), and the presence of CD8+ve T-cells (TE) alone also trends towards a higher OS using the median to split the cohort as high or low (40 vs 52 months (p=0.13)). Combining these to compare PD-L1 high/CD8 high against PD-L1 low/CD8 low found a significant difference in predicted OS (55 vs 35 months, χ^2 value 4.708 p=0.03), as shown in Fig 7.2.2. Interestingly, if using the same ≥50% TPS dichotomous division by the monoplex deep learning algorithm instead of manual scoring, only ficw cases are changed (all decreased from manual score of strong ≥50% TPS to weak 1-49% TPS PD-L1 expressers), but the prognostic outcome is slightly further in favour of PD-L1 strong/CD8 strong, with greater statistical significance, (57 vs 35 months, χ^2 value 5.896 p=0.015). The

significance is not seen if using Triplex PD-L1 data to define PD-L1 high (56 vs 40 months, χ^2 value 2.325, p=0.127).

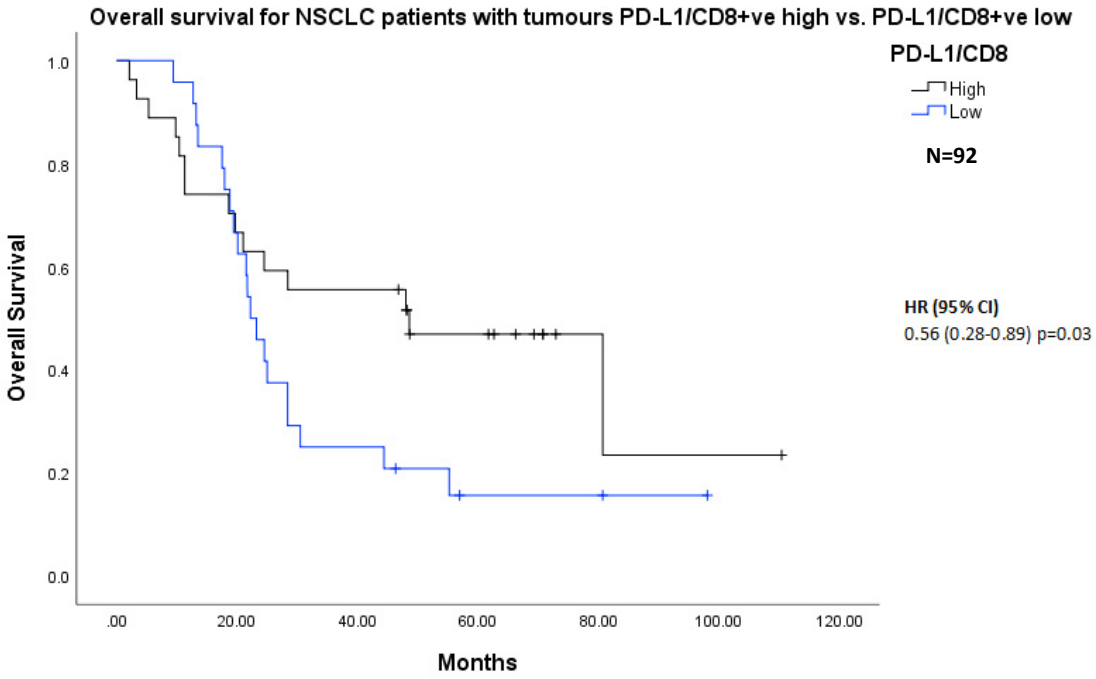


Fig 7.2.2. Kaplan-Meier (Log-Rank) survival curve for NSCLC patients grouped into PD-L1 expression high and CD8 expression high vs. PD-L1 expression low and CD8 expression low. A significant difference in OS is seen in favour of the high group. (55 vs 35 months p=0.03)

7.2.10 MIF for PD-L1 assessment

mIF was compared to manual PD-L1 scores in 18 cases. There was no significant difference in average PD-L1 expression between manual TPS scoring and mIF (20 vs 25 p=0.372), but the OPA for using clinical categories is 0.39 with non-significant intra-class coefficient and Cohen's kappa. This could be improved if using a dichotomous cut-off of ≥50% TPS. Results are shown in Table 7.2.13.

Case	PD-L1 Manual	PD-L1 IF	Difference	Change category?
1	10	25	15	N
2	0	5	5	Y
3	1	10	9	N
4	0	53	53	Y
5	5	3	-2	N
6	100	61	-39	N
7	50	30	-20	Y
8	25	13	-12	N
9	0	50	50	Y
10	0	6	6	Y
12	30	35	5	N
13	1	23	22	N
14	0	15	15	Y
15	70	44	-26	Y
16	0	8	8	Y
17	70	15	-55	Y
19	0	47	47	Y
20	0	7	7	Y

Table 7.2.13. 18 NSCLC cases selected for mIF showing comparison of manually scored PD-L1 IHC (SP263) compared to mIF score for PD-L1, with a change in category taken as any difference resulting in tumour crossing clinical categories of negative, weak or strong positive (<1%, 1-49%, ≥50% TPS)

7.2.11 mIF compared to multiplex IHC for immune cell markers

mIF cell densities were compared to multiplex IHC cell densities for PD-L1, CD8 and CD68 for the 11 cases which underwent both sets of assays. Correlation was highly significant for PD-L1 (Correl. 0.867 $p<0.0001$, R^2 0.788 $p<0.0001$), CD8 in the TC (Correl. 0.909 $p<0.0001$, R^2 0.919 $p<0.0001$) and CD8 in the TE only (Correl. 0.891 $p<0.0001$, R^2 0.827 $p<0.0001$), but not for CD68 (Correl. 0.427 $p=0.190$, R^2 0.161 $p=0.221$). Each cohort was divided into 'high' and 'low' categories for each immune marker by the median value. Illustrative examples are shown in Fig 7.2.3. Grouping each case into 'high' or 'low' for each marker was compared between mIF and multiplex IHC, with agreement in 9 cases (81%) for PD-L1, 9 cases (81%) for CD8 in the TC, 8 cases (73%) for CD8 in the TE only, and 7 cases (64%) for CD68. Results are summarised in Table 7.2.14

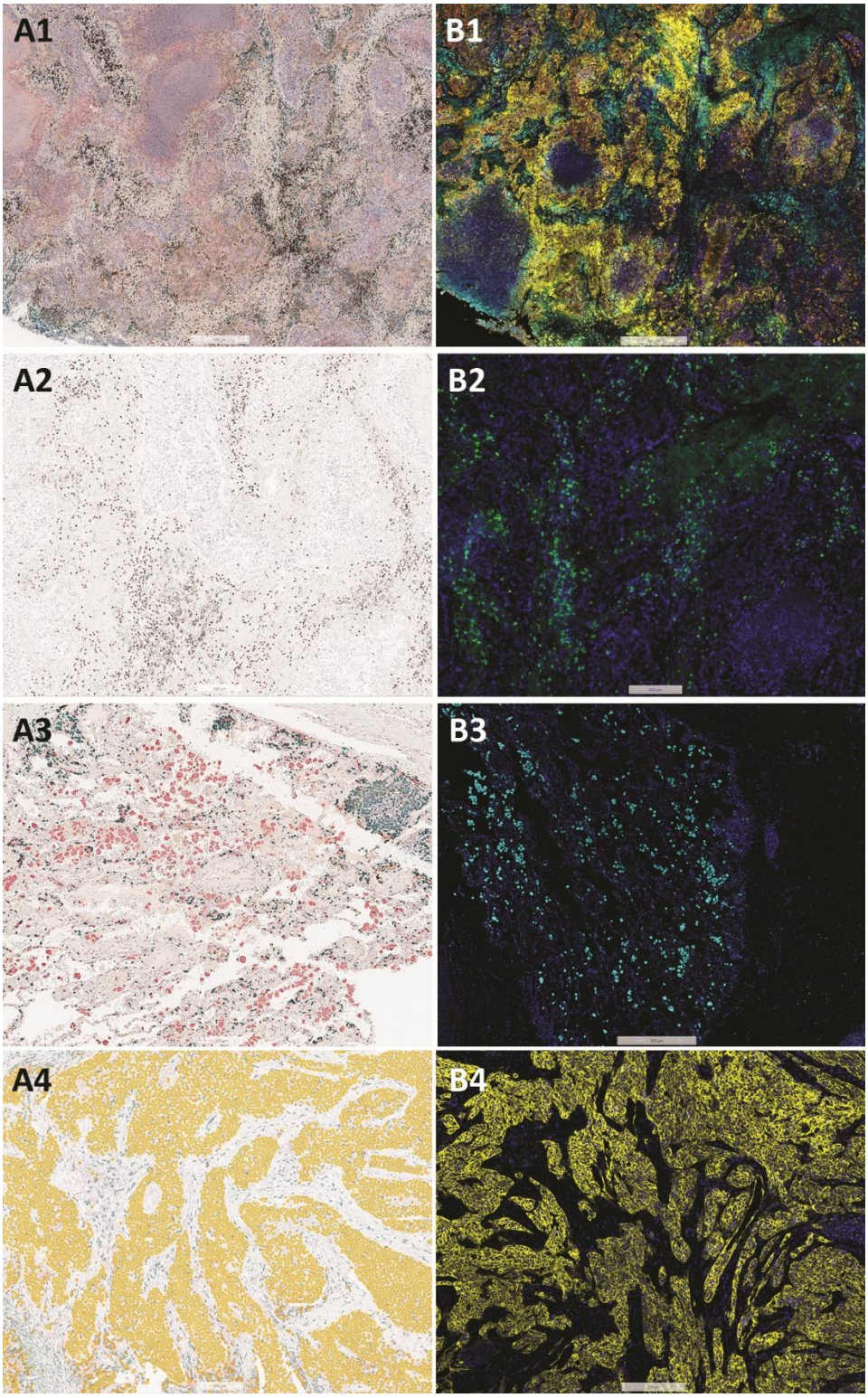


Fig 7.2.3. Multiplex IHC and mIF for immune markers. **A1-A4** Multiplex IHC, **B1-B4** mIF. **A1** – CD68, CD8, PD-L1 **B1** - DAPI (blue), pan-CK (orange) PD-L1 (yellow), CD8 (green) and CD68 (turquoise). **A2 & B2** – CD8 alone. **A3 & B3** – CD68 alone. **A4 and B4** – PD-L1 alone (Multiplex IHC has overlaid annotations for each cell (yellow)

	OPA	PPV	NPV	ICC	Significance
Both	0.39	0.44	0	0.554	N (0.053)
≥1%	0.39	0.55	0		
≥50%	0.72	0.33	0.8		

Table 7.2.14 Correlations and predictive values of image analysis of mIF for assessing PD-L1 expression compared to manual scoring of PD-L1 IHC (SP263). Improvement is achieved using a 1% or 50% dichotomous cut-off value only.

	PD-L1	CD8 TC	CD8 TE	CD68 TC
Average density IHC	1364	526	252	257
Average density IF	1632	290	185	232
Average Difference	461	-237	-67	-25

Table 7.2.15. Average densities and differences for immune cell markers by either multiplex IHC or mIF.

mIF produced both falsely high and falsely low PD-L1 expression in several cases. An example of falsely reduced PD-L1 expression relative to the monoplex PD-L1 (SP263) IHC stain is shown in Fig 7.2.4. H&E at higher power shows a number of tumour cells with appreciable understaining for IF Pan-CK, and a corresponding decrease in cells considered PD-L1 positive tumour cells within the pseudo-IHC stain for PD-L1. An example of falsely raised PD-L1 expression relative to the monoplex PD-L1 (SP263) IHC stain is shown in Fig 7.2.5. Low power H&E shows extensive necrosis throughout the section, with necrotic and non-viable tumour cells, as well as cellular debris staining for PD-L1 via the mIF, with a corresponding falsely raised PD-L1 in the pseudo-IHC stain for PD-L1. Extensive necrosis was a common feature for all cases with a falsely raised PD-L1 expression level by mIF. Aberrant expression of PD-L1 by necrotic cells is a well-documented cause of false positive PD-L1, as illustrated in Fig 7.2.6.

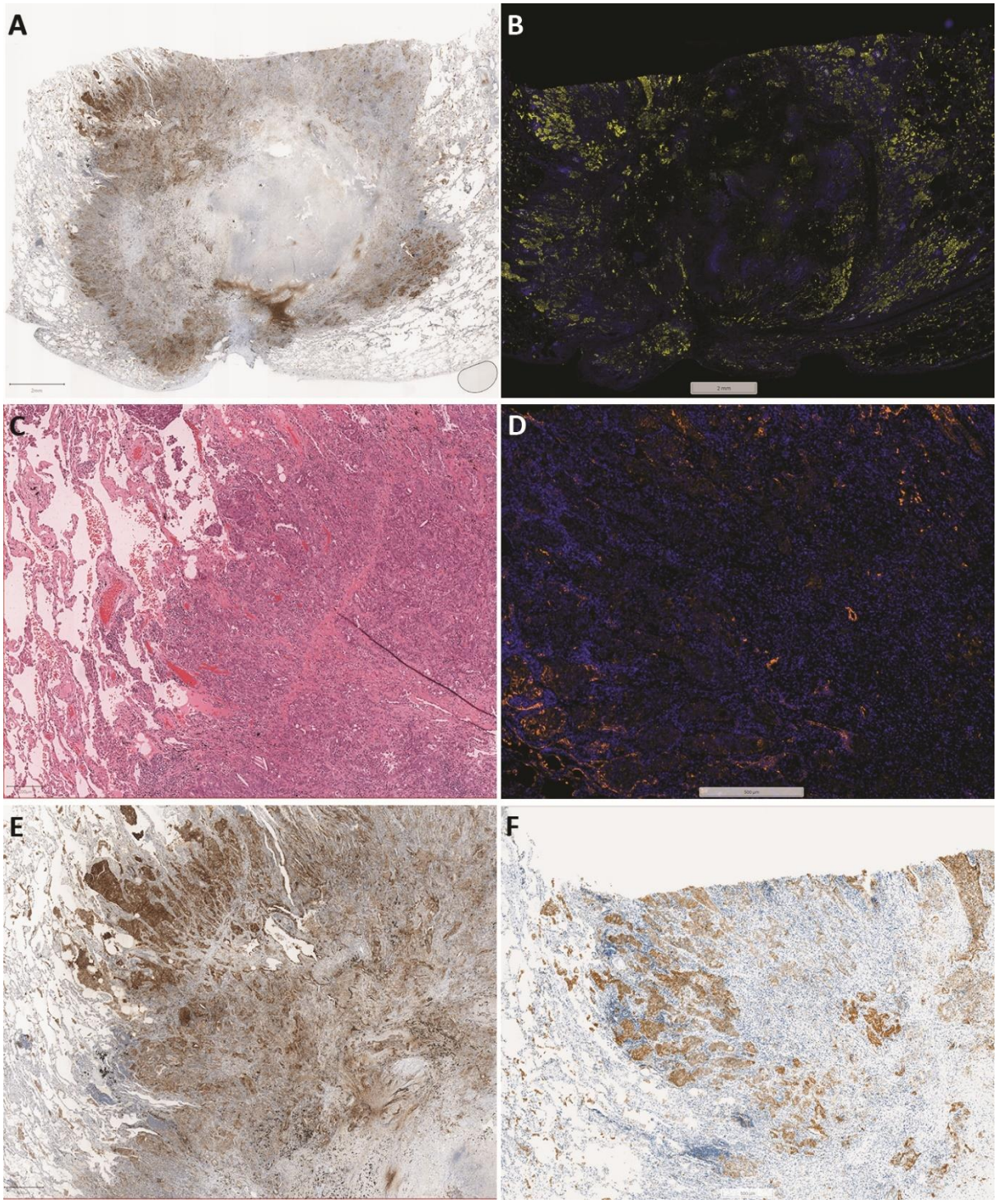


Fig 7.2.4. Under-expression of PD-L1 by mIF. **A** – Low power of PD-L1 IHC (SP263) showing widespread PD-L1 expression around a core of necrotic tumour. **B** – mIF showing PD-L1 (Yellow) with a similar distribution to that in A. **C** – H&E view of tumour showing tumour cells at invasive margin. **D** – The same area as C with DAPI positive cells widely apparent (blue) but with a large number of the tumour cells not staining for pan-CK (orange). **E** – PD-L1 IHC stain for same area showing mostly positive tumour cells. **F** – Pseudo-IHC for PD-L1 from mIF at the same area showing fewer PD-L1 positive tumour cells.

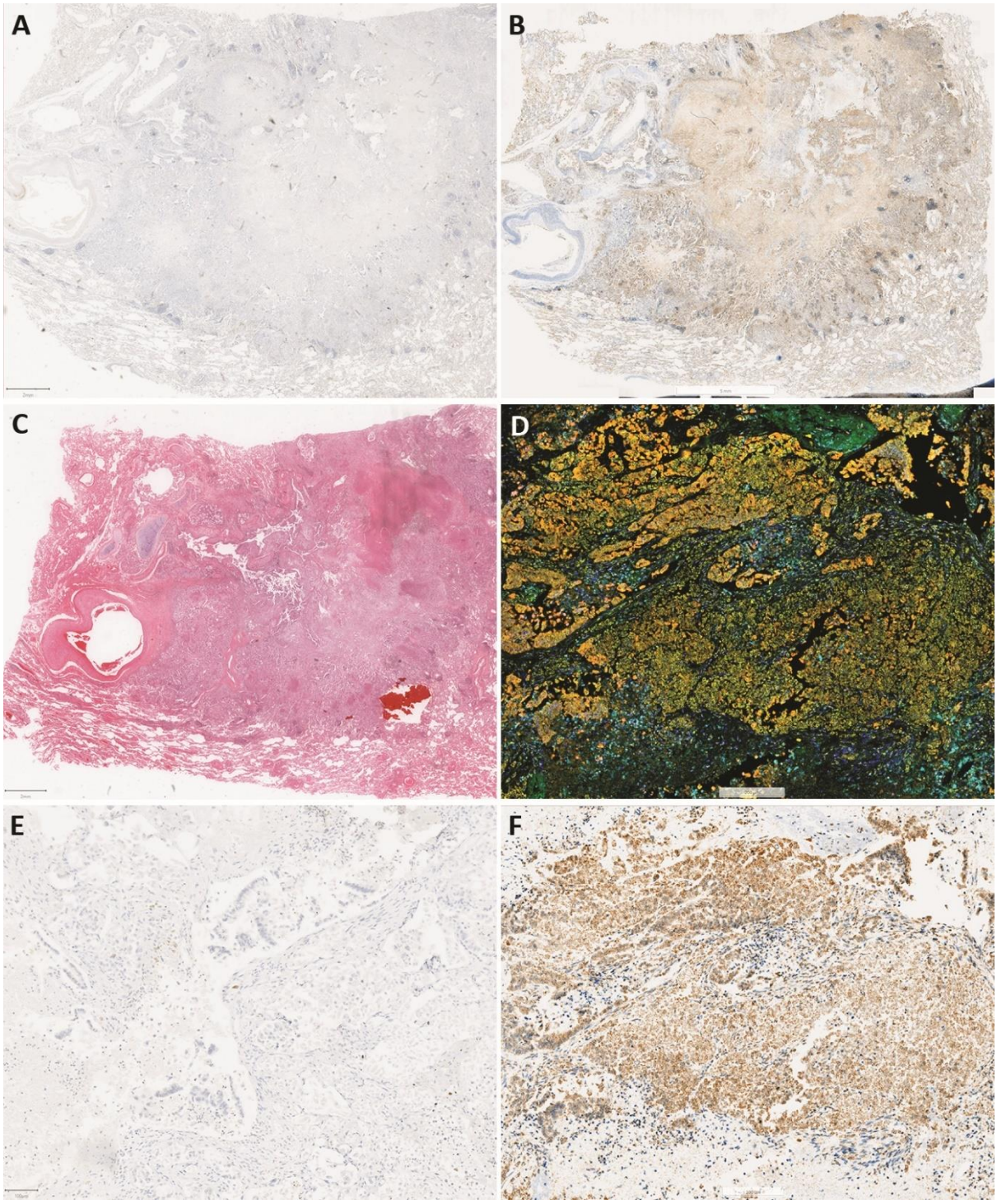


Fig 7.2.5. Over-expression of PD-L1 by mIF. **A** – Low power view of PD-L1 IHC (SP263) showing a largely negative tumour for PD-L1 expression. **B** – Pseudo-IHC for PD-L1 from mIF showing widespread false positivity. **C** – H&E view of tumour showing widespread necrosis. **D** – mIF area of necrotic cells and debris staining positive for PD-L1 (yellow) **E** – PD-L1 IHC stain for same area as D showing negative tumour cells. **F** – Pseudo-IHC for PD-L1 from mIF for same areas as D showing mostly positive 'tumour cells'.

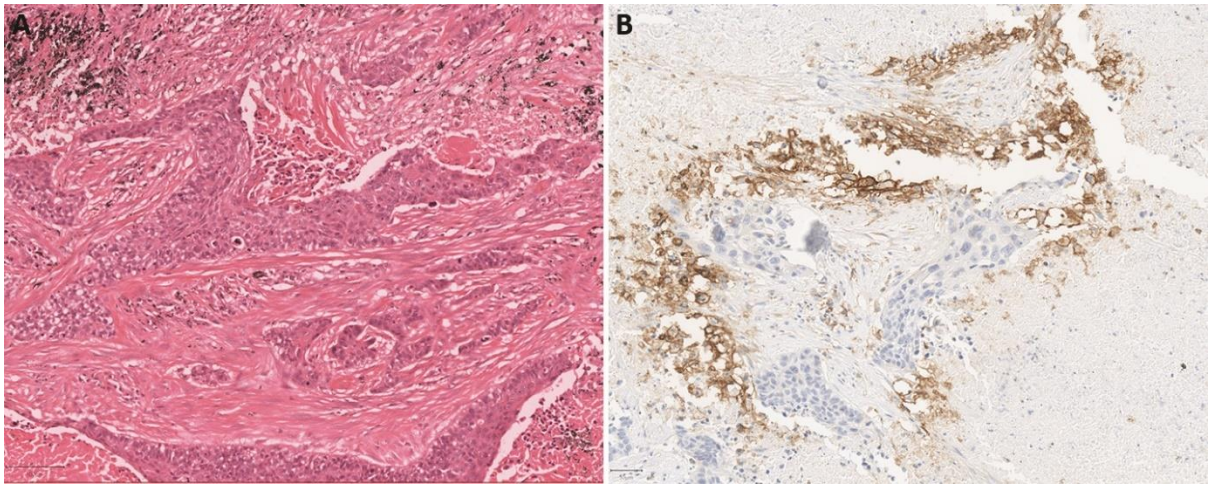


Fig 7.2.6. Aberrant PD-L1 expression in necrosis **A** – H&E of NSCLC demonstrating viable tumour in close proximity to fibrotic and necrotic tumour cells. **B** – Same area showing the viable tumour cells negative for PD-L1, with necrotic and non-viable tumour cells aberrantly expressing PD-L1 (SP263)

7.2.12 Classification of TMEs by mIF

Tumours were classified into TIL status TME as per Fig 7.1.5 (using CD8 alone) by mIF and by PD-L1 status (Table 7.2.16). There was no significant difference in CD68 levels between the TMEs (131 vs 151 vs 114 p=0.669). There was a non-significant trend towards increased inter-tumoural heterogeneity for infiltrated TMEs (TPS difference 21 vs 0.5 vs 3 p=0.248) and a non-significant trend towards decreased intra-tumoural heterogeneity for infiltrated TMEs (COV 126 vs 361 vs 181 p=0.253)

Survival analysis found immune desert TMEs had a significantly poorer predicted OS than immune excluded and immune infiltrated TMEs (21 vs 83 vs 70 χ^2 value 6.383 p=0.041).

TME	Cases	PD-L1	Desert (%)	Excluded (%)	Infiltrated (%)
Desert	7	**Negative**	5 (71)	1 (50)	2 (20)
Excluded	2	**Weak**	1 (14)	1 (50)	4 (40)
Infiltrated	9	**Strong**	1 (14)	0	3 (30)

Table 7.2.16. PD-L1 expression in 18 NSCLCs divided into clinical categories for different immune TMEs by mIF.

Tumours were classified by predicted response to PD-1/PD-L1 IM therapy as per Fig 7.1.7 (CD8 only, FoxP3 could not be utilised) by mIF and PD-L1 status. The good predicted response category had significantly more inter-tumoural PD-L1 expression heterogeneity than the no response group (TPS

change 45 vs 0 p=0.019), and trended towards less intra-tumoural heterogeneity than the partial and no response groups (27 vs 210 vs 189 p=0.085). There was no significant difference between the groups for CD68. Results are summarised in Table 7.2.17.

	Good response	Partial response	No response	Significance
Number	3	12	3	n/a
COV (Intra)	27	210	189	N (p=0.085)
TPS change (Inter)	45*	7	0	Y (p=0.022)
Group Change	1	4*	0	Y (p=0.026)
CD68	260	219	185	N (p=0.342)

Table 7.2.17. Comparison of different groups of NSCLCs categorised by potential response to PD-1/PD-L1 IMs for intra-tumoural and inter-tumoural PD-L1 expression heterogeneity and immune cell markers by mIF. * denotes significantly higher value.

Survival analysis found the good response trended non-significantly towards a superior outcome than partial and no response groups (88 vs 48 vs 46 months OS respectively, χ^2 value 2.153, p=0.341)

7.3 Discussion

The TME is a vastly complex biological entity, comprised of multiple different components, including the tumour cells themselves as well as a large number of other features that include the 'background' host tissue that the tumour cells have originated from or are invading into, and other cell types such as vascular and immune cells that may both be hindering or helping the growth of the malignant process. Categorising aspects of the TME to group them by type or behaviour is therefore a challenging process, and this chapter has attempted to focus on a limited number of immune cell types than are associated with both pro-tumoural and anti-tumoural activity. Current attempts to categorise the immune component TME of tumours include the pattern of TILs, the use of immune cell markers and genetic signatures.[594, 595, 598] However, an inherent problem when dealing with an environment as complicated as the TME is the difficulty in striking a balance between simple, easily useable and clinically relevant groups, that have a risk of over-fitting or misclassifying tumours that don't neatly fit into a small number of categories, and the use of a sliding scale or spectrum of properties that may more precisely characterise the tumours, but can result in too complicated an approach for routine clinical application. This project has explored the ability of defining TMEs by various immune cell markers, and the pros and cons of using multiplex data and image analysis approaches to do so, with a particular focus on understanding and categorising different types of PD-L1 expression, with the ultimate objective of refining complex polyfactorial data to a small

number of measurable data that still reflects the underlying TME but maybe translatable to the clinical setting.

7.3.0 Immune TMEs and their relationship to PD-L1 expression heterogeneity

In this chapter, several different approaches to categorising TMEs using a mixture of PD-L1 expression and immune cell markers was made, generally with a good overlap of similar characteristic TMEs regardless of the method used. A consistent observation was that increased immune activity, defined as infiltrating TILs, cytotoxic activity and/or reactive PD-L1 expression is associated with increased inter-tumoural PD-L1 expression heterogeneity but decreased intra-tumoural heterogeneity of PD-L1 expression. Conversely, low minimal immune activity, defined as immune desert, immune excluded, immunosuppressive and/or constitutive PD-L1 expression were generally associated with increased intra-tumoural, but reduced inter-tumoural PD-L1 expression heterogeneity. Overall this suggests that when PD-L1 expression is being used as an active mechanism of immune escape, it has a greater ability to up or downregulate PD-L1 expression in metastatic cells in response to variations within the metastatic TME compared to the primary lung tissue. This is also true even for tumours with reactive PD-L1 expression at low overall levels. Previous work has shown that the majority of PD-L1 expression in NSCLC is reactive, rather than constitutive,[563, 564, 566] explaining why a higher number of PD-L1 positive tumours are characterised as immune active. It has also been shown that immune and genetic markers can be down or upregulated as tumours metastasise, indicating that this is a common approach to survival for tumour cells when infiltrating different tissue types,[577, 578, 580, 626, 627] and that the immune properties of the TME can affect metastatic potential of tumour cells.[628-630] Macrophages in particular are associated with pro-metastatic behaviour[631] and in this project the more immune-active tumours are typically associated with higher levels of CD68. All tumours in this cohort are metastatic, however, so CD68 may well be at higher levels within this cohort than a mixed or non-metastatic NSCLC cohort, and CD68 alone cannot differentiate between M1 and M2 macrophage subtypes, which have differing roles in tumour promotion, but TAMs are typically thought to have pro-tumour M2 phenotypes,[130, 132] and thus the high overall levels of macrophages in the tumours with most inter-tumoural PD-L1 expression heterogeneity suggests macrophages have both a pro-metastatic function and a potential impact on the metastasising cells to regulate PD-L1 expression. Studies have shown the role in B-cell derived interleukins and neutrophils in contributing towards metastases,[632, 633] and whilst neither of these immune cell populations were studied directly in this project, it further suggests that specific TME immune functions can effect metastatic potential. The observation that tumours with an active and anti-tumoral immune TME more frequently demonstrate larger amounts of PD-L1 expression inter-tumoural heterogeneity suggests the immune

TME is not only important for metastatic potential *per se*, but also the behaviour and interaction of the tumour cells when they have metastasised.

Interestingly, when PD-L1 expression is reactive there is less intra-tumoural heterogeneity than when PD-L1 expression is constitutive. The former is likely explained by the fact that when PD-L1 expression is being successfully deployed as an immune escape mechanism, the tumour cells without the ability to express PD-L1 are eradicated, and the remaining cells are therefore more consistent in this expression. Eradication of tumour cells without favourable properties in an immune active TME has been previously described, resulting in a dominance of clones with a particular feature.[578, 579, 634] At low levels of PD-L1 expression this is still observed, suggesting even when PD-L1 is only part of the mechanism of immune escape, and only required at low levels, the ability of the tumour cells to increase PD-L1 expression may still be active, and indeed PD-L1 expression has been to shown to be upregulated in tumour cells in response to immune stimuli.[136, 196]. The high levels of intra-tumoural PD-L1 expression heterogeneity seen when PD-L1 expression is generally constitutive is probably due to PD-L1 expression not being driven by the immune process but by genomic changes within the tumour cells, which would understandably see no relationship to immune cell markers. Considerable intra-tumoural genetic heterogeneity has been observed in a wide variety of tumours,[547, 635] and if survival benefit is only minimally driven by anti-tumoural immunity, the more random distribution of PD-L1 is likely a result of this genetic diversity. Tumours with constitutive expression would also be less likely to change PD-L1 levels in response to differing TMEs within the lymph nodes, and would explain the lower inter-tumoural heterogeneity observed.

7.3.1 TMEs and their clinical relevance

PD-L1 expression can be thought of as reactive or constitutive, with relevance to tumoural heterogeneity as described, but a more clinically orientated categorisation might be thought of in terms of potential response to PD-1/PD-L1 IMs. In this project, patients with a potentially good response to PD-1/PD-L1 IMs are defined as having tumours with reactive PD-L1 expression in the context of a suitably active immune response targeted against the tumour, and for PD-L1 expression to be the major mechanism of immune escape. Non-responders are either a result of no PD-L1 expression, wholly constitutive PD-L1 expression, or insufficient/no appropriate immune response targeted against the tumour, which may be due to immunosuppressive factors or lack of presenting antigenic stimulation. Partial responders to treatment could be a mix of these features: a poor or partially inhibited immune response, where PD-L1 expression is weak or has constitutive components, where PD-L1 is only one mechanism of immune escape, if PD-L1 is achieving immune escape through suppression of cytokines rather than through direct T-cell inhibition, or a

combination of these features. By combining different properties of the immune signatures within the TMEs, patients in this cohort was thus divided into potential responders, and it has been previously demonstrated that good responders to PD-1/PD-L1 IM therapy have been shown to have specific immune and genetic signatures.[636]

By utilising these categories to divide this cohort of patients, a trend towards improved overall survival is seen, in keeping with previous observations that immune active, anti-tumour TMEs will result in a more favourable prognosis.[603, 604] Interestingly, a significant improvement in predicted OS was seen when using just two features within the TME: a high PD-L1 in the context of high CD8+ve T-cells within the tumour epithelium. At its simplest, this suggests the presence of an active infiltrating cytotoxic immune response with PD-L1 expression as the main mechanism of immune escape conveys a superior outcome for patients even without IM treatment. This does not necessarily translate to the best predictive power of response to PD-L1 IMs however; high PD-L1 levels even in TMEs infiltrated by CD8+ve T-cells may have dysfunctional T-cells, and conversely PD-L1 IM may be successful even in the presence of high quantities of Tregs and low Granzyme B, because even moderately exhausted T-cells can be stimulated by PD-1/PD-L1 IMs.[246] However, on average, high CD8+ve and high PD-L1 expressing tumours would be expected to respond well to PD-1/PD-L1 IMs, and if this phenotype is already associated with superior outcome, it may suggest why a subset of NSCLC patients respond so favourably to these therapies.[28, 637]

The use of a similar approach utilising a smaller number of immune markers and interpretation of H&E slides could thus be used to classify patients into similar potential responders to therapy in a fashion that does not involve the complicated multiplex assays and digital image analysis.

7.3.2 Assessment of multiplex and digital image analysis

A practical problem with multiplex approaches is the interpretation; mIF fluorophores can be relatively easily quantified by an automated approach, but the quantification of chromophores is more challenging.[339, 623, 624] In addition, components of the TME must be studied in spatial relationship to each other; knowing the total cell density of CD8+ve T-cells is not sufficient; where the T-cells are in relation to other cells is also necessary. The use of digital image analysis approach to provide these metrics was used to help overcome this, but relatively novel approaches must be treated with caution. Even simple monoplex approaches can see considerable variation: PD-L1 IHC can be vary between clones, assays and platforms,[266, 331, 332] and so it seems reasonable to assume more complex multiplex technologies may also suffer a lack of concordance.

Despite these potential issues, the comparison of data provided between multiplex IHC and mIF in this project was generally reasonable for PD-L1 and CD8 by overall cell densities, though the former were generally over-represented, and the latter generally under-represented by IF compared to IHC. CD68, a marker of macrophages, correlated very poorly between multiplex IHC and IF however. Different clones for the same marker can be notoriously difficult in achieving concordance, and this may account for some of the differences seen between IHC and IF. In addition, despite a good correlation of PD-L1 expression between IHC and IF in terms of cell density, the ability of the mIF to provide equivalent TPS to the manual scoring is poor. This is probably partially due to limitations in the mIF approach itself: necrotic and fibrotic areas in particular seemed to return false positives, and false negatives may be a result of tumour cells that appeared to be aberrantly negative for pan-CK in some cases (Fig 7.2.4 and 7.2.5). Indeed false positives for IF dyes is a well-documented area,[638] as are false positives for PD-L1 IHC in necrotic cells.[282] Converting manual cell densities to TPS equivalent to clinical categories was also problematic; multiplex IHC for PD-L1+ve, CD8/CD68-ve cells provided a reasonable surrogate TPS when compared to manual monoplex scoring, but generally underscored TPS when there was a difference and could not reliably classify tumours by clinical categories, though this is partially likely a result in keeping with the general observation that SP142 stains fewer tumour cells than SP263.[297, 304] mIF was also poor at classifying tumours into clinical categories by the same method. When considering the TPS scores provided by digital image analysis of the monoplex IHC compared to manual scoring of the same, however, this showed a very high correlation by both overall TPS and clinical categories, suggesting the image analysis approach *per se* is an accurate one, but is only as good as the images provided.

A simple approach was thus adopted by using a simple dichotomous division of immune cell markers using the median from this cohort into 'strong' and 'weak' for each marker. By so doing a good level of concordance was seen when dividing TMEs by TIL patterns by either H&E or the multiplex IHC data, suggesting this approach to interpreting the data in this project is reasonable. The biggest limitation of this approach is lack of a secondary dataset to compare values to. Whilst the selected tumours represent a wide array of typical NSCLC morphologies, stages, demographic details and so forth, it is unknown if these are typical in terms of immune cell markers. Dividing the cohort into 'high' or 'low' for each immune marker is thus purely dependent on an internal measure, and it is possible several tumours would be classified differently if using median cut-offs determined from a different cohort. The difficulty presented in turning absolute cell values into a percentage score is similarly effected and would suggest a far larger set of data is required before determining absolute values for high or low that could be applied to any given sample. The other main limitation of this approach, and a general issue in this area, is the division of PD-L1 into categories based on

percentage cut-offs of 1% and 50%. Whilst this allows for relatively simple classification with predictive power in the routine clinical situation, PD-L1 expression is more likely a continuous variable and is more likely to have clinical implications when considering other factors within the TME than a TPS alone.[603, 639] As a result, when using clinical categories, tumours barely expressing PD-L1 at 1% are grouped with tumours expressing PD-L1 in nearly half of all cells, and tumours expressing PD-L1 on only half their cells are grouped with tumours expressing PD-L1 in all their cells, which intuitively feels like oversimplifying the situation. My approach has been to consider categorisation in a more dynamic fashion; weak PD-L1 expressers could therefore be grouped in with negative tumours or strong tumours, but whilst this more flexible approach may be more suitable for diving tumours into biological relevance, it is probably too ambiguous for the routine clinical setting.

7.3.3 Other findings

An unexpected observation was high CD8+ve T-cell and low Granzyme B TMEs has a superior OS compared to other TMEs. This would imply that patients with tumours harbouring exhausted TMEs survive longer than those with TMEs that are active and functional, whereas the opposite has been previously demonstrated in hepatocellular carcinoma.[640] However, only a small number of cases fit into this category (n=13) and when stratifying for PD-1 (high levels of which would also be anticipated in exhausted T-cells) the significance is lost, and when splitting tumours by Granzyme B in any other scenario, this observation is not maintained. This could therefore be a statistical anomaly.

Image analysis continues to be a reliable method of aiding the interpretation of tumours for immune cell markers and PD-L1 expression in general. The limiting factor is the quality and substance of the images provided. Multiplex approaches provide an attractive platform for performing image analysis, but have technical and financial issues associated with them, and the use of differing assays, molecules and platforms to look for immune markers makes their application in the routine clinical situation difficult to envisage in the foreseeable future. Importantly, however, they do provide reasonably consistent categorisation of TMEs, regardless of the approach used, and the findings that pertain to inter-tumoural and intra-tumoural PD-L1 expression heterogeneity, and general prognosis are consistently seen in both multiplex IHC and mIF in this project.

The challenge will be to apply a simplified version of the approaches developed in this chapter to a range of typical specimens used in NSCLC diagnostic and molecular profiling, in order to best predict clinical response to PD-1/PD-L1 IMs.

Chapter 8 – Predicting response to PD-1/PD-L1 immunomodulatory therapy

8.0 Introduction

The main objective of this thesis has been to identify ways in which NSCLC patients can be accurately assessed so as to predict their response to PD-1/PD-L1 IMs. As such, the main focus has been on aspects of PD-L1 IHC, the only clinically validated predictive biomarker for this purpose to date, and considered ways in which to optimise its use. However, the majority of the work has been performed on the LLP cohort; an excellent cohort of NSCLC patients with generous quantities of resected NSCLC tissue and for whom all of which have nodal disease and clinical outcomes recorded. The major limitation of the LLP cohort, however, is that none of the patients have been treated with PD-1/PD-L1 IMs. Therefore although I have been able to assess changes in expression of PD-L1 by IHC, and, where relevant, to consider the prognostic power of these findings, I have not been able to translate that to predicting response to IM therapy.

As a result, the data, outcomes and hypotheses generated thus far should be applied to the CCC cohort, which consists of NSCLC patients treated with PD-1/PD-L1 IMs and for which full clinical data is available. In so doing, an attempt to generate predictive data can be made, as well understanding the limitations of using the typically smaller specimens used in the routine clinical setting of diagnosing NSCLC patients. The approximately 300 patients in the CCC cohort represent 'real-world' data and are a relatively homogenous group in terms of being stage III/IV NSCLC (as per AJCC (American Joint Committee on Cancer) guidelines)[641]. Work performed on the CCC cohort has been limited to date, with several future studies planned (explored in more detail in Chapter 9 (The Next Frontier)), but an overview of the challenges of using this cohort, and the data of interest generated thus far, and how this fits with current evidence for PD-L1 IHC as a predictive biomarker will be given here.

8.0.0. PD-L1 cut-offs for predicting response

In the various PD-1/PD-L1 IM clinical trials for NSCLC, different PD-L1 IHC clones were used with different cut-offs to define patients as 'PD-L1 positive'. Despite sharing the same target of PD-1, (e.g. pembrolizumab, nivolumab,) or PD-L1 (e.g. atezolizumab, durvalumab) there remains a disparity of the value of using PD-L1 IHC to define patients as 'PD-L1 positive' from clinical trial data in NSCLC.[46-48, 254] This is reflected in the guidance of prescribing PD-1/PD-L1 IMs, in which PD-L1 may be a companion diagnostic, complementary diagnostic, or not required, depending on the specific IM and clinical status of the patient. In regards to defining PD-L1 positive patients, the UK advice from NICE provides guidance for the treatment of NSCLC with PD-1/PD-L1 IMs which uses the categories of

negative, weak and strong expressers defined as <1%, 1-49% and ≥50% TPS respectively.[567] European guidance from ESMO also uses the same groups in NSCLC[642], as does American FDA guidance[11] and Japanese guidance.[643] Other cut-offs are occasionally used, most typically ≥25%, but these other cut-offs do no always act as a powerful predictor of response.[644, 645]

These clinical guidelines are largely based on the observation that PD-L1 positive patients, defined as ≥1% TPS, are generally associated with superior outcomes with PD-1/PD-L1 IMs versus the standard treatment of care, with further benefit in 'high' PD-L1 expressers also seen.[47, 299] Furthermore, ≥50% TPS is generally used to define eligibility for IM therapy in the 1st line setting.[567, 646] The caveat that these patients are wildtype for the commonest targetable single driver mutations in NSCLC is also taken into consideration. [299, 647] However, in some clinical trials, PD-L1 expression was of minimal predictive value, and so PD-1/PD-L1 IMs can still be prescribed even for PD-L1 negative patients in certain circumstances.[260] These trials are discussed in more detail within Chapter 1, but importantly, there are no head-to-head trials comparing different PD-1/PD-L1 IMs, and so data is limited as to why PD-L1 positive patients vary in terms of response to treatment. This variation in response is likely to include a combination of clinical differences between patient cohorts, pharmacological variation between the IMs, factors that affect PD-L1 IHC and differing cut-offs to define 'PD-L1 positive'.

An important question therefore, is whether the current cut-offs of ≥1% and ≥50% TPS best reflect the biological mechanisms that underpin PD-L1 mediated immune escape, and therefore provide the best predictive power for grouping NSCLC patients as responders to IMs. As has been explored earlier in this thesis, a tumour of 0.5% and 1% PD-L1 expression are likely to have more in common from an immune escape perspective than a tumour of 1% and 45% PD-L1 expression, yet the latter pairing are typically grouped together and the former are not. The expression of PD-L1 can be thought of numerically (e.g. a TPS) but also biologically; for example, if PD-L1 is being expressed reactively or constitutively. As Chapter 7 (The tumour microenvironment) explored, the presence or absence of specific T-cell populations within a TME will likely be a helpful parameter in determining PD-L1 expression as reactive or not, but the feasibility of whether testing for these in the typically small tissue or cytology specimens used to diagnose NSCLC is a realistic possibility is an important aspect to consider, and one this chapter will attempt to address.

8.0.1 Measuring clinical response

Measuring clinical response to therapy is not necessarily a straightforward task. Traditionally clinical trials used OS as a measurement in cancer therapy, as this is an objective measure of response measured as a clear clinical endpoint (i.e. patient death), which takes into account the impact of

treatment and any irAEs with a generally agreed upon outcome.[648] However other endpoints have been increasingly utilised, including PFS which has a number of potential advantages over OS: a smaller sample size is required, follow-up is shorter, it is a more practical metric for disease states requiring very long follow-up, it may take better account of stable disease and could deal with patient cross-over more effectively, as well as providing a meaningful quality of life metric for treatment decisions.[305, 649] However, PFS may introduce bias, is harder to measure than OS and does not necessarily correlate with OS, particularly in advanced NSCLC or immunotherapy trials.[648-651] ORR (overall response rate) is another outcome measure that typically uses RECIST and iRECIST criteria to measure patient response[652, 653](Table 8.1.0) and is particularly useful in single-arm trials, as it requires no control arm for comparison.[648, 649] ORR is defined as the proportion of patients with complete or partial response to treatment. Other potential measures of clinical response in clinical trials include durable response, quality of life endpoints, surrogate/biomarker endpoints and milestone survival, with advantages and limitations to all.[648] It is outside the scope of this thesis to provide a comprehensive overview of clinical measure outcomes, but where 'real-world' data is concerned, a mixture of OS and PFS is typically used, with many papers also integrating clinical prognostic markers, such as ECOG PS (Eastern Cooperative Oncology Group performance status)[654] that has been shown to have a consistent effect on predicting survival in NSCLC patients treated with IM therapy.[655-657]

Predicting toxicity and irAEs is also an important outcome which would be hugely valuable to oncologists. It is generally well established that the occurrence of irAEs positively correlates with response to treatment by PD-1/PD-L1 IMs,[121, 658-662] and therefore a predictor of response to PD-1/PD-L1 IM might be a useful predictor of irAEs.

This chapter will therefore look to explore the feasibility and challenges of using a variety of non-resection specimens for predicting response to PD-1/PD-L1 IMs by the interrogation of the immune TME and PD-L1 expression. The challenge of using these 'routine diagnostic specimens' that generally contain far smaller quantities of tissue than resections specimens is a key translational issue. I will also attempt to measure clinical response to IM therapy by a multitude of factors, including OS, PFS, and measurements involving RECIST criteria, whilst utilising other metrics such ECOG PS, number of IM cycles received, irAEs, demographic data and so forth.

8.1 Methods

8.1.0 Analysis of CCC clinical outcomes

The CCC cohort included data gathered by oncologist colleagues from the CCC (Led by Dr Ana Ortega-Franco and Dr Carles Escriu) and includes demographic and clinical details as shown in Table 8.2.0. OS and PFS were used as a measure of response to PD-1/PD-L1 IM therapy, with irAEs and the specific grade of irAE were used as a measure of toxicity. Outcomes and events were defined as per Table 8.1.0.

Outcome/event	Definition
OS	Time between date of metastatic disease diagnosis and date of death.
PFS	Time between date of metastatic disease diagnosis and last date of no detectable progression of disease by clinical or imaging measures.
Best response:	Best response during treatment according to Recist 1.1:[652]
CR	Complete response: Disappearance of all target lesions. Involved lymph nodes require reduction in short axis to <10mm
PR	Partial response: At least a 30% reduction in the sum of target lesions
PD	Progressive disease: At least a 20% increase in the sum of target lesions with a minimum absolute increase of ≥5mm
SD	Stable disease: No change or change insufficient to meet criteria for PR or PD.
ORR	Overall response rate: the proportion of patients with CR or PR.
Rapid Progression	Decrease in ECOG PS* of ≥1 point within 4 weeks of commencing IM therapy
Hyperprogression	Increase of 40% of sum of target lesions with a minimum absolute increase of ≥10mm **or** an increase of 20% with new metastatic lesions
IM cycles	The number of IM cycles received by patient before termination due to toxicity, death or other adverse event.
irAE	Immune related adverse events of any grade toxicity
Grade 3/4 Toxicity	Severe, significant or life threating AEs requiring intervention as per NCI guidelines

*ECOG PS = Eastern Cooperative Oncology Group performance status[654]

Table 8.1.0 Definitions of outcomes and events used as markers of clinical response or measures of toxicity in NSCLC patients treated with PD-1/PD-L1 IM therapy.

8.1.1 Analysis of survival data in the CCC cohort

In order to best match cases, only patients receiving pembrolizumab were included. All patients were Stage III/IV (as per AJCC guidelines), were wild-type for EGFR and ALK and were considered as both 'whole cohort' and divided into a sub-cohort of patients receiving IM either in 1st line setting or in the 2nd/3rd line setting ('2nd-line plus').

First, comparison of OS and PFS within the CCC cohort using the conventional 'weak' (1-49%) and 'strong' (≥50%) TPS groups was performed. OS and PFS was then looked at using a ≥25% TPS cut-off to define 'PD-L1 positive'. Due to the observations pertaining to the 2nd-line plus cohort, OS and PFS relationship to PD-L1 expression in the 2nd-line plus cohort was then looked for in a series of exploratory analyses by using 10% PD-L1 TPS incremental groupings to either divide the cohort dichotomously (≥10% as 'positive', then ≥20% as 'positive', then ≥30% as 'positive' etc.) or into 10% PD-L1 TPS groups (1-10%, 11-20%, 21-30% etc.). Predicted survival data and hazard ratios for each group was thus captured and analysed.

To ascertain the impact of specific features that may account for differences observed between OS and PFS, both the entire CCC cohort and the cohort divided by 1st-line or 2nd-line plus IM treatment were assessed for the association between survival data and PD-L1 expression with irAEs, ECOG status and the number of IM cycles received. Multivariate linear analysis which included PD-L1 expression, number of IM cycles, ECOG status, gender, patient age, tumour morphology, irAEs and whether IM therapy was received 1st line or 2nd-line plus was then looked for in regards to predicting survival outcome.

8.1.3 Pathological analysis of CCC cohort

For cases from the CCC cohort which had H&E and PD-L1 sections available, these were assessed for PD-L1 TPS and for TIL infiltration pattern as per Chapter 7 (Desert, excluded and infiltrated TMEs). Grouping these cases by TME on H&E assessment into immune TME subtypes was dependent on the type and site of specimen. In generous biopsy samples of primary tumour, the presence of lymphocytes mixed throughout the tumour cells marked the sample as immune infiltrated (Fig 8.1.0 A1-A2), whereas lymphocytes confined to the stroma marked the sample as immune excluded (Fig 8.1.0 B1-B2). A generous biopsy of tumour with minimal immune presence could indicate an immune desert (Fig 8.1.1 A1-A2), but a small biopsy, deficient in tumour stroma, whilst technically sufficient to diagnose a patient and assess PD-L1 expression, is likely uninformative and unrepresentative of the rest of the tumour in regards to its TME (Fig 8.1.1 B1-B2). Biopsies of metastatic deposits which have not formed a discrete tumour mass could also not be accurately subtyped into immune TMEs, but an attempt to quantify some type of immune response can be made in generous samples from these sites: for example, in pleural biopsies, infiltrating tumour cells may elicit no apparent immune response, or be associated with lymphoid aggregates close to their proximity (Fig 8.1.3). For metastatic deposits which form clear tumour masses, immune TME classification may be determined with sufficient sampling (Fig 8.1.2). In samples from lymph nodes, the presence of lymphocytes is to be expected, and therefore their presence is not helpful in determining an immune TME type. Cytology specimens from sites other than lymph nodes, such as

pleural fluid, may also be difficult to classify an immune TME due to the admixing of tumour cells, immune cells and mesothelial cells (Fig 8.1.4). As such, subtyping of immune TMEs was limited to samples of primary NSCLC with preserved architecture and sufficient quantities of tumour stroma. Specimens that do not fit these criteria, including cytology specimens, metastases and small biopsies with limited tissue were categorised as being either unsuitable for immune status categorisation (e.g. lymph node aspirate), or divided into 'lymphocytes present' or 'lymphocytes absent'.

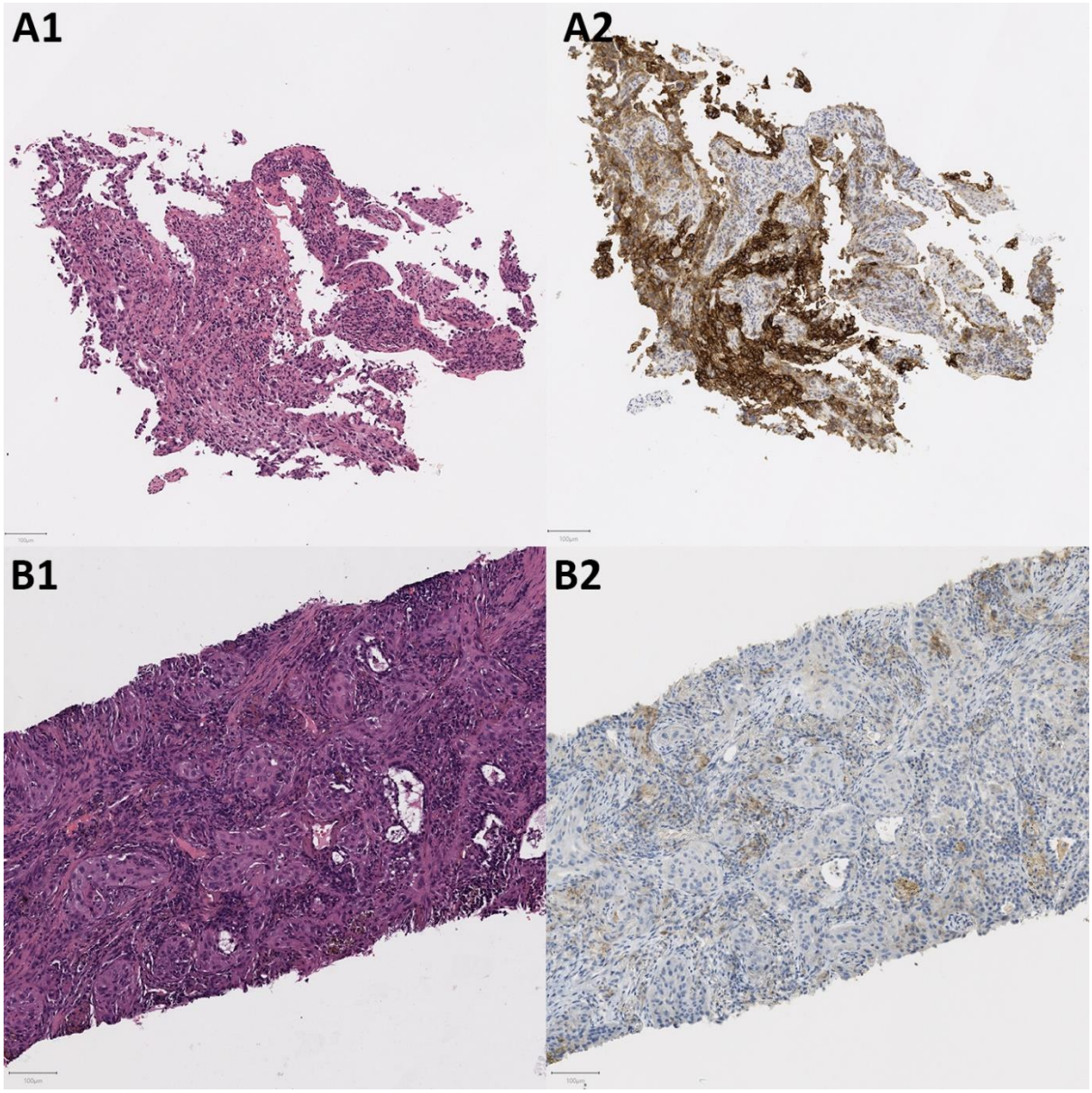


Fig 8.1.0 Examples of NSCLC biopsies for which immune TMEs can be accurately depicted. **A1-A2 -** Infiltrated TME on H&E (**A1**) and PD-L1 (22C3) (**A2**). Note the admixing of TILs within the tumour epithelium. **B1-B2 -** Excluded TME on H&E (**B1**) and PD-L1 (22C3). Note the TILs are confined to the stroma of the tumour. PD-L1 expression is limited to the TILs.

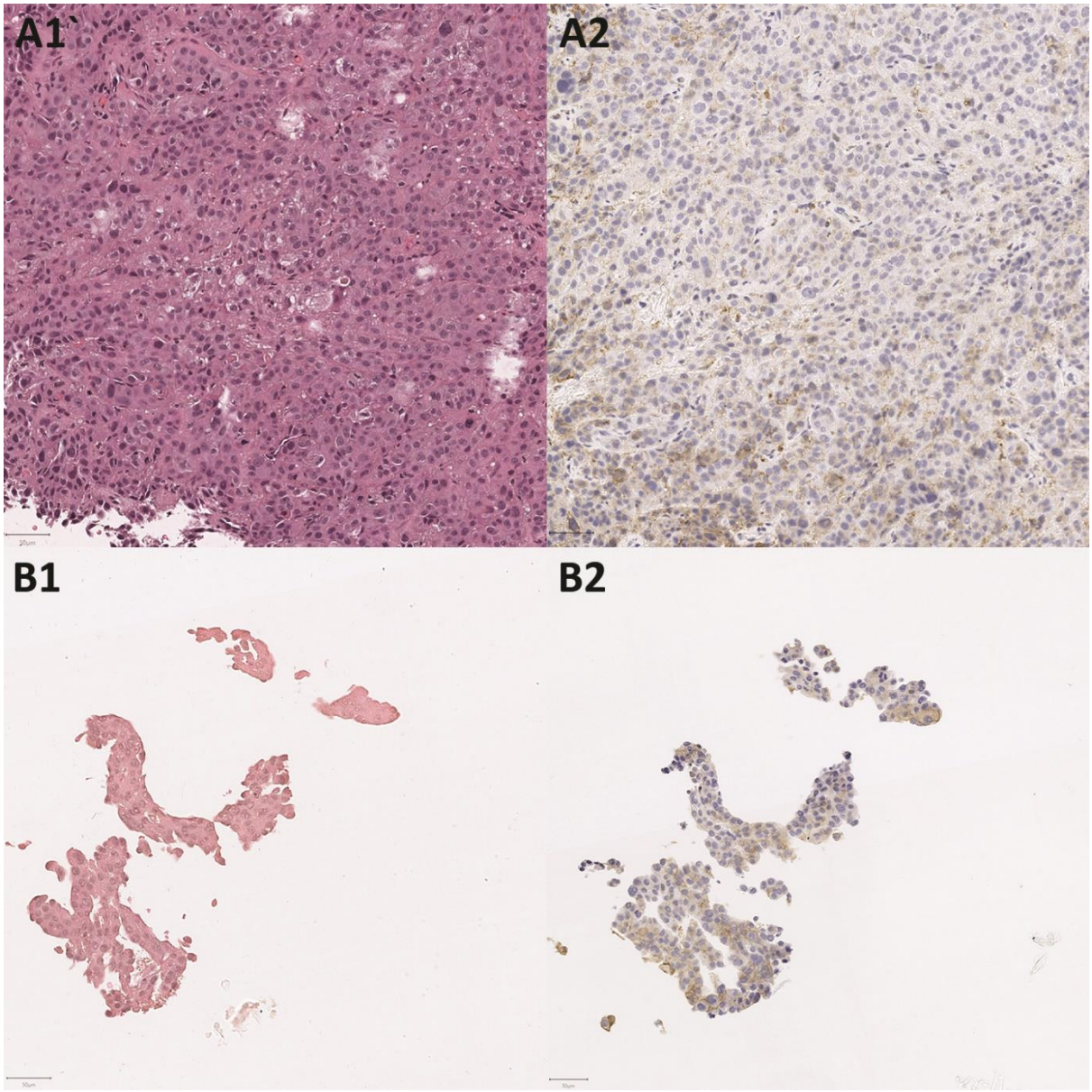


Fig 8.1.1 Examples of NSCLC biopsies with minimal immune cell presence. **A1-A2** – Generous biopsy with minimal TILs can be subtyped as an immune desert TME on H&E (**A1**) and PD-L1 (22C3) (**A2**). **B1-B2** – Minimal tissue in small biopsy with no tumour stroma by H&E (**B1**) and PD-L1 (22C3) (**B2**). Despite being technically sufficient to make a diagnosis and to analyse PD-L1 expression, there is insufficient tissue to accurately define the TME as an immune desert due to the absence of tumour stroma.

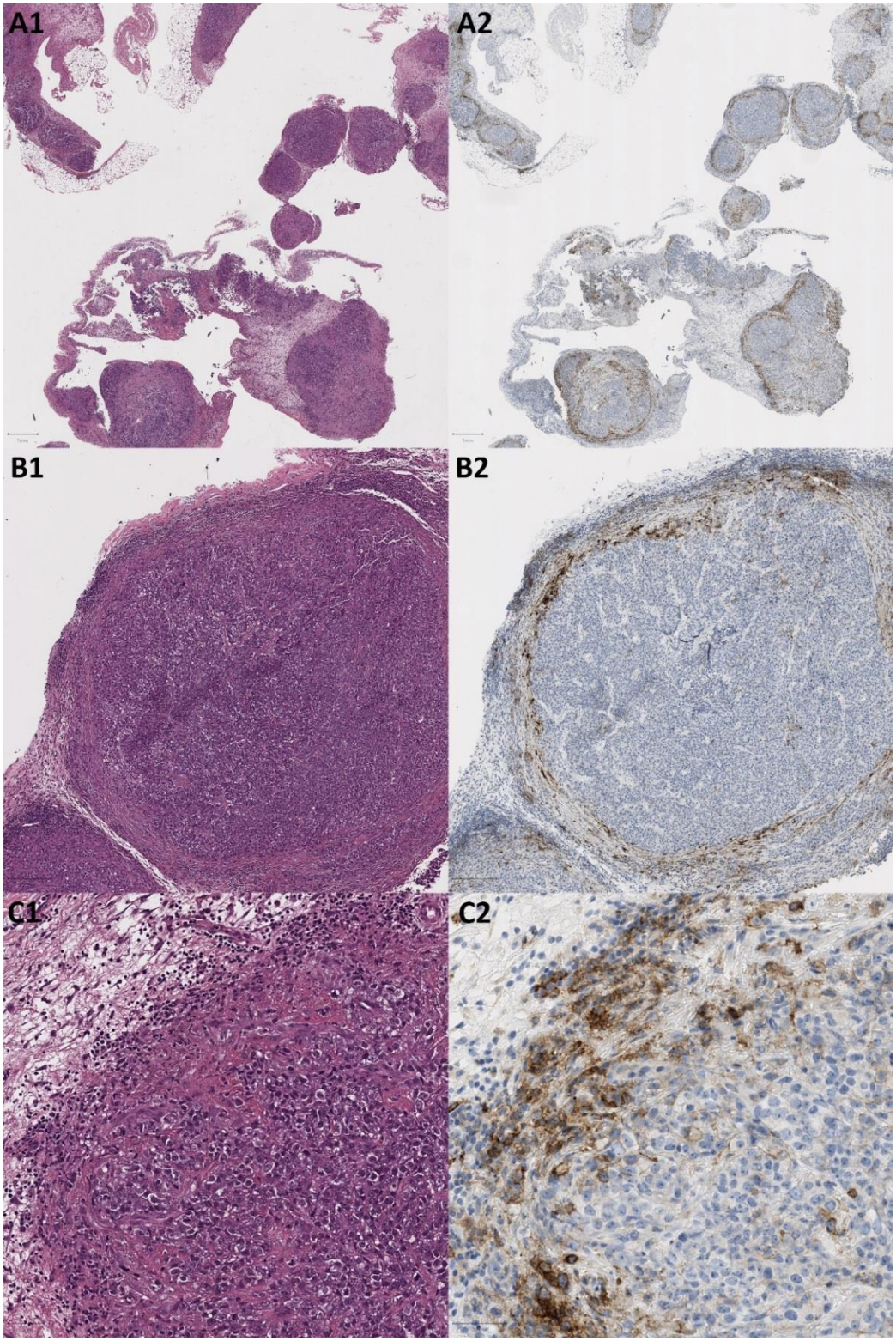


Fig 8.1.2 Examples of pleural biopsy with striking example of immune exclusion and PD-L1 expression at the interface. **A1-C1** – H&E **A2-C2** – PD-L1 (SP263) expression. **A –** Low power view showing every deposit of metastatic NSCLC demonstrates the same pattern of immune exclusion. **B-C** – Higher power views. PD-L1 expression is mostly in TILs but there is also expression of PD-L1 in tumour cells at the interface between the tumour and excluded immune cells.

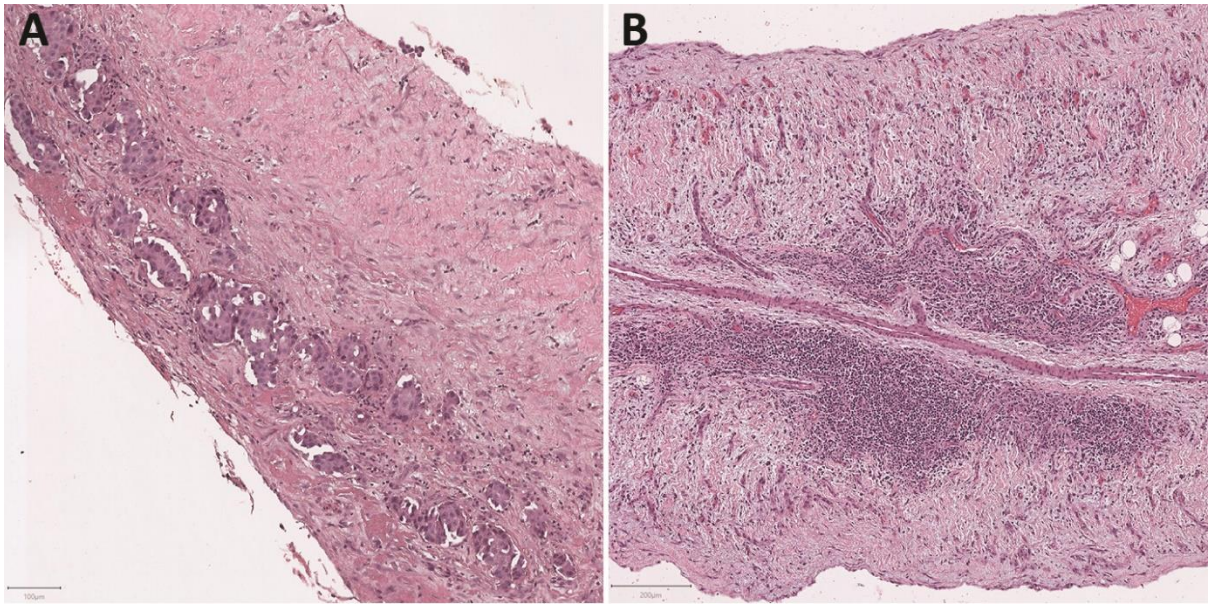


Fig 8.1.3 Example of pleural biopsy samples with infiltrating NSCLC metastases. **A** – Groups of tumour cells eliciting no apparent immune response. **B** – Infiltrating tumour cells with dense areas of lymphocytic infiltrate nearby.

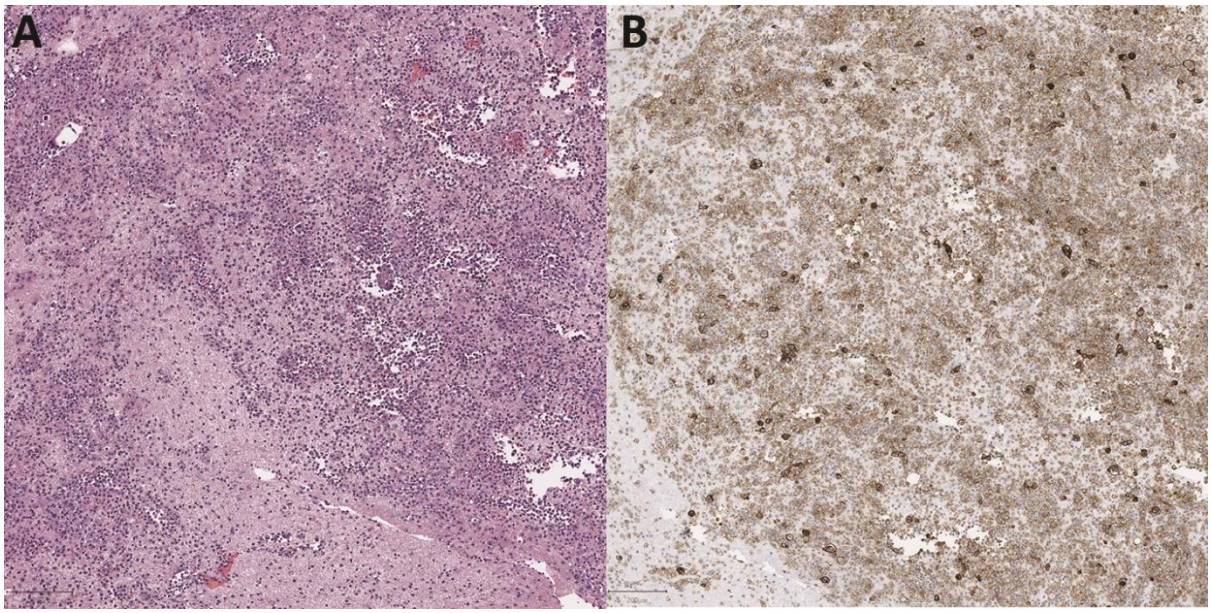


Fig 8.1.4 Admixed tumour and immune cells from pleural fluid sample by H&E (**A**) and PD-L1 (22C3) (**B**). This specimen is not suitable to classify the TME due to admixing of cells and loss of tissue architecture.

8.1.4 Statistics

Associations between PD-L1 expression as a continuous variable and outcome/events as defined in Table 8.1.0 was looked for by Spearman's correlation and linear regression. Use of PD-L1 as a categorical variable was compared using cross tabulation and chi-squared analysis with Bonferroni's corrected post-hoc analysis for multiple groups. Survival analysis was performed using Log Rank statistics to compare factors within Kaplan Meier analysis, and Cox Regression to generate hazard ratios. Forest plots for hazard ratios were generated for exploratory analyses of novel PD-L1 categories. Univariate linear regression was performed in SPSS with subsequent multivariate linear regression to determine significant factors for the dependent variable (survival data). R^2 values, β co-efficient values and p values are acquired from this method. All significances are taken as $p<0.05$.

8.2 Results

8.2.0 Demographic and pathological details of the CCC cohort

A total of 308 patients from the CCC cohort were initially available. A total of 31 cases were excluded from further analysis as per exclusion criteria (PD-L1 negative (20) or insufficient tissue for PD-L1 IHC (11) and/or IM therapy other than pembrolizumab (15) and/or harbouring EGFR mutations (4)). All remaining 277 patients were stage III or stage IV (all types) as per AJCC guidelines.[641] Finally the cases were considered as both whole cohort, or divided into patients whom had received pembrolizumab first line (107 patients) and those whom had pembrolizumab in the second or third line setting following XCT (170 patients). A summary of the CCC cohort is shown in Table 8.2.0. For further tissue analysis, 168 blocks across 131 patients were available for retrieval, with 72 cases having H&E and PD-L1 (22C3) slides available for review.

8.2.1 Assessment of PFS and OS in CCC cohort by conventional cut-offs

The OS and PFS for the entire CCC cohort divided into patients with tumours being weak (1-49% TPS) or strong (≥50% TPS) expressors of PD-L1 is shown in Fig 8.2.0 and 8.2.1. The cohort defined as tumours with strong PD-L1 expression had a significantly superior OS (15.7 vs 12.3 months p=0.005) and PFS (12.9 vs 8.1 months p<0.001) than weak PD-L1 expression.

	N	%
Median age (at diagnosis)	69 (Range 33-87)	
Gender		
Male	147	53
Female	130	47
ECOG Status		
0	72	26
1	136	49
2	64	23
3	5	2
OS status - dead	188	68
Rapid progressors	48	17
Hyperprogressors	12	4
Mean number of IM cycles	6 (range 1-40)	
Pembrolizumab 1st-line	107	
Pembrolizumab 2nd-line	170	
irAE (any grade)	113	41
Grade 3/4 Toxicity	35	13
Morphology		
ADC	176	64
SCC	91	33
Other	10	4

Table 8.2.0 Demographic and pathological details of the 277 included patients from the CCC cohort.

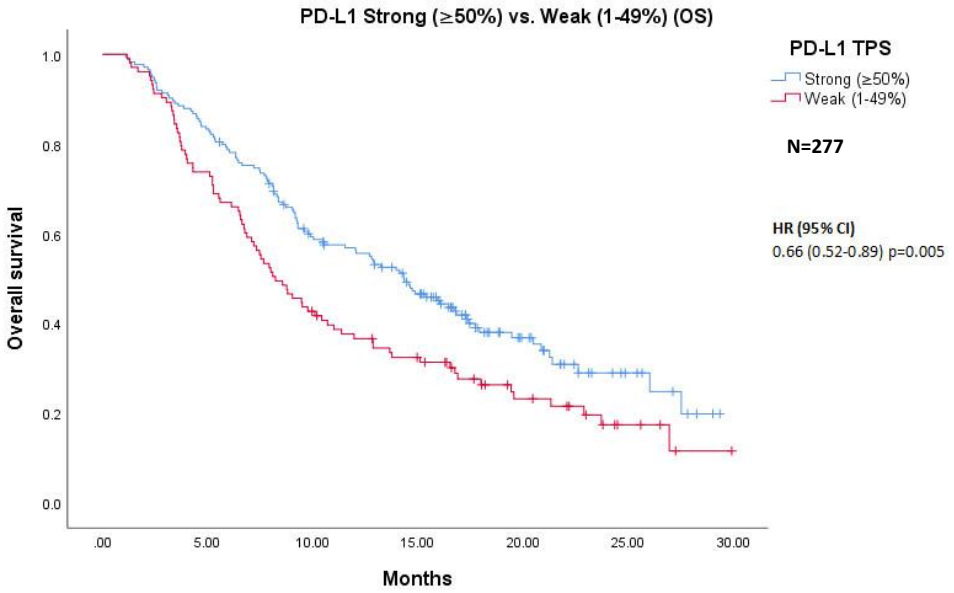


Fig 8.2.0 OS in PD-L1 strong vs PD-L1 weak NSCLC patients treated with pembrolizumab. There was a significant improvement of OS in the strong expressors compared to the weak expressors (15.7 vs 12.3 months p=0.005).

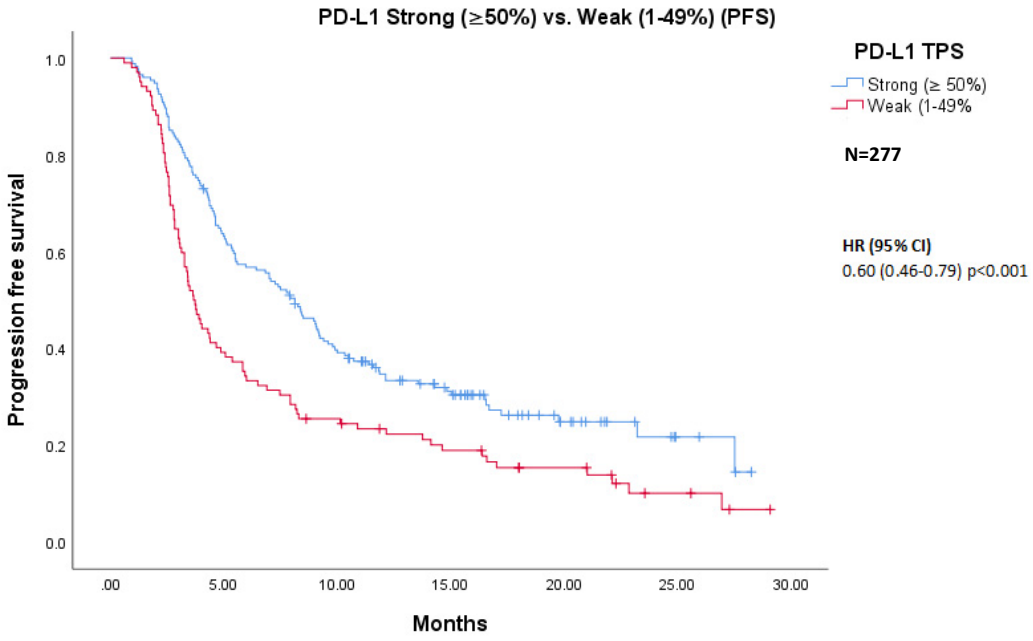


Fig 8.2.1 PFS in PD-L1 strong vs PD-L1 weak NSCLC patients treated with pembrolizumab. There was a significant improvement of PFS in the strong expressors compared to the weak expressors (12.9 vs 8.1 months p<0.001).

Utilising a ≥25% TPS to categorise patient tumours as PD-L1 positive showed a similar result: PD-L1 'positive' patients had a superior OS (15.8 vs 14.4 months p<0.001) and PFS (11.7 vs 10.4 months p<0.001) than PD-L1 'negative' patients, as shown in Fig 8.2.2 and 8.2.3.

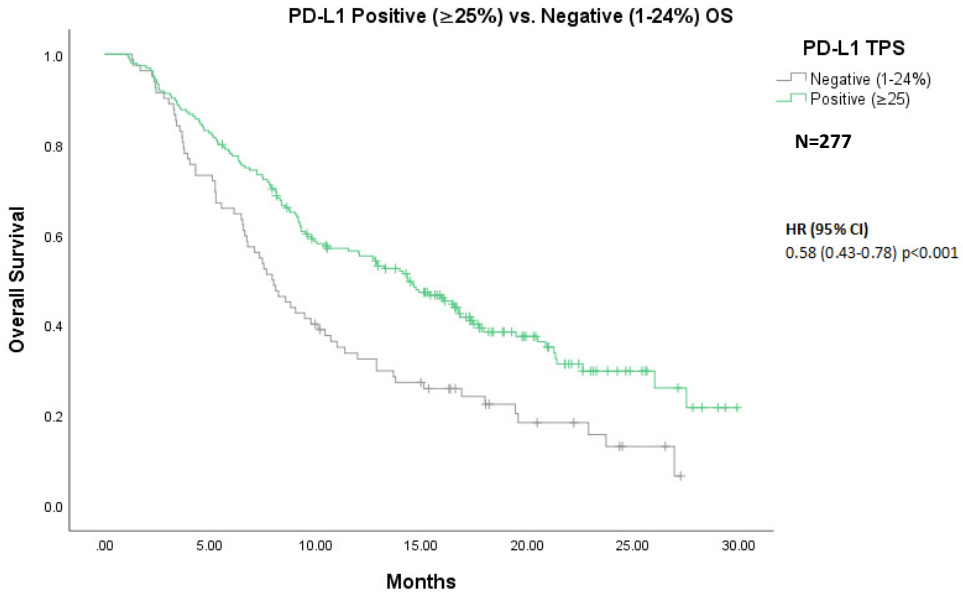


Fig 8.2.2 OS in PD-L1 'positive' (≥25% TPS) vs PD-L1 'negative' (1-24% TPS) NSCLC patients treated with pembrolizumab. There was a significant improvement of OS in the 'positive' expressors compared to the weak expressors (15.8 vs 14.4 months p<0.001).

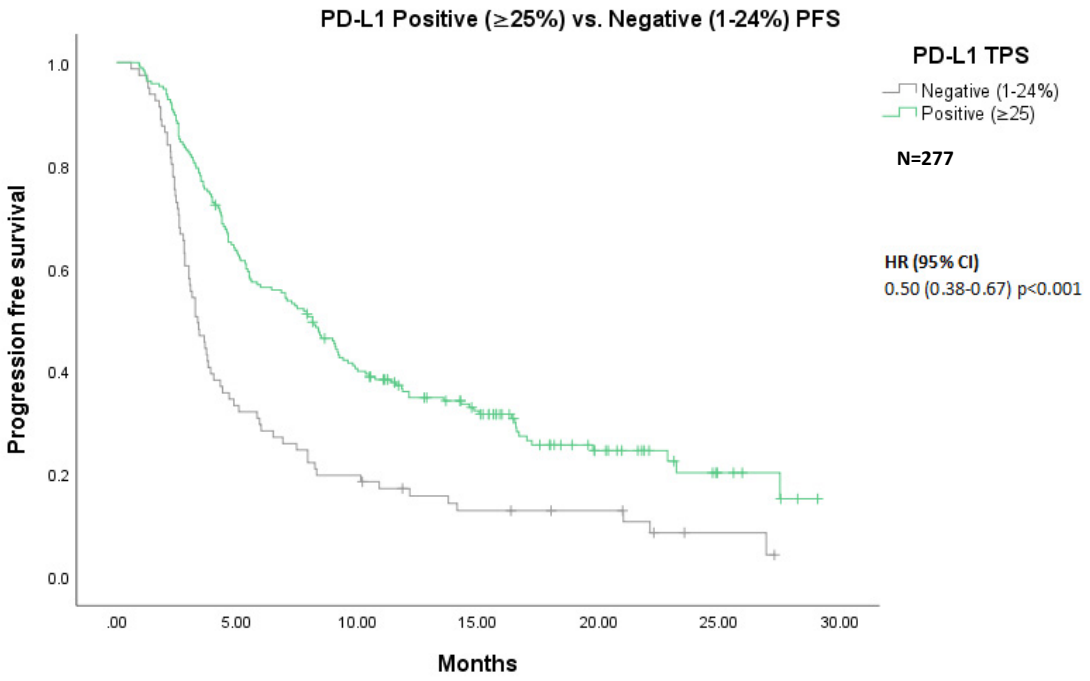


Fig 8.2.3 PFS in PD-L1 'positive' (≥25% TPS) vs PD-L1 'negative' (1-24% TPS) NSCLC patients treated with pembrolizumab. There was a significant improvement of PFS in the 'positive' expressors compared to the weak expressors (11.7 vs 10.4 months p<0.001).

These results confirmed the expected outcome that the clinically validated cut-offs of PD-L1 expression predict response to IM therapy in a NSCLC cohort.

Only patients with a TPS of ≥50% received pembrolizumab first line, so no comparison could be made using the same cut-offs (1-49% and ≥50% TPS, or ≥25% TPS) for the 1st-line IM treated patients. However, this comparison was made in patients receiving IM therapy in the second or third line setting (2nd-line plus). The analysis of 2nd-line plus IM therapy group found there was no significant difference in OS (13.5 vs 12.3 months p=0.350) or PFS (9.7 vs 8.1 months p=0.127) in patients with tumours expressing 'strong' or 'weak' amounts of PD-L1. However, when utilising the ≥25% TPS cut-off to define 'positive' in patients treated with pembrolizumab in the 2nd-line plus setting, the difference in OS approached but did not quite reach a significant improval in PD-L1 'positive' patients (14.3 vs 11.1 months p=0.051), but did demonstrate a significant improval in PFS (10.4 vs 6.9 months p<0.004) (Fig 8.2.4)

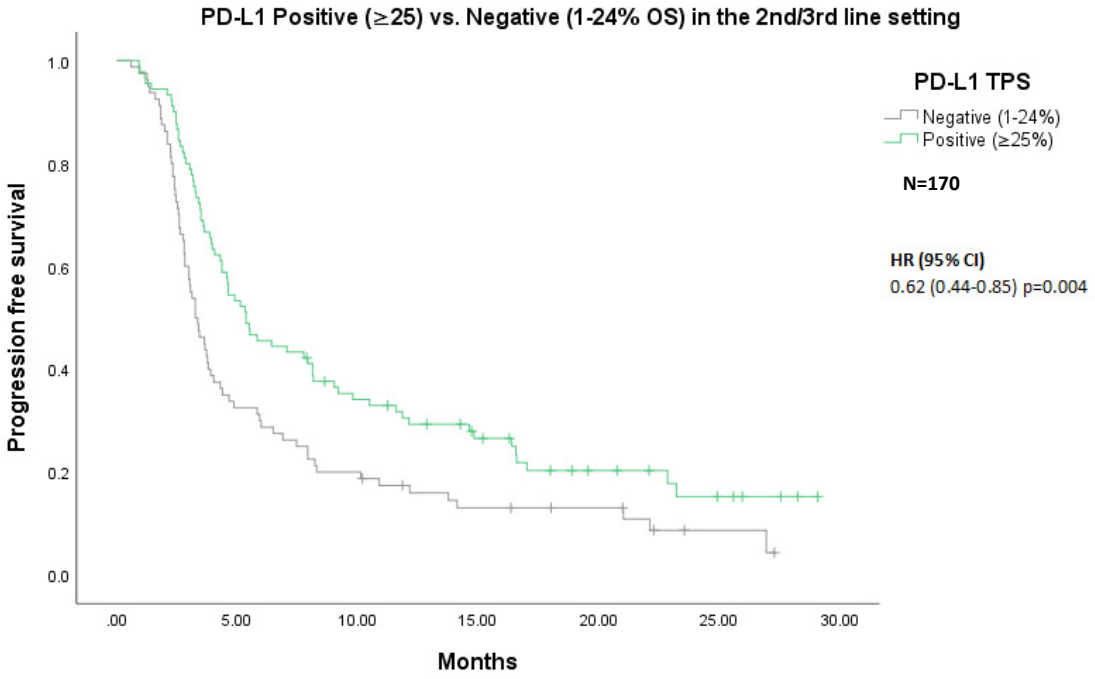


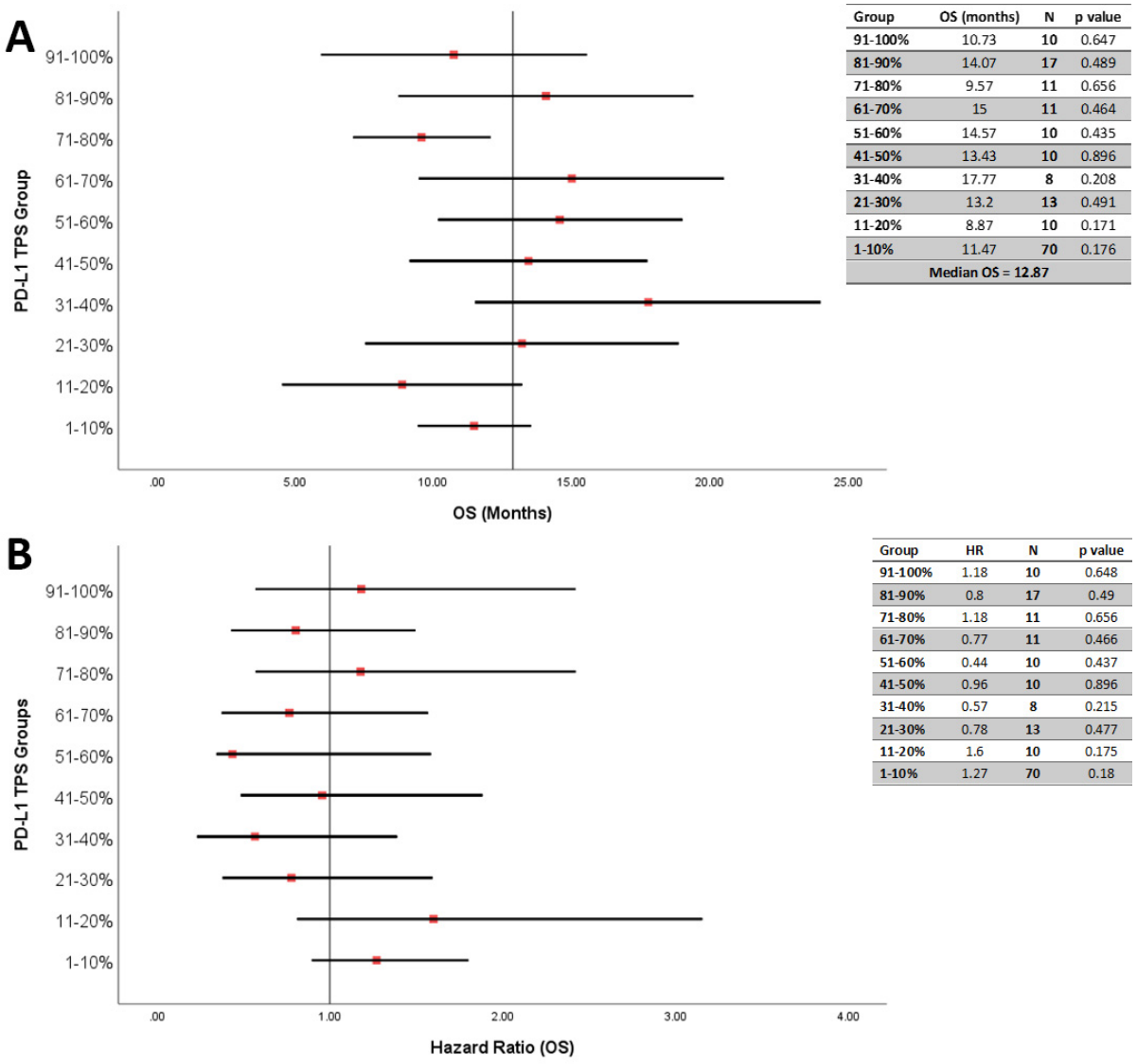
Fig 8.2.4 PFS in PD-L1 'positive' (≥25% TPS) vs PD-L1 'negative' (1-24% TPS) NSCLC patients treated with pembrolizumab in the 2nd or 3rd line setting. There was a significant improvement of PFS in the 'positive' expressors compared to the weak expressors (10.4 vs 6.9 months p<0.004).

To correctly interpret the following section of results, a brief explanation follows: as the patients of the CCC cohort were treated with pembrolizumab utilising the 22C3 clone to determine PD-L1 status of the tumours, they most closely resembled data from the Keynote-024 and Keynote-042 trials.[46, 47] However, it was only when utilising the ≥25% TPS cut-off, which was largely pioneered in the context of durvalumab and the SP263 clone,[50] that a significant benefit in PFS (with a trend that neared significance for OS) was seen in the 2nd-line plus setting.

This suggested that predictive power of PD-L1 IHC may not be limited to specific clones for specific PD-1/PD-L1 IM therapies, and that there may be advantages to utilising alternative PD-L1 'cut-offs' in specific clinical settings, and therefore an exploratory analysis to further investigate this was performed.

8.2.2 – Exploratory analysis of PD-L1 expression for OS and PFS

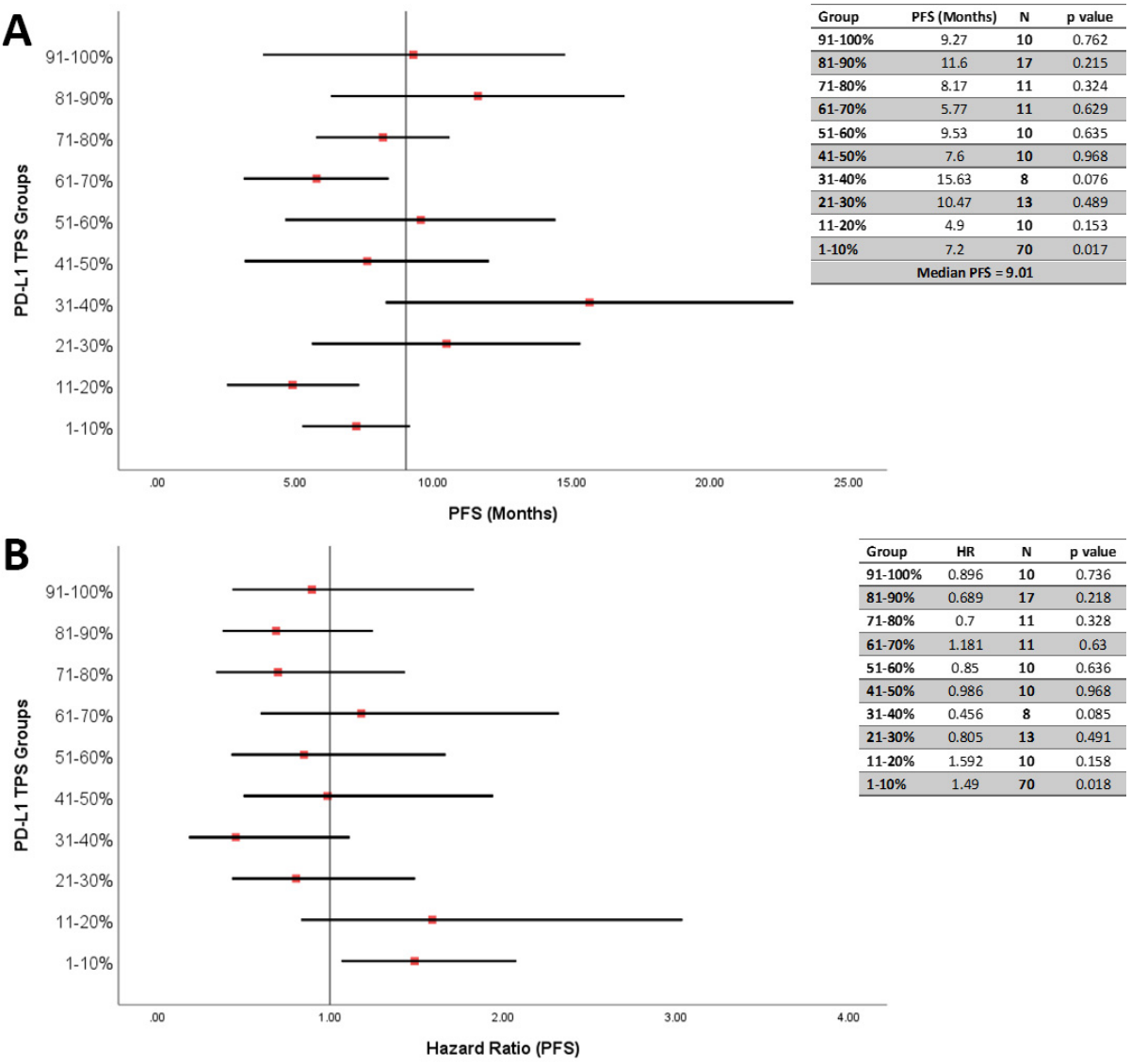
The 170 patients from the CCC cohort whom received pembrolizumab in the 2nd or 3rd line setting were divided by their tumour PD-L1 TPS into 10% incremental groups with the OS and corresponding HR for each group calculated and plotted (Fig 8.2.5), followed by the PFS and corresponding HR (Fig 8.2.6).



Group	OS (months)	N	p value
91-100%	10.73	10	0.647
81-90%	14.07	17	0.489
71-80%	9.57	11	0.656
61-70%	15	11	0.464
51-60%	14.57	10	0.435
41-50%	13.43	10	0.896
31-40%	17.77	8	0.208
21-30%	13.2	13	0.491
11-20%	8.87	10	0.171
1-10%	11.47	70	0.176
Median OS = 12.87			

Group	HR	N	p value
91-100%	1.18	10	0.648
81-90%	0.8	17	0.49
71-80%	1.18	11	0.656
61-70%	0.77	11	0.466
51-60%	0.44	10	0.437
41-50%	0.96	10	0.896
31-40%	0.57	8	0.215
21-30%	0.78	13	0.477
11-20%	1.6	10	0.175
1-10%	1.27	70	0.18

Fig 8.2.5. The OS (**A**) and HR (**B**) for NSCLC patients divided by PD-L1 expression of their tumours into 10% incremental groups. The midline represents the median OS (**A**) or a HR of 1 (**B**). Although no statistical significance is seen, the trends show that above 20% there is a general improvement in OS with a slight reduction in the very highest PD-L1 expressors (71-80%, >90%).

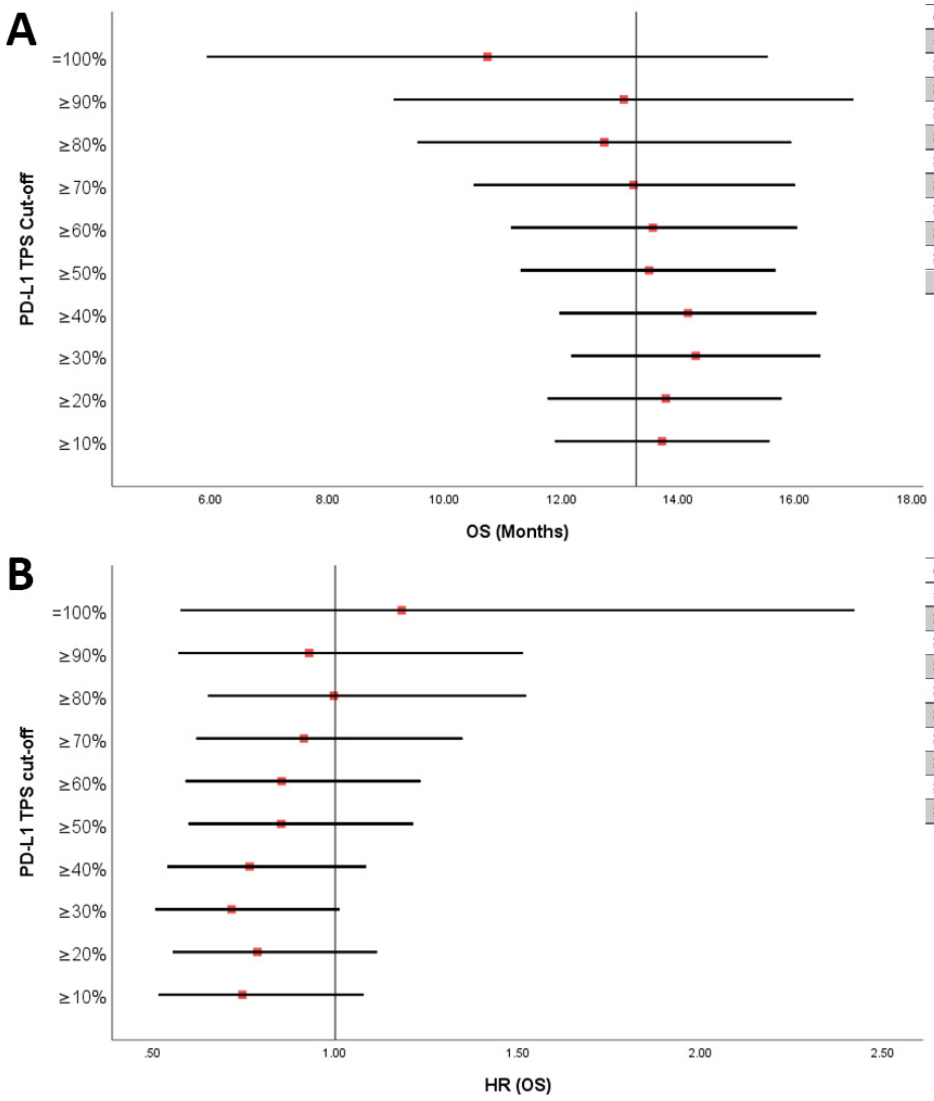


Group	PFS (Months)	N	p value
91-100%	9.27	10	0.762
81-90%	11.6	17	0.215
71-80%	8.17	11	0.324
61-70%	5.77	11	0.629
51-60%	9.53	10	0.635
41-50%	7.6	10	0.968
31-40%	15.63	8	0.076
21-30%	10.47	13	0.489
11-20%	4.9	10	0.153
1-10%	7.2	70	0.017
Median PFS = 9.01			

Group	HR	N	p value
91-100%	0.896	10	0.736
81-90%	0.689	17	0.218
71-80%	0.7	11	0.328
61-70%	1.181	11	0.63
51-60%	0.85	10	0.636
41-50%	0.986	10	0.968
31-40%	0.456	8	0.085
21-30%	0.805	13	0.491
11-20%	1.592	10	0.158
1-10%	1.49	70	0.018

Fig 8.2.6. The PFS (**A**) and HR (**B**) for NSCLC patients divided by PD-L1 expression of their tumours into 10% incremental groups. The midline represents the median PFS (**A**) or a HR of 1 (**B**). The 1-10% group has a statistically significant worse PFS. Higher categories of PD-L1 show only trends, with a general increase in PFS seen in the >20% groups. Unlike OS for the same categories, there is no trend towards a decreasing PFS seen in the highest expressors.

The same cohort of 2nd-line plus treated patients were then divided into 'positive' for PD-L1 by dichotomous division of the cohort based on their tumour TPS expression using 10% increments, with the OS and corresponding HR for each group calculated and plotted (Fig 8.2.7), followed by the PFS and corresponding HR (Fig 8.2.8).

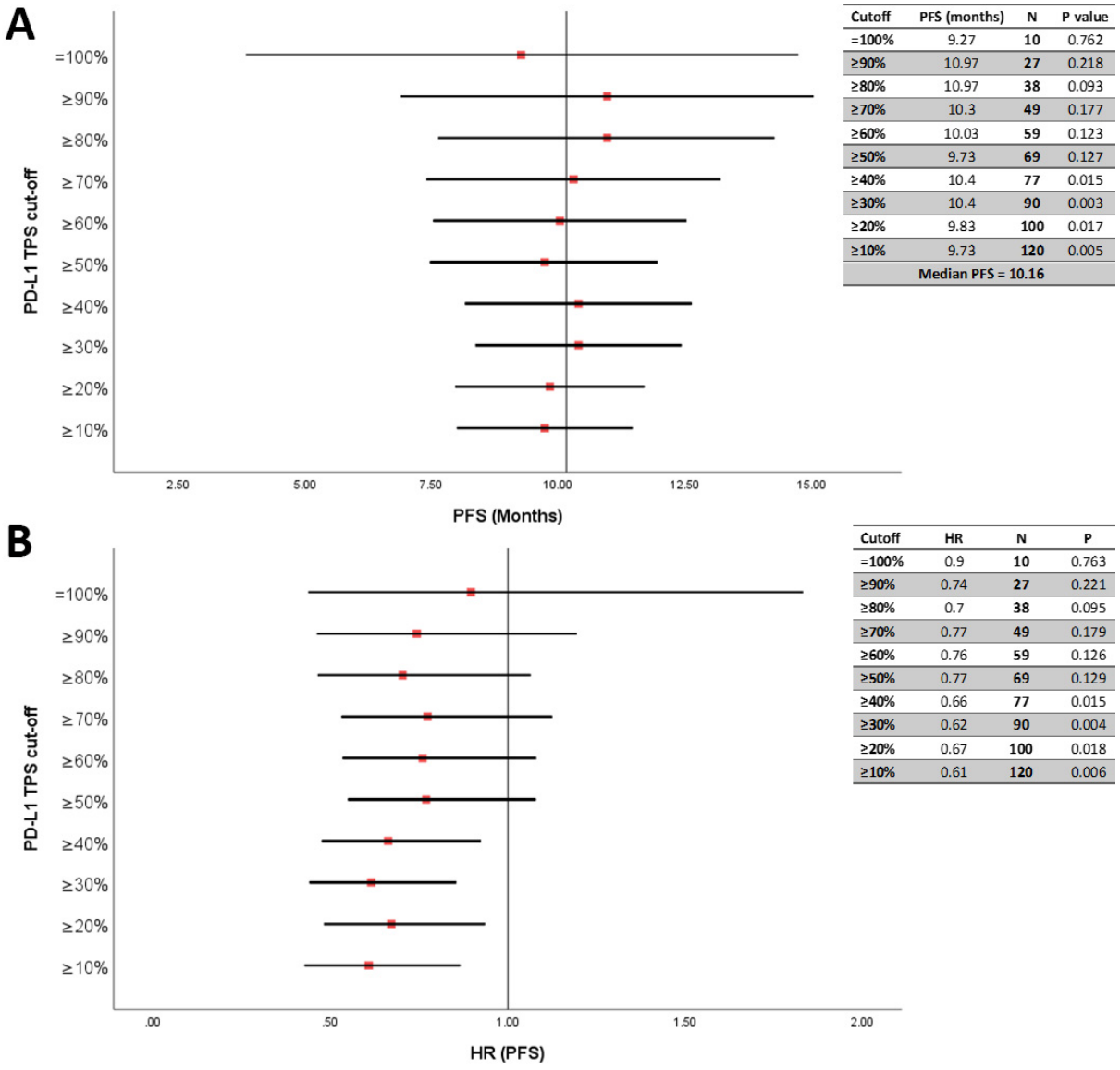


Cut-off	OS (months)	N	p value
=100%	10.73	10	0.647
≥90%	13.07	27	0.767
≥80%	12.73	38	0.984
≥70%	13.23	49	0.65
≥60%	13.57	59	0.397
≥50%	13.5	69	0.373
≥40%	14.17	77	0.131
≥30%	14.3	90	0.057
≥20%	13.79	100	0.176
≥10%	13.72	120	0.117
Median OS = 13.28			

Cutoff	HR	N	p value
=100%	1.18	10	0.656
≥90%	0.93	27	0.765
≥80%	1	38	0.984
≥70%	0.91	49	0.648
≥60%	0.85	59	0.393
≥50%	0.85	69	0.372
≥40%	0.77	77	0.131
≥30%	0.72	90	0.059
≥20%	0.79	100	0.18
≥10%	0.75	120	0.126

Fig 8.2.7. The OS (**A**) and HR (**B**) for NSCLC patients divided into 'positive' by increasing thresholds at 10% intervals. The midline represents the median OS (**A**) or a HR of 1 (**B**). No statistical significance is seen, but when raising the threshold of 'positive' the very high expressors have a slight trend towards a lower OS.

The data generated from the exploratory analysis fits with the earlier observation that a ≥50% TPS threshold may not be the optimal cut-off in predicting response to IM therapy specifically within the 2nd line setting, and that other thresholds may yield finer predictive power. Interestingly, when splitting the cohort into 10% TPS groups, there is a general improvement seen above low levels of PD-L1 expression, but there is no further improvement in OS or PFS with increasing levels of PD-L1; the relationship between PD-L1 expression and survival is not linear. Indeed the very highest expressors showed perhaps a slight reduction in OS.



Cutoff	PFS (months)	N	P value
=100%	9.27	10	0.762
≥90%	10.97	27	0.218
≥80%	10.97	38	0.093
≥70%	10.3	49	0.177
≥60%	10.03	59	0.123
≥50%	9.73	69	0.127
≥40%	10.4	77	0.015
≥30%	10.4	90	0.003
≥20%	9.83	100	0.017
≥10%	9.73	120	0.005
Median PFS = 10.16			

Cutoff	HR	N	P
=100%	0.9	10	0.763
≥90%	0.74	27	0.221
≥80%	0.7	38	0.095
≥70%	0.77	49	0.179
≥60%	0.76	59	0.126
≥50%	0.77	69	0.129
≥40%	0.66	77	0.015
≥30%	0.62	90	0.004
≥20%	0.67	100	0.018
≥10%	0.61	120	0.006

Fig 8.2.8. The PFS (**A**) and HR (**B**) for NSCLC patients divided into 'positive' by increasing thresholds at 10% intervals. The midline represents the median PFS (**A**) or a HR of 1 (**B**). A statistically significant improval in PFS and HR is seen when using lower thresholds to determine PD-L1 'positive' relative to PD-L1 'negative' in the same cohort (≥10%, ≥20% ≥30% and ≥40%) with no change between 'positive' and 'negative' patients seen when using higher thresholds.

To further investigate this, a novel categorisation of 'Very Weak' (1-20% TPS), 'Moderate' (21-79% TPS) and 'Very Strong' (80-100% TPS) PD-L1 expressors based on patients' tumour PD-L1 expression levels was generated to explore the possibility that further subdivisions, particularly at the highest levels of PD-L1 expression, may yield additional predictive value.

The OS for the moderate PD-L1 expressors was significantly longer than the very weak PD-L1 expressors (15.2 vs 11.2 months p=0.029) and trended non-significantly towards longer OS than the very strong expressors (15.2 vs 12.8 months p=0.301)There was however no significant difference in

OS between the very strong and very weak PD-L1 expressors (12.8 vs 11.2 months p=0.408). Results shown in Fig 8.2.9.

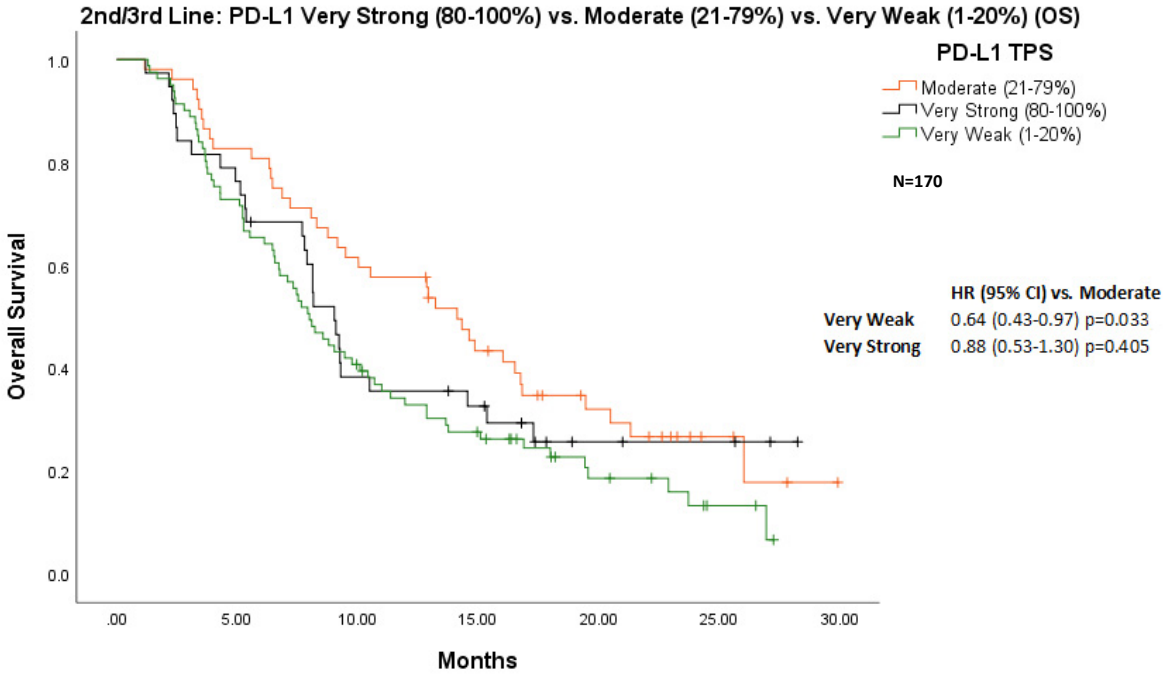


Fig 8.2.9 OS in PD-L1 'Very Strong' (80-100% TPS) vs PD-L1 'moderate' (21-79% TPS) vs PD-L1 'Very Weak' (1-20% TPS) NSCLC patients treated with pembrolizumab in the 2nd or 3rd line setting. There was an overall trend towards different OS (15.2 vs 12.8 vs 11.2 months p=0.098) for all groups with significantly longer OS between the moderate and weak expressors (p=0.029) and no significant difference in OS between the very weak and very strong (p=0.408)

The PFS for the moderate and very high expressors were both significantly higher than the very weak expressors (10.0 vs 11.0 vs 6.9 months p=0.013), but there was no significant difference between very strong and moderate expressors for PFS (11.0 vs 10.0 months p=0.652). Results shown in Fig 8.2.10.

The 1st-line IM treated cohort was then assessed using similar categories based on their tumours demonstrating moderate (50-79%) or very strong (80-100%) PD-L1 expression . The very highest PD-L1 expressors in the 1st-line setting had a significantly improved OS (19.2 vs 13.3 months p=0.009) and PFS (15.3 vs 8.0 p=0.001) compared to the moderate expressors. Results shown in Fig 8.2.11 and Fig 8.1.12.

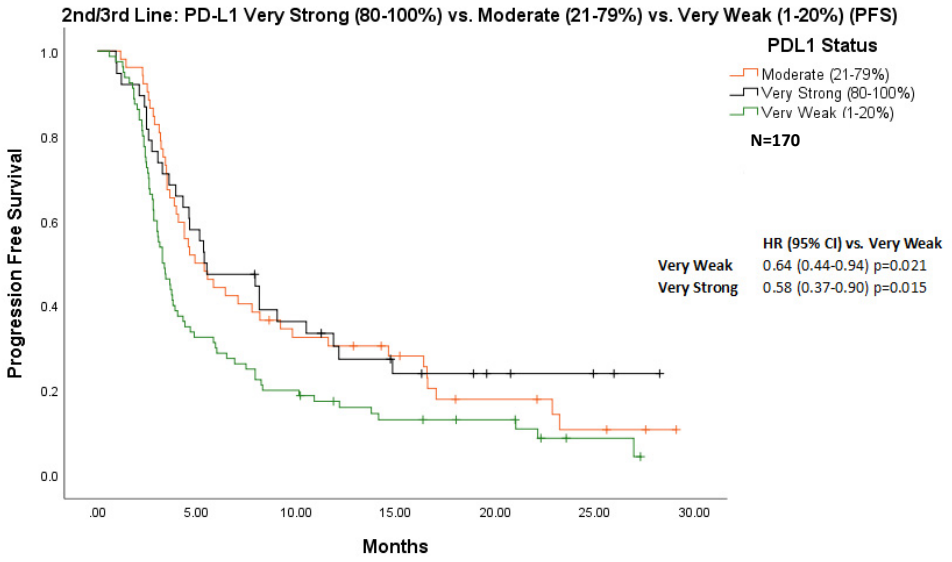


Fig 8.2.10 PFS in PD-L1 'Very Strong' (80-100% TPS) vs PD-L1 'moderate' (21-79% TPS) vs PD-L1 'Very Weak' (1-20% TPS) NSCLC patients treated with pembrolizumab in the 2nd or 3rd line setting. There was an overall significant difference in PFS (11.0 vs 10.0 vs 6.9 p=0.013) for all groups with significantly longer PFS between the moderate and weak expressors (p=0.019) and the very strong and very weak expressors (p=0.019) but no significant difference in OS between the moderate and very strong (p=0.652)

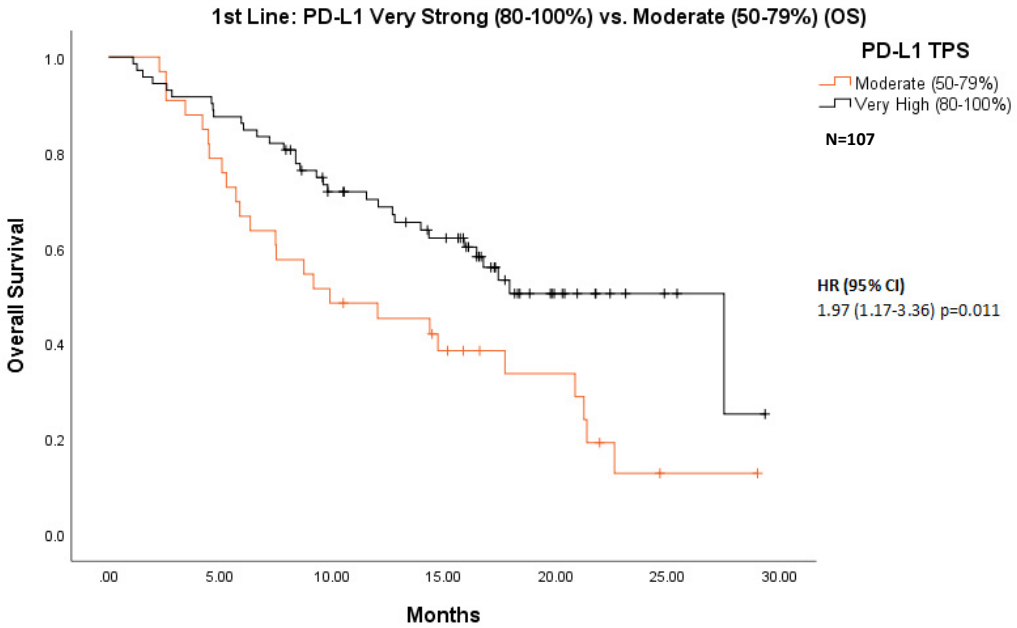


Fig 8.2.11 OS in PD-L1 'Very Strong' (80-100% TPS) vs PD-L1 'moderate' (50-79% TPS) NSCLC patients treated with pembrolizumab in the 1st line setting. There is a significant improval in OS for the very high expressors (19.2 vs 13.3 months p=0.009).

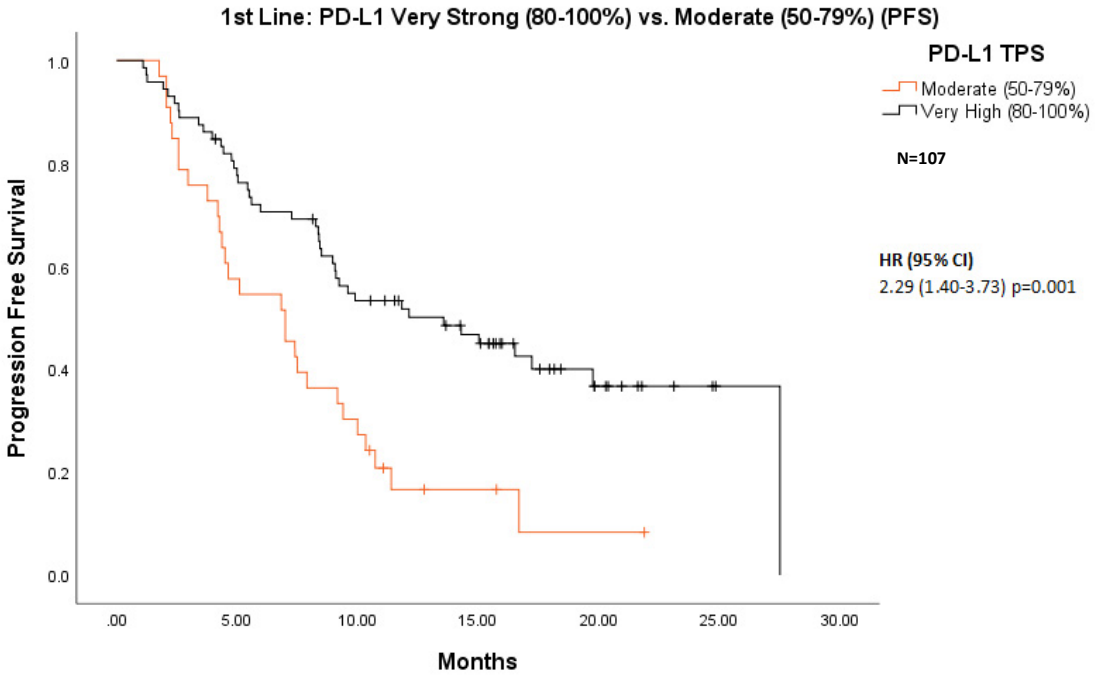


Fig 8.2.12 PFS in PD-L1 'Very Strong' (80-100% TPS) vs PD-L1 'moderate' (50-79% TPS) NSCLC patients treated with pembrolizumab in the 1st line setting. There is a significant improval in PFS for the very high expressors (19.2 vs 13.3 months p=0.009).

8.2.3 Relationship of irAEs, PD-L1 expression and treatment response

The entire CCC cohort was split into patients who had an irAE and those that did not. There was a significant increase in average OS (14.5 vs 10.1 months p<0.0001) and average PFS (10.9 vs 7.0 months p<0.001) for the group that suffered irAEs. There was no difference between low grade (Grade 1/2 toxicity) and high grade (Grade 3/4 toxicity) in predicting OS (14.6 vs 14.7 months p=0.997) or PFS (10.9 vs 11.1 months p=0.985).

The 2nd-line plus treated cohort also demonstrated a significant improval in average OS (13.8 vs 9.3 months p<0.0001) and PFS (9.5 vs 6.1 months p<0.001) for the patients that suffered irAEs compared to those that did not.

The entire CCC cohort also demonstrated a significant difference in irAEs when split by RECIST criteria, with progressive disease associated with fewer irAEs relative to the other groups (PD 23.3% vs CR 40%, PR 61%, SD 49% p<0.0001). Partial responders also had statistically significant higher rates of irAEs (61% p=0.001).The 2nd-line plus IM treated cohort demonstrated a significant difference in irAEs when split by RECIST criteria, with progressive disease associated with fewer irAEs relative to the other groups (PD 25.7% vs CR 62.5%, PR 58.3%, SD 56.4%, p=0.002) and a non-

statistically significant trend towards higher rates of irAEs in the partial responders (58.3% p=0.231) and complete responders (62.5% p=0.368).

When considering the relationship between PD-L1 expression and irAEs in the 2nd-line plus treated cohort, by dividing the cohort into 10% TPS groups, there was no significant relationship noted (Chi-Squared p=0.248), with around 40-50% of all patients suffering some form of irAE regardless of PD-L1 expression (Table 8.2.1)

			irAE		Total
			N	Y	
PDL1 Status	1-10%	Count	36a	35a	71
		Adjusted Residual	-1.6	1.6	
	11-20%	Count	6a	4a	10
		Adjusted Residual	0.1	-0.1	
	21-30%	Count	11a	2b	13
		Adjusted Residual	2.0	-2.0	
	31-40%	Count	5a	3a	8
		Adjusted Residual	0.3	-0.3	
	41-50%	Count	5a	5a	10
		Adjusted Residual	-0.5	0.5	
	51-60%	Count	5a	5a	10
		Adjusted Residual	-0.5	0.5	
	61-70%	Count	8a	3a	11
		Adjusted Residual	1.0	-1.0	
	71-80%	Count	4a	7a	11
		Adjusted Residual	-1.5	1.5	
	81-90%	Count	13a	4a	17
		Adjusted Residual	1.6	-1.6	
	91-100%	Count	6a	4a	10
		Adjusted Residual	0.1	-0.1	
Total		Count	99	72	171
Pearson Chi-Square p=0.248					

Table 8.2.1 NSCLC patients treated 2nd-line plus with pembrolizumab divided into groups determined by their tumours expressing PD-L1 in 10% TPS increments compared for the rate of irAEs. No statistically significant change in the rate of irAEs is seen across the cohort when split by PD-L1 TPS.

8.2.4 Relationship of ECOG-PS and IM cycles, PD-L1 expression and treatment response

The entire CCC cohort was divided by ECOG status; there was no significant difference in average OS (11.6 vs 12.0 vs 11.8 vs 14.1 months p=0.885) or PFS (8.6 vs 8.7 vs 8.3 vs 11.9 months p=0.723). It also worth noting, however, that only 5 patients in the entire cohort were ECOG status of three. The 2nd-line IM treated cohort was also divided by ECOG status (excluding the single patient with ECOG

status of 3), and saw a non-significant trend in decreasing OS (11.9 vs 10.9 vs 10.1 months p=0.399) and a statistically significant reduction in PFS (8.1 vs 7.3 vs 6.7 p=0.04) with poorer performance status.

When considering the relationship between PD-L1 expression and ECOG status in the 2nd-line plus IM treated cohort, by dividing the cohort into 10% TPS groups, there was no significant relationship noted (Chi Squared p=0.464) with a generally equal split of ECOG status in each group (Table 8.2.2)

			ECOG			Total
			.00	1.00	2.00	
PDL1	1-10%	Count	24a	34a	12a	70
		Adjusted Residual	0.3	0.3	-0.8	
	11-20	Count	3a	7a	0a	10
		Adjusted Residual	-0.2	1.5	-1.6	
	21-30	Count	4a	4a	5a	13
		Adjusted Residual	-0.2	-1.2	1.7	
	31-40	Count	3a	4a	1a	8
		Adjusted Residual	0.3	0.2	-0.5	
	41-50	Count	2a	4a	4a	10
		Adjusted Residual	-0.9	-0.5	1.6	
	51-60	Count	6a	2a	2a	10
		Adjusted Residual	1.9	-1.8	0.0	
	61-70	Count	2a	7a	2a	11
		Adjusted Residual	-1.1	1.1	-0.2	
	71-80	Count	2a	7a	2a	11
		Adjusted Residual	-1.1	1.1	-0.2	
	81-90	Count	6a	6a	5a	17
		Adjusted Residual	0.2	-1.0	1.0	
	91-10	Count	4a	5a	1a	10
		Adjusted Residual	0.5	0.2	-0.8	
Total		Count	56	80	34	170
Pearson Chi-Square p=0.464						

Table 8.2.2 NSCLC patients treated 2nd-line plus with pembrolizumab divided into groups determined by their tumours expressing PD-L1 in 10% TPS increments and compared by ECOG status. No statistically significant change in performance status as PD-L1 TPS changes was noted.

The association with the number of IM cycles received and its association with survival data was looked for in the entire CCC cohort and showed that higher numbers of IM cycles received was associated with improved OS (adjusted R^2 0.522, β 6.316 p<0.0001) and PFS (adjusted R^2 0.661, β 2.422 p<0.0001). There was also a statistically significant but very weak positive correlation between the number of cycles received and PD-L1 TPS (Spearman's correlation 0.178 p=0.003).

The number of IM cycles received and the association with survival data was looked for in the same way in the 2nd-line plus IM treated cohort and showed that higher numbers of IM cycles received was associated with improved OS (adjusted R^2 0.619, β 5.582 p<0.0001) and PFS (adjusted $R^2$0.697, β1.715 p<0.0001). There was no statistically significant relationship between the number of cycles received and PD-L1 TPS in the 2nd-line IM treated cohort, however. (Spearman's correlation 0.103 p=0.181).

8.2.5 Univariate and multivariate linear regression modelling

In the entire CCC cohort, the association of PD-L1 expression with survival data was looked for with univariate linear regression modelling, and was found that increasing PD-L1 expression was a significant predictor of improved OS (adjusted R^2=0.018, p=0.014) and increased PFS (adjusted R^2=0.05 p<0.001).

Multivariate linear analysis included PD-L1 expression, number of IM cycles, ECOG status, gender, patient age, tumour morphology, irAEs and whether IM therapy was received 1st line or 2nd-line plus was then looked for in regards to predicting survival outcome.

The occurrence of an irAE, PD-L1 status and the number of IM cycles received were all statistically significant predictors of PFS (irAE β 2.16 p<0.0001; PD-L1 β 1.58 p=0.05; IM cycles β 0.57 p<0.0001). Interestingly, although receiving IM therapy 1st or 2nd line was not found to be a statistically significant predictor (β 0.64 p=0.284), removing it from the model overestimated the impact of PD-L1 expression (PD-L1 β 2.06 p=0.002), highlighting the value of this data point. Gender, morphology, age and ECOG status were not statistically significant predictors of PFS. The impact of combining the combined significant predictors and the non-significant predictors of PFS are shown in Fig 8.2.13.

For OS, the occurrence of irAEs and number of cycles received were both statistically significant predictors of OS (irAE β 2.96 p<0.0001; IM cycles β 0.531 p<0.0001) but PD-L1 expression was not a predictor of OS (β 0.367, p=0.714). Removing 1st/2nd line IM data again saw a trend towards overestimating the impact of PD-L1, but did not reach significance. (β 0.978 p=0.243) Gender, morphology, age and ECOG status were not statistically significant predictors of OS.

In the 2nd-line plus IM treated cohort, the association of PD-L1 expression and survival was looked for with univariate linear regression modelling, and was found to have no association with OS (adjusted R^2=0.018, p=0.674) or PFS (adjusted R^2=0.015 p=0.294).

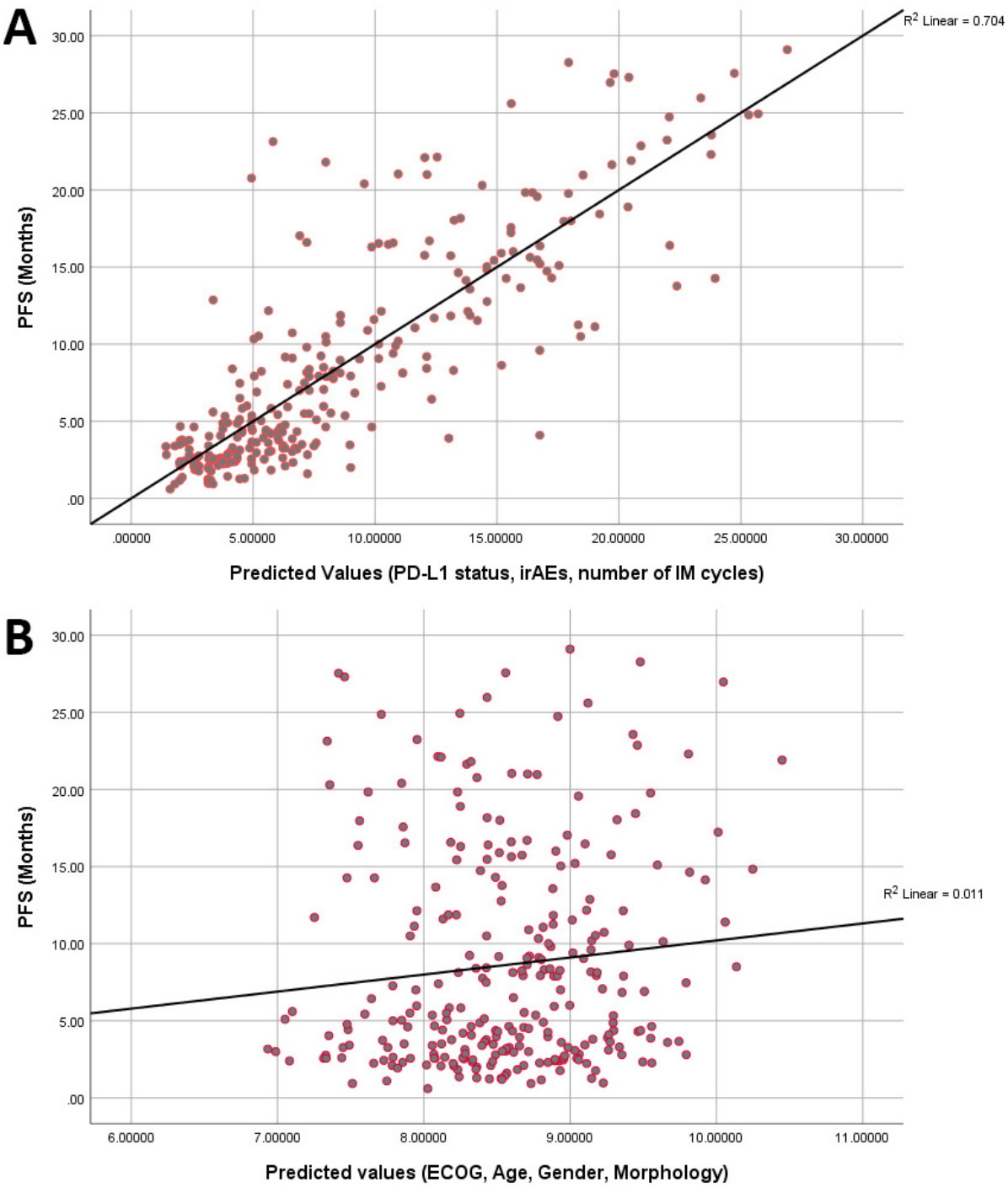


Fig 8.2.13 Multivariate analysis predicted values as a function of predicting PFS by combining the significant factors (**A**) and the non-significant factors (**B**). A model that accounts for 70.4% of variance can be achieved using the variables PD-L1 status, irAEs and number IM cycles.

Multivariate linear analysis included PD-L1 expression, number of IM cycles, ECOG status, gender, patient age, tumour morphology and irAEs was then looked for in regards to predicting survival outcome.

Only the number of cycles was a statistically significant predictor of PFS (β 0.620 p<0.0001), but both the number of cycles and the occurrence of irAEs were statistically significant predictors of OS (cycles β 0.598 p<0.0001; irAE β2.111 p=0.004). PD-L1 expression, gender, morphology, ECOG status and age were not significant predictors of survival outcome in the 2nd-line plus IM treated cohort in this multivariate analysis.

8.2.6 Pathological assessment of CCC cohort tumours

A total of 72 cases from the CCC cohort had tissue stained for 22C3 and for which the slides were available for analysis. These represented a variety of cases (Table 8.2.3). In addition to a modest number of resections, these samples were composed of biopsies of primary tissue, as well as some cytology specimens, pleural biopsies and samples from distant metastases.

The distinction of immune TMEs based on H&E was limited to primary NSCLC biopsy or resection specimens, and other specimens simply noted as having lymphocytes present or absent, if suitable. Of the 72 cases, only 60 were of primary tissue, but only 34 (47%) met the inclusion criteria for subtyping the immune TME. Dividing the cohort into immune infiltrated, excluded and desert TMEs found no significant difference in PFS (26 vs 25 vs 22 months χ^2 value = 1.05 p=0.594) or OS (34 vs 27 vs 33 months, χ^2 value = 2.42 p=0.298), and further subdivisions of the TMEs integrating PD-L1 status had no significant change on PFS or OS. Immune excluded TMEs had fewer irAEs than immune infiltrated and immune desert TMEs, (37% vs 44% vs 50% respectively, p=0.689) and had no Grade 3/4 irAEs, (0% vs 15% vs 17% respectively, p=0.182), but neither of these trends were significant.

In regards to best response, there was no statistical difference between immune TMEs for complete response, partial response, stable disease or progressive disease by RECIST criteria (p=0.394), although it is interesting to note that all cases of complete response were immune infiltrated TMEs. The hyperprogressive disease cases were all either immune infiltrated or immune excluded, although this was not statistically significant (13% vs 10% vs 0% p=0.648 infiltrated, excluded, desert TMEs respectively). There was also no statistical difference in the number of cycles of IM therapy tolerated by patients between immune TMEs (average IM cycles tolerated 11 vs 14 vs 10 p=0.552 infiltrated, excluded and desert TMEs respectively). Further subdivisions of the TMEs integrating PD-L1 status had no significant change on any of these outcomes.

A simpler classification of lymphocytes 'present' or 'absent' within the specimens was looked for across all samples, with 40 cases (56%) meeting the inclusion criteria for this. There was no significant difference in 'lymphocyte present' cases for PD-L1 high (≥50%) compared to PD-L1 low (1-49%): (68% vs 63% p=0.717). Very high PD-L1 expressers (≥61%) had fewer 'lymphocyte present' cases than moderate PD-L1 expressers (10-60%), but this was not significant, (10% vs 33% p=0.068). There was also no difference in OS (13.8 vs 13.7 months p=0.815) or PFS (11.5 vs 10.8 months p=0.821) when dividing the cohort into lymphocytes 'present' or 'absent'.

Specimen Type	N	%
Primary Biopsy	41	57
Primary Resection	16	22
EBUS FNA	3	4
Pleural Fluid	2	3
Pleural Biopsy	8	11
Distant metastases	2	3
Cellular Content		
Good	19	26
Moderate	21	29
Low	32	44
TME (1ry only)		
Infiltrated	21	36
Excluded	17	29
Desert	7	12
Unassessable	14	24
Lymphocytes (Other)		
Present	5	33
Absent	3	20
Unassessable	7	47
PD-L1 status		
<1%	4	6
1-19%	25	35
20-79	28	39
≥80%	15	21

Table 8.2.3 NSCLC cases from the CCC cohort with specimens available for PD-L1 analysis and immune TME subtyping. Specimen location, cellular content, PD-L1 expression and immune TME subtypes are shown.

Finally, differences in toxicity by TME type were looked for. There was no statistical difference between immune infiltrated, excluded and desert TMEs for irAEs (43% vs 22% vs 40% p=0.358) or grade 3/4 toxicities (10% vs 6% vs 15% p=0.625). There was no statistical difference between lymphocytes 'present' or 'absent' for irAEs (33% vs 40% p=0.613) or for grade 3/4 toxicities (8% vs 15% p=0.379).

8.3 Discussion

Predicting response to PD-1/PD-L1 IMs so as to guide management of patients before any systemic treatment is commenced is a vitally important piece of information to acquire from initial diagnostic material. Unfortunately, the nature of sampling lung masses in the routine clinical setting often produces specimens that have limited tissue and are therefore difficult to perform extensive testing on. In addition, even if relatively generous, the use of specimens that have lost morphological architecture, contain no stroma, are from lymph node metastases or other metastatic sites, are all potential limitations in the ability to interrogate the type of immune TME. Simple H&E is unlikely to provide the required detail to subtype TMEs even if of primary NSCLC. A tumour with dense lymphocytic infiltration, even in the context of high PD-L1, does not necessarily convey a high probability of response to IM therapy, as the nature and type of the T-cells present are unknown without extensive further immune cell markers. As diagnostic material may be small, the precious tissue available must be used in the most efficient manner so as to predict response to treatment. As such, anything than can improve predictive power that utilises current methods without requiring additional tests is a valuable asset.

8.3.0 PD-L1 positive – What cut-off?

PD-L1 IHC is the only clinically validated biomarker for predicting response to PD-1/PD-L1 IMs, but an important question remains unanswered; why is PD-L1 expression by IHC inconsistent as a predictive biomarker? This thesis has explored many reasons for this: pre-analytical, analytical and post-analytical variables, tumour heterogeneity and PD-L1 expression in the context of different immune TMEs, but the fact remains that the predictive power of PD-L1 expression by IHC varies between clinical trials and patient cohorts. [46, 47, 49, 663]

A potential solution may lie in a simple question: are we defining 'PD-L1 positive' correctly? In NSCLC a ≥1% TPS cut-off is typically used, with ≥50% used to define 'strong positive', and it might therefore be reasonable to assume that particularly high expressers drive the majority of the most favourable responses to IM therapy. In this study, results across the entire cohort fitted in with the previous findings of the seminal clinical trials that higher PD-L1 expression is broadly associated with better response to IM therapy. [46, 47, 49] When splitting the cohort into 1st line only patients, further evidence that the highest expressors responded most favourably to IM therapy was seen. However, the 2nd-line plus IM treated cohort was more difficult to predict. Very weak expressors (approximately less than 20-25% TPS) were indeed associated with poorer response to IM therapy, both in regards to OS and PFS. However, there was no further benefit noted above this cut-off, and indeed data from this sub-cohort suggested that perhaps the very highest PD-L1 tumours conveyed a slightly poorer OS to

patients treated with PD-1 IM, even if the PFS remained relatively unchanged. Using PD-L1 as a continuous variable also demonstrated it had no statistically significant predictive power in the 2nd-line plus IM treated cohort.

Taking these findings together, it would suggest that, in the 2nd-line setting particularly, there is not a linear relationship between PD-L1 TPS and response to IM therapy. If patients were categorised by their tumours by PD-L1 expression into a 'moderate TPS' group within the 2nd-line plus IM cohort, it was noted that they were associated with a superior OS compared to the 'very high' expressors, albeit, non-significantly. In addition, whilst the 'moderate' expressors had a superior OS compared to the 'very weak' expressors, the 'very strong' group did not. Again, a reduction in PFS was noted only for the 'very weak' group. These findings further suggest that the 'high' PD-L1 expressors should not be considered as a homogenous group in terms of response, and that treatment response is not necessarily driven by the highest of PD-L1 expressing tumours.

A change in OS but not PFS would suggest survival differences that are not entirely attributable to tumour response to therapy. Therefore the difference in OS is likely to be in part related to three areas in the context of IM therapy: irAEs, co-morbidities and the general immune status of the patient. The latter is a difficult metric to quantify, and indeed all cancer patients could arguably be considered as having impaired immune responses.[664] However, irAEs and co-morbidities/performance status were both measured in this cohort. ECOG PS was associated with a change in PFS, but not OS, in the 2nd-line plus cohort, but there was an equal distribution of ECOG status amongst differing PD-L1 expressors. A similar observation was noted for irAEs; whilst their presence was generally associated with superior survival outcomes, there was an equal occurrence of irAEs amongst the different PD-L1 expressing groups. PD-L1 thus remains a biomarker related to, but independent of, these other factors.

This provides evidence that the current clinical approach of using PD-L1 expression as a categorical variable will provide good predictive power, but that using PD-L1 as a linear continuous variable is less robust. However, whilst the weakest PD-L1 expressing tumours are generally consistent in their relatively poor response to IM therapy, these findings also suggests the need to further investigate the observation that the highest PD-L1 expressors are potentially associated with poorer survival outcomes relative to moderate PD-L1 expressors in some patient cohorts.

8.3.1 PD-L1 TPS as a marker of reactive or constitutive expression

Intuitively very high PD-L1 expressers should represent a group of tumours for which PD-L1 expression represents the main method of immune escape. However, although the PD-L1/PD-1 mediated inhibition of T-cells is the main target of PD-1/PD-L1 IMs, long-term inhibition of cytotoxic T-cells can result in extreme T-cell exhaustion and anergy, unrecoverable even with the aid of immune checkpoint inhibition,[246, 665, 666] and very high expressers of PD-L1 may therefore reflect more instances of this extreme T-cell dysfunction. In addition, PD-L1 mediated signalling has functions outside of direct T-cell inhibition and, importantly, by PD-1 independent mechanisms in some circumstances; including inducing anti-apoptotic survival,[166] acting as a direct defence against IFN-γ,[167] and inducing mTOR mediated survival in tumours cells[169, 170] and so very high PD-L1 expressing tumours may also reflect tumours resistant to IM therapy in a PD-1 independent fashion. It is also worth noting data from Chapter 7 that demonstrated in the LLP cohort that PD-1 levels had the weakest correlation with PD-L1, and there was no significant difference in PD-1 levels between high and low categories of PD-L1 expression, in contrast to other T-cell markers.

As such, very high levels of PD-L1 expression may reflect a TME that contains three factors that will minimise response to IM therapy: severely dysfunctional T-cells unrecoverable by IMs, non PD-1 mediated tumour cell survival, and relatively low levels of PD-1, all of which will limit the efficacy of treatment by PD-L1 IMs, but the latter two will particularly limit the function of PD-1 IMs, an important point as the included patients from the CCC cohort were treated by pembrolizumab, an anti-PD-1-mAb. In addition, a dominant tumour cell clone with constitutive expression of PD-L1 may result in a homogenous, very high level of PD-L1 expression, but would not be a candidate for response to treatment by IMs. As the majority of PD-L1 expression is thought to be reactive,[563, 564] however, this is only likely to play a minor role in the very high expressing group, but this may still reflect a small but important number of tumours within this category.

Conversely, moderate PD-L1 expressers might represent tumours for which PD-L1 expression is reactive and plays a major immune escape role, but in which T-cells are not so dysfunctional they cannot be reactivated by IM therapy, and in which there is less pro-tumoural activity by PD-L1 via alternative mechanisms. Very low PD-L1 expressers are on average likely to only partially respond to PD-1/PD-L1 IMs on the basis that these represent tumours for whom PD-L1 expression is part of their mechanism of immune escape, but also includes tumours for which PD-L1 expression is minimal as a result of small numbers of clones constitutively expressing it.

The weakest expressors of PD-L1 have remained as those least likely to respond to PD-1/PD-L1 therapy, most likely as a result that PD-L1 is only a minor part of their immune escape mechanism.

Indeed this observation was true across the entire cohort. However, an important observation is that, in contrast to the 2nd-line plus IM treated patients, the 1st-line IM treated patients appeared to have a consistent relationship between PD-L1 expression and treatment response. The cause for the discrepancy between 1st-line and 2nd-line plus IM treated patients is not clear, but may be related to the fact that 2nd-line treated patients have had their tumours for longer, and thus potentially allowing tumours to develop more alternative immune escape mechanisms and resistances. For example, T-cell exhaustion takes time, [612, 667] and it is therefore expected that 'older' tumours may have more robust or complex TMEs that are less susceptible to treatment with single agent IMs. In addition, the major clinical difference is that the 2nd-line plus IM treated patients have received XCT first. Whilst the impact on XCT on PD-L1 expression specifically has showed mixed results, [418-420] it is generally accepted that XCT impairs the overall immune function of patients receiving it.[668-670] Contemporary studies looking at the potential of stimulating immune response prior to IM therapy with XCT acknowledge the complexity of this process. [671] It is therefore likely that a combination of these factors at least partially contribute to the difference between 1st-line and 2nd-line plus cohorts. As 2nd-line patients generally have poorer outcomes, it is an important observation that PD-L1 expression should perhaps be treated slightly differently in these patients.

Although by necessity this cohort has excluded the PD-L1 negative tumours, (<1% TPS), it seems reasonable to state that genuinely negative tumours are likely to remain poor candidates for response to PD-1/PD-L1 IMs. It is an intriguing notion that further subdivision of the high PD-L1 expressing tumours may better yield predictive power, but in this cohort many of the trends were not significant, and in reality additional biomarkers beyond PD-L1 expression will still be required.

8.3.2 Using the TME to predict response

In Chapter 7, I proposed a method of differentiating reactive from constitutive PD-L1 expression by the presence of various TILs within the TME. A logical extension of this is to consider how to apply these findings to specimens available in routine clinical practice. Perhaps the most important observation was how few specimens had suitable tissue for subtyping the immune TME. Excluding lymph node samples and biopsies without stroma as a minimum, only around half of cases were suitable, and if excluding resection specimens (for which PD-L1 is now rarely performed in the advanced disease setting and clearly not an option for initial diagnostic tissue) less than a third of specimens were suitable. Whilst attempting to score TILs in solid tumours, including lung ADCs, has been the subject of much previous work, [672, 673] with it having the potential to be combined with digital pathology and algorithm assessment, [674] almost the entirety of these studies were performed on significant quantities of tumour, and in the lung cancer setting specifically virtually always on

resection specimens. Small samples, such as cytology specimens and small biopsies, present a robust challenge: an initial diagnosis is largely possible but there may be insufficient tissue for molecular profiling of driver mutations and PD-L1 expression, and despite careful handling, a small number of specimens remain unsuitable for these analyses.[675, 676] Furthermore, due to tumoural heterogeneity, these very small samples may not be fully informative of the tumour, particularly in regards to predicting response to treatment.[386, 677] As more immune-checkpoint inhibitors are developed, it seems reasonable to assume the number of predictive assays may increase, and the value of determining the type of immune TME may increase as a pan-IM predictor of response. Combined with the push towards NGS panel testing for molecular profiling, which requires more tissue than single-driver analysis, there is a strong argument to push for significant increases in tissue quantity when sampling NSCLC patients.[678, 679] Cost, access and patient eligibility are all things that need to be taken into consideration, but it is interesting to note the recent coronavirus (COVID-19) pandemic has seen a sharp rise of CT-guided biopsies at our centre, in lieu of the normally large number of EBUS acquired samples, with the former generally providing more generous samples of tumour. A change in sampling born of necessity could perhaps be an argument to change approaches in sampling for optimised profiling.

In the meantime, however, the subtyping of immune TMEs is likely to be accurate only in very generous primary tumour biopsies or patients undergoing surgical resection. As such, PD-L1 IHC remains a standalone test that can be easily applied to all tissue; therefore if the observation holds that moderate expressers of PD-L1 do indeed represent a group most likely to benefit from PD-1 (and perhaps PD-L1) IM therapy, in the 2nd-line plus setting, this would be an important and simple improvement in predictive profiling.

8.3.3 Predicting toxicity from PD-L1 expression

irAEs are common in patients treated with IMs. Incidence of irAEs vary considerably between studies and databases, but as many as 70-90% of patients will suffer some form of irAE, with the commonest being skin rashes, thyroiditis, pneumonitis and colitis, resulting in a broad array of complications ranging in severity, including a number of relatively uncommon fatal irAEs including pancreatitis, stroke, MI, sepsis, PE, myocarditis and long-term conditions including diabetes mellitus and pancreas dysfunction.[680-683]

Much work on IM therapy has looked at CTLA4 inhibitors specifically (as they have been in clinical use longer than PD-1/PD-L1 IMs), although CTLA4 and PD-1/PD-L1 IMs typically have distinct pattern of irAEs to each other[680, 684, 685] and whilst frequency of irAEs are associated with clinical response to PD-1/PD-L1 IMs, this relationship is much less consistent for CTLA4 inhibitors.[680, 683] It is a general

observation, however, that the combination of PD-1/PD-L1 IMs with CTLA4 IMs significantly increase the risk of irAEs over monotherapy.[686, 687] There are few head to head comparisons of PD-1 and PD-L1 IMs for irAEs, although a meta-analysis of 19 RCTs found that although treatment with anti-PD-1 mAbs resulted in superior PFS and OS than anti-PD-L1 mAbs, there was no statistical difference in safety profiles or toxicity.[688] Nonetheless, despite these distinct patterns of irAEs between IMs, and the differences in normal functioning between PD-1 and CTLA4, there are likely to be some overlaps in the mechanisms of irAEs between IM types,[680, 683] and important lessons to be learnt from all classes of IMs.

irAEs likely occur as a result of auto-immune dysfunction from the deregulation of normal immune checkpoint function subsequent to treatment with IMs. Normal immune tolerance is characterised by a lack of T-cell activity against host cells whilst maintaining the ability to mount a robust defence against foreign cells.[689] This process is primarily regulated by the thymus on naïve T-cells, but is also controlled on peripheral T-cells by a variety of mechanisms, including the expression of immune checkpoints by certain host cells. Several mechanisms have been proposed as to how treatment with IMs results in auto-immunity dysfunction. One is that certain patients have underlying, subclinical auto-immunity characteristics, and that the IMs tip the balance in favour of overt auto-immune disease, although there is little evidence that analysis of pre-existing autoantibodies and single-nucleotide polymorphisms associated with auto-immune disease can predict response to irAEs.[680] Other proposed mechanisms of toxicity are related to the function of anti-tumour immune cell behaviour evoked by IM therapy. CTLA4 IMs results in T-cells being easier to activate, and has been shown to result in increased diversification of circulating T-cells, a process which correlates with increased irAEs in patients treated with CTLA4 IMs, but is also linked to improved response to treatment.[690, 691] A particularly interesting proposed mechanism of toxicity is that of loss of self-tolerance as a result of epitope spreading.[680, 692, 693] Epitope spreading is the concept that initial immune responses targeted against a specific epitope of a protein can be diversified to include other epitope regions on the same or different proteins in order to enhance an immune response, and has been associated with deleterious effects in auto-immune disorders, and beneficial outcomes, for example following the administration of a vaccine.[692] Epitope spreading maybe beneficial to an immune response targeted against a tumour, in which multiple neoantigens can become the target of T-cells, but may also result in an increased risk of self-antigens becoming targets of T-cells, and the auto-immune disease process that this entails.[680]

In both of these latter instances, the mechanisms that contribute to anti-tumoural efficacy of IMs also contribute to the increased likelihood of auto-immune activity, and this fits the data in this

cohort in that the same patients who respond most favourably to PD-1 IMs are also the patients with the highest risk of suffering irAEs, an observation seen in several previous studies.[121, 658-662]

8.3.4 Predicting toxicity from the TME

As with predicting response to treatment, the prediction of irAEs may well be further enhanced via subtyping of immune TMEs. Previous studies have highlighted the importance of immune cells and markers for irAEs; Oh *et al*. 2017 found that an immediate diversification of T-cells after commencement of CTLA4 inhibitors may be an indicator of irAEs,[691] Schindler *et al.* 2014 and Tarhini e*t al.* 2015 found increased IL-17 and eosinophils were related to irAEs from CLTA4 inhibitors in melanoma patients,[694, 695] and Ke *et al.* 2020 has demonstrated a potential role for IL-6 for predicting irAEs in NSCLC.[696] Robust and reliable biomarkers of irAEs, particularly for PD-1/PD-L1 IMs, remain elusive, and like predicting response to treatment in this cohort, division of immune TMEs by H&E alone did not see any predictive power in anticipating irAEs. It is probable further interrogation of the TME could result in better prediction of irAEs, but small biopsies and cytology specimens are unlikely to provide the necessary quantity and quality of tissue to provide such clarification. Several major recent reviews and meta-analyses have summarised potential markers of response to irAEs, but PD-L1 expression does not appear to have been looked for as a potential candidate, or these data have not been reported.[682, 683, 697, 698] However, given the high correlation of response to PD-1/PD-L1 IMs and irAEs, an eloquent and simple solution maybe the use of moderate PD-L1 expressing categories to predict toxicity, as well as clinical response. As discussed already, relatively few samples in the CCC cohort would be suitable for such in-depth analysis, but this is an area of potential future work.

An intriguing additional possibility is the use of circulating biomarkers in conjunction with tissue to predict response and toxicity. This is an area explored in more detail in Chapter 9, but the role of circulating markers in predicting irAEs has shown promise in several studies,[699-701] and may prove to be a key metric to use alongside tissue pathology.

8.3.6 Limitations

Limitations with work in this chapter include the nature of the specimens available to me. This does however neatly reflect the challenge of using routine diagnostic specimens from NSCLC patients and emphasise the issues with the often paltry quantities of tissue we receive. The use of novel categories for defining tumour PD-L1 expression as ‘moderate’, ‘very strong’ and ‘very weak’ is based largely on trends within a relatively modest sized cohort, many of which are not significant, and the possibility these trends may not be observed in larger, independent cohorts is a possible

scenario. Nonetheless, it is well established that some patients with strongly PD-L1 positive tumours may not respond to PD-1/PD-L1 IM therapy, so the discussion as to why this may be, and how perhaps the highest expressors are the exemplars of specific resistance mechanisms, remains a valid area.

8.3.5 Conclusion

The CCC cohort is arguably the most powerful cohort I have access to, and certainly provides the greatest data on predicting response to treatment by PD-1/PD-L1 IMs. In-depth and complex analysis of the TME of tumours will hopefully reveal greater predictive power of clinical response and irAEs, but it is likely many specimens received for new diagnoses of NSCLC will be unsuitable for these analyses, even if the technologies were feasibly applicable in the routine clinical laboratory. It is my hope that more complex studies applied to the CCC cohort can provide data that may suggest a small number of easily measurable metrics that can be routinely applied to these specimens.

This study has confirmed the powerful nature of PD-L1 expression as a predictive biomarker, but has yielded the intriguing notion that yet further categorisation can be exploited to better predict response to treatment. In particular, the division of high PD-L1 expression tumours into 'moderate' and 'very strong' may help predict both response and toxicity in the 2nd-line setting, with the significant caveat this would want further investigation, ideally in a wholly distinct cohort of NSCLC patients. The predictive power of PD-L1 is a boon to the clinical decision pathway, but even taken to the theoretical limits of this, additional biomarkers will undoubtedly hold the key to the most optimal prediction of response to IM therapy.

Chapter 9 – The Next Frontier: Improving the prediction of response to PD-1/PD-L1 IMs

In the previous chapter, the benefits and limitations of using the PD-1/PD-L1 IM treated CCC cohort of NSCLC patients in order to predict response to therapy was explored. The work performed to date on this cohort has revealed important information that challenges the current paradigm of categorising patients as PD-L1 positive by conventional cut-offs, but has also demonstrated the difficulty in using routine diagnostic NSCLC specimens composed of small biopsies and cytology samples.

Therefore the following future projects are based on data generated during this thesis in order to ensure there is a good rationale for using this precious resource. Some of the hypotheses generated during this thesis, and for which I wish to confirm or refute these, will be explored in further detail in this chapter, as well as an overview of some of the potential future predictive biomarkers of response and how these may fit into routine clinical practice. It is important to note that the process of generating hypotheses to test in future work is both limited and informed by the accessibility of specific platforms and technologies, the knowledge of specific biomarkers, and the general principles and concepts which are backed up by evidence to varying degrees. This process is an ongoing and evolving one, and some of the main themes, biomarkers and hypotheses are shown in Fig 9.1.0

9.1 Novel use of established biomarkers: PD-L1 expression in the pleura

9.1.0 PD-L1 in pleura and pleural fluid as a function of immune escape

In Chapter 4 (Pre-Analytics), the distribution of PD-L1 expression by anatomical location was found to be generally consistent, with the exception that specimens from pleural fluids (and the small number of pericardial fluids) had a disproportionately and significantly increased number of PD-L1 strong expressers. This trend was also seen to a lesser degree for pleural biopsies. The observation that pleural fluid specimens have more high expressers of PD-L1 over primary NSCLC specimens has been seen in other studies,[318, 472, 702, 703] and whilst the difference in some studies is relatively modest, the concordance with primary tissue is far from perfect.[313, 702] Most studies looking at this phenomena have focused on the concordance of pleural fluid in regards to the value of using it to analyse PD-L1 expression. My hypothesis is that the difference between these specimens is not due to artefact, but is in fact a genuine immune escape related difference between tumours that metastasise to body cavities and those that do not.

Categorisation of NSCLC as 'advanced' refers to stage III/IV tumours which is based on TNM staging and can be considered advanced by primary tumour ('T' = primary tumour grows to a large enough

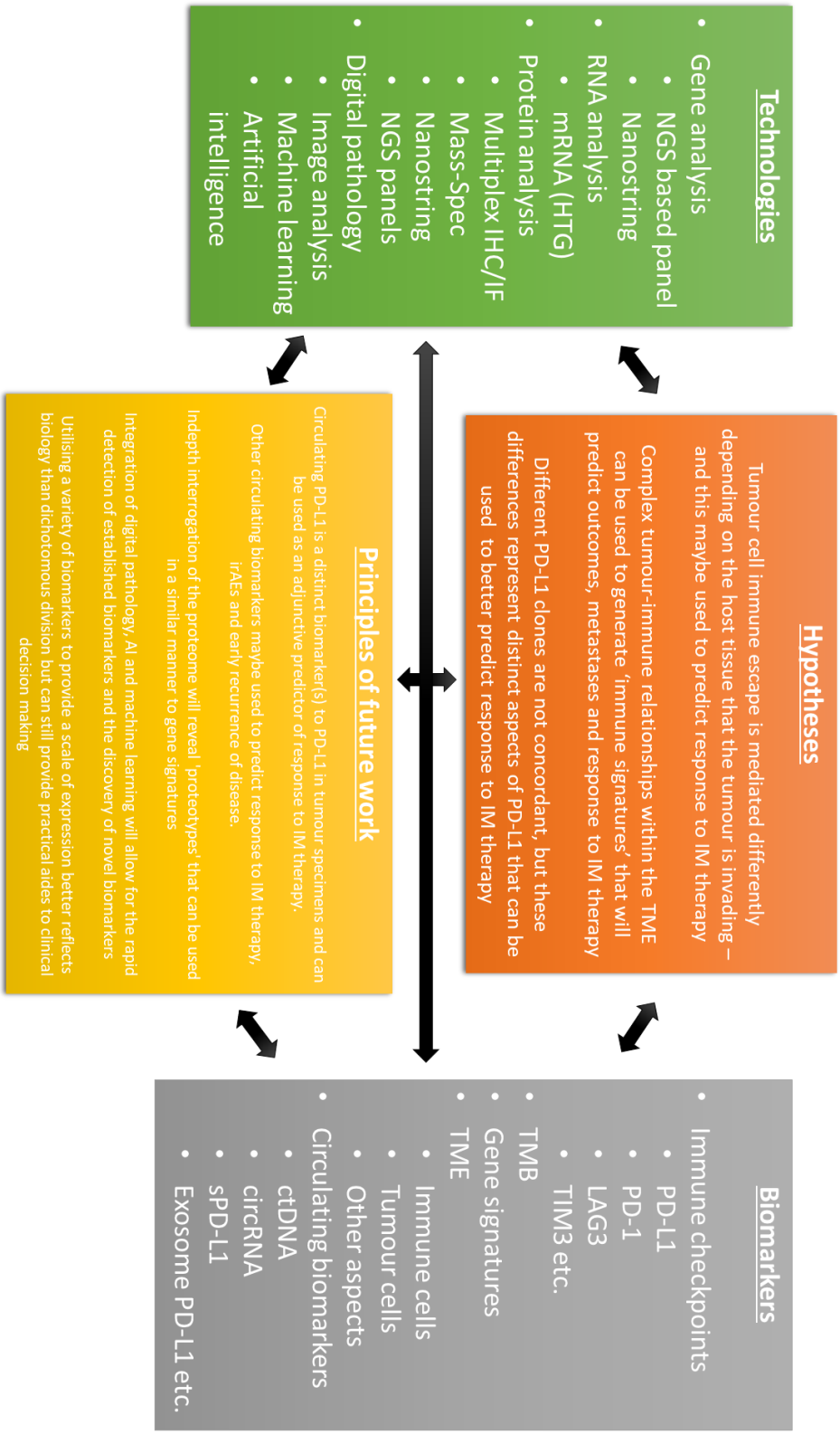


Fig 9.1.0 Flow chart of 4 areas for developing future projects. Available and accessible technologies, the specific biomarkers of interest and the basic principles of future work are used to inform each other and to generate, refine and improve hypotheses, which in turn effects the principles of what and how to look for biomarkers.

size or is near to anatomical landmarks or has direct invasion through non-parenchymal tissue), spread to lymph nodes ('N', spread to nodes that drain the thoracic cavity) or by metastases ('M', tumour in both lungs, pleura/pericardial discrete masses, malignant pleural/pericardial effusion, or spread to sites out with the thoracic cavity, including distant lymph nodes).[395] Staging is important for prognostic data and treatment decisions, but shared categories can still represent distinct subtypes of advanced NSCLC; for example, a stage IV patient may be so because of contralateral lung tumour, distant nodal disease, malignant pleural effusion, or a combination of these features. Thus, many advanced tumours do not display a preference for metastasising to thoracic cavities, and for those that do, it may be that specific immune-escape related mechanisms are required for effective spread to these cavities (or that effective spread is driven by these mechanisms) and the high rate of strong PD-L1 expressers seen in these specimens is a reflection of this.

Several studies have looked at immune markers that vary between benign and malignant pleural effusions, perhaps suggesting specific immune-escaped mediated processes that convey a survival benefit for malignant cells in the pleural cavity. In one study, CARD9 (Caspase recruitment domain-containing protein 9) was found to be significantly upregulated in NSCLC malignant pleural effusions.[704] CARD9 is a diverse adapter protein playing a regulatory role in a number of cell survival and apoptotic pathways, including the MAPK pathway and NF-kB signalling.[705] It has also been shown to have a role in the innate immune response, and, amongst other pathways, can stimulate dendritic cells via the upregulation of cytokines including TNF-A and IL-23 and IL-12.[706] The authors of the pleural fluid paper concluded one or more of these CARD9 mediated pathways was a key determinant in ADCs metastasising and surviving in the pleural fluid, and was therefore a potential biomarker of malignancy for these specimens,[704] but this finding also provides data that suggests certain immune related proteins expressed on tumour cells may be required for metastases and/or survival within the pleural space. HMGB1 (High-mobility group box 1) has also been found to be highly upregulated in malignant pleural effusions compared to benign pleural effusions, alongside IL-6 and IL-8.[707] HMGB1 has been shown to be overexpressed in NSCLC patients' tumour tissue and serum specimens[708, 709] and probably plays a key role in NSCLC development via mediation of the MAPK pathway[710] or NF-kB signalling[711] as well as playing a key role in NSCLC metastases specifically.[712] HMGB1 has also been shown to drive immune and inflammatory processes in malignant and autoimmune diseases.[713-715]

For both CARD9 and HMGB1, upregulation may play a role in tumour cell survival, but they also both play key roles in regulating immune and inflammatory responses. Antibodies against HMGB1 have been shown to provide a protective benefit against inflammatory mediated cell damage[716] and pulmonary inflammation specifically.[717] Therefore, although proteins such as these convey survival

benefit to tumour cells through mediation of cell cycle and apoptotic pathways, they are also likely to generate significant inflammation and immune responses, for which tumour cells will require an immune-escape pathway, and for which it is possible PD-L1 mediated immune-escape plays a particularly successful role.

PD-L1 is also known as B7 homolog 1 or B7-H1. Its relative in the B7 family, B7-H3, has also been shown to be upregulated in malignant pleural effusions secondary to NSCLC and associated with a poorer survival for these patients.[718] It is less well characterised than PD-L1, but it does appear to play a key role in immunosuppression in NSCLC[719-721] and has been shown to be elevated in the serum of NSCLC patients[720] and associated with increased TILs in NSCLC tumours.[722] B7-H3 appears to play a similar role to PD-L1 in some respects in that it has a co-inhibitory role in T-cells leading to immune escape for tumour cells[721] and has thus retrieved some attention as a potential target for IM therapy.[74, 723] Its full role is yet to be elucidated, but if the observation holds that B7-H3, like PD-L1, is consistently upregulated in pleural effusions, it may suggest that B7 family mediated mechanisms of immune escape maybe particularly effective, and potentially unique, for NSCLC tumour cells that metastasise to the pleural space.

9.1.1 Specific immune escape functions for specific tissue invasion

Certain tissues in the body have particular immune properties that require a specific set of immune escape mediated pathways for tumour cells to survive. This is perhaps best illustrated in the case of brain metastases. Tumour cells metastasising to the brain have to overcome several barriers including the blood-brain-barrier and the unique architecture, blood flow and oxygen availability of the brain.[724] Whilst molecular features of tumour cells that achieve brain metastases vary between tumour types (e.g. Upregulated alterations in *ALK*, *BRAF* and HER2 in NSCLC, malignant melanoma and breast cancers respectively),[725, 726] a number of shared features are required to survive once they have invaded. In particular, the unique challenges of the immune response within the brain require a shared and unusual set of properties, such as immune escape from microglial cells (a CNS specific cell with immune functions) which have been shown to destroy cancer cells.[727] If the tumour cells can successfully achieve immune escape, the microglia can actually provide a survival benefit to the tumour cells[728, 729] and so these shared immune escape properties can significantly enhance the benefit to these metastasised cells. Metastases to the brain thus sets the precedent that tumour cells require unusual characteristics to both successfully invade and to survive in particular environments, including a unique set of immune-escape mediated properties, and therefore the same maybe true of tumour cells invading the pleural space.

Mesotheliomas are malignant tumours that originate in the pleura, pericardium or peritoneum. Mesotheliomas are distinct to NSCLC in terms of cellular origin, molecular properties, aetiology and most other factors, but the fact both may survive within the pleural cavity raises the intriguing possibility that immune escape processes that are successful for mesotheliomas may also be successful for NSCLCs metastasising to the pleural space. Indeed several studies have shown PD-L1 to be expressed in a significant proportion of mesotheliomas or malignant pleural effusions secondary to mesotheliomas, varying between studies in terms of incidence, but generally around 20-50% of specimens being positive for PD-L1 regardless of the clone used.[730-734] The prognostic significance of PD-L1 positive mesotheliomas, as with NSCLC, varied between studies.[731, 735] Although response to PD-1/PD-L1 IM therapy has seen mixed response in mesotheliomas, with no IM therapy yet approved[736, 737] it does still suggest that PD-L1 expression may play a key role in achieving immune escape within the pleural cavity or pleural tissue, regardless of the malignant cells' origin. Indeed, as PD-L1 has been shown to induce immunosuppression by multiple methods, including non-PD-1 mediated pathways, the lack of response to PD-1/PD-L1 IM therapy in mesotheliomas, despite the prevalence of PD-L1 expression, may suggest that PD-L1 mediates a unique immune escape pathway in the pleural area, and therefore these differences would explain why many NSCLC patients with PD-L1 positive pleural effusions do not respond to PD-1/PD-L1 IM therapy.[738]

In order to explore the hypothesis that high PD-L1 expression in pleural specimens is a result of specific immune escape processes utilised preferentially by NSCLC cells that metastasise to the pleural space, specimens from the CCC cohort that represent distinct subtypes of 'advanced' tumour can be compared. By comparing advanced tumours that have not metastasised, spread only to lymph nodes, metastasised to the pleura or pleural space or metastasised to other sites, these NSCLC tumours can be separated on the basis these differing behaviours of advanced tumours may also reflect distinct immune escape profiles, and so provide insight into the underlying mechanisms of immune escape as well as help to better guide PD-1/PD-L1 IM therapy.

One potential pathway to do this is via an immune-oncology (IO) mRNA-based NGS panel such as that offered by HTG.[739] It is outside the scope of this thesis to provide an in-depth review of the strengths and limitations of mRNA studies for protein expression, but despite mRNA not always correlating with protein expression by other methods, including PD-L1,[335, 336, 355] it does provide an excellent opportunity to compare these potentially distinct groups of tumours by looking at gene expression profiles of proteins related to immune functions (such as B-cell function, T-cell function, TNF, NK function, ILs and various cytokines and chemokines) and other important roles relating to normal and malignant cell function (e.g. cell adhesion, apoptosis, cell cycle, cell signalling and proteasomes) including members of the CARD family and HMGB1.

mRNA as a marker of gene expression is not perfect, however, and the direct study of proteins themselves may yield more informative data, as explored in Chapter 7 (The tumour microenvironment) when using multiplex IHC to define the immune TME.

9.2 Immune TME signatures to predict response to IMs

A recurring issue in predicting response to IM therapy is the reduction of very complex biology to simple categorisation of patients into potential 'responders' or 'non-responders'. As our knowledge of immune escape, tumour heterogeneity, and the TME increases, it is increasingly apparent that even a single biomarker means different things in different contexts. PD-L1 expression in the context of a primary lung NSCLC biopsy with large quantities of CD8+ve, PD-1+ve, functional TILs infiltrating the tumour epithelium is probably distinct from PD-L1 expression in a biopsy of metastatic disease with an immune desert TME, for example. These scenarios likely reflect a variation in response to IM therapy and the distinct biological mechanisms of PD-L1 expression.

A simple classification of 'PD-L1 positive' is therefore insufficient to accurately predict treatment response, as is the relatively simple division of tumours into infiltrating, excluded or immune deserts based on H&E. Furthermore, patients do not simply 'respond' or 'not-respond' – some will respond very well, some poorly, some will progress rapidly, others more slowly and so forth.

Predictive assays that utilise multiple markers provide an opportunity to extract additional information from the specimen so as help define the significance of expression of any single biomarker, such as the presence of cytotoxic TILs in the context of PD-L1 expression given in the example above. However, these assays also run the risk of providing clinicians with reams of additional data that are not helpful in making a clinical decision. An example outside of immunotherapy is NGS panels that screen multiple genes for TKI therapy. These can return data on actionable mutations, but may also show the tumour to harbour only rarer mutations of unknown clinical significance, the knowledge of which is of dubious clinical benefit.

Even when predictive assays utilise multiple signals to define a pre-determined 'cut-off', it is far from simple in being a consistent tool. For example, the use of TMB to predict response to IM therapy is hampered by questions surrounding the best approach to measuring TMB, the variable definitions of 'high TMB' as well the mixed clinical responses seen in these patients.[267, 740-743] The biggest limitation to this approach is the categorisation of patients or tumours as 'having' or 'not having' a particular feature (TMB is 'high' or 'low', PD-L1 is 'negative' or 'positive' etc.) when these are often continuous variables with expression more accurately reflected as a spectrum. In clinical practice, it is useful to have biomarkers that provide clear guidance for treatment decisions, and the division of patients

into clear groups is the only practical way of running clinical trials by the established methodology. Nonetheless, this simple division does not reflect basic biological principals, oversimplifies an incredibly complex area and results in the inevitable exclusion of patients who would benefit, and inclusion of patients who do not, based on what is ultimately an arbitrary division of 'positive/negative' or 'high/low'.

Despite these issues, there has been success in utilising multiple markers to define useful groups for predicting response to IM therapy. Gene expression profiles have been used to create 'immune signatures' in which the presence of increased levels of mRNA for certain proteins has seen some success in predicting response to IM therapy. Ayers *et al.* 2017 devised a 18 gene signature using Nanostring technology that is related to IFN-γ mediated T-cell inflammation of tumours and was associated with superior response to PD-1 IM in HNSCCs,[744] and Hwang *et al*. 2020 found two gene signatures, one related to TILs, and one related to M1 macrophages, that were both associated with superior response to PD-1 IM therapy in NSCLC.[745]

Therefore, the creation of immune TME 'signatures' based on the presence, location and density of TIL types and other protein based immune markers within the TME may well provide a balance between oversimplification and providing too much data. Furthermore, although this thesis has focused on PD-1/PD-L1 IMs specifically, CTLA4 IMs and many other immune checkpoints being actively considered as potential targets of IM therapy may benefit from immune TME signatures. For example, a TME signature that can reliably demonstrate that the major mechanism of immune escape is immune-checkpoint mediated via direct tumour-cell/APC and T-cell interactions may only require the additional information of which checkpoint(s) are involved to identify the optimal IM therapy approach.

Current attempts to utilise multiple markers within the immune TME that produce divisions such as subtyping TMEs into 'hot' and 'cold' immune types has seen some success[746, 747] but, like TMB, the assays to define this, the specific metrics to use and the actual cut-off of 'hot' is variable. Furthermore, it is probable this dichotomous division does not sufficiently take account of the variable functionality of immune cells, the knowledge that tumours treated with IMs are characterised by a spectrum of responses and resistances, and that the TME may be regulated by multiple methods of immunosuppressive and cytotoxic activities.

As shown by the multiplex and image analysis study in Chapter 7, there are multiple ways to subtype immune TMEs depending on which markers are used and the priority they are given. The subtypes of immune TME suggested in Chapter 7 are only prototypes, but the advantage of utilising multiple immune markers, quantifying their expression and describing their spatial relationship not only more

accurately reflects our biological understanding of the processes that lead to immune escape, but allows for finer stratification of patients by response to IM therapy and the potential likelihood of irAEs. Far from overcomplicating the decision making process for routine clinical practice, this stratification of patients via an immune TME signature may provide us with more personalised patient information, as well as potentially identifying the best patients eligible for novel treatments and clinical trials as they are developed.

One potential approach to further developing the notion of an immune TME signature is applying TME immune subtypes to the PD-1/PD-L1 IM treated CCC cohort based on the data collected in Chapter 7. The largest limitations of the CCC cohort, and routine NSCLC diagnostic specimens in general, are the small biopsies and cytology specimens, which potentially lack stroma, may have no preserved architecture present and can involve considerable admixing of different cell types. As already discussed, this is an issue for defining immune TMEs. However, it is possible that immune markers present within many specimens will still be suitable for refining the subtyping of TMEs, and ultimately provide specific markers that can be used to predict response to PD-1/PD-L1 IMs. The CCC cohort would therefore act as a cohort to test the feasibility of creating an immune TME signature by this approach in routine NSCLC specimens, as well as providing patient outcome data to help identify which immune features of the TME are useful in predicting response to treatment, and if any distinct markers can better identify patients likely to suffer irAEs.

This novel approach to utilising multiple protein expression by multiple IF or IHC, rather than mRNA or gene based assays, could be a useful and powerful tool in better predicting response to IM therapy, but requires the careful and accurate application of digital pathology and image analysis tools.

9.3 Predictive power of image analysis

Image analysis combined with digital pathology and other machine learning and AI based techniques is undoubtedly the future of pathology. Current attempts to replicate the human pathologist's ability to interpret routine H&E sections are limited to relatively simple distinctions at present,[536, 748] and the subtle and nuanced distinction of much of pathology, combined with the requirement to combine the microscopic image with the macroscopic specimen findings, as well as integrating complex clinical understanding and radiological findings means that pathologists are still likely to be valuable assets for routine clinical diagnostics for many years hence. The true value of digital pathology and image analysis at present lies in its ability to augment pathologists' work. These can range from the simple but convenient advantages of digital pathology such as the visualisation of larger quantities of microscopic tissue over conventional microscopy, easy measurement tools and

rapid case sharing, to the more complex image analysis techniques required for interpreting multiplex immunohistochemical and immunofluorescent stains so as to accurately quantify markers as described in Chapter 7. In many centres digital pathology is a reality, with conventional microscopy and glass slides used only when there had been failing of the digital service or is a strict requirement for viewing a particular assay,[507, 508, 512] though it has yet to be fully integrated into many labs, even within wealthy countries such as the UK. Nonetheless, as these services are developed, it can be hoped that simple additions to digital pathology can be routinely added in a seamless fashion.

Despite these positives, many questions still surround the use of digital pathology, and image analysis and machine learning approaches in particular. One such question is the responsibility of an automated algorithm. By conventional pathology, a particular assay scored by a named individual leaves a clear route in the event of an error, such that re-education and re-training can be initiated. An unfortunate reality of modern medicine, however, is a culture of 'blame', in which mistakes are not always seen as an opportunity to learn and improve. In the theoretical situation that an automated algorithm makes a mistake which ultimately results in a patient coming to harm, the question of who or what to 'blame' has no certain answer. Furthermore, one of the core concepts of machine learning and AI is that experience informs the algorithms such that they improve overtime.[749] Given the tight nature of the regulatory process of new medicines and *in-vitro* diagnostics used in medical practice, it is difficult to see how an automated algorithm can be given regulatory approval if the specific device is a constantly changing and evolving process, with no guarantee that each iteration will necessarily be superior to the previous one.

As such, a pragmatic approach will likely lie in the use of semi-automated algorithms, in which pathologist input is directly required for their use, or the augmentation of the pathologist's workload by automated algorithms, such that the ultimate oversight of the case still remains the purview of the pathologist. Despite the fantastic advances seen in machine learning, the reality is that biology is hard to understand, extremely complicated and notoriously inconsistent, and thus far no machine learning tool can replicate the subjective and qualitative abilities of a skilled pathologist. Nonetheless, humans are generally bad at certain tasks that computers can process extremely rapidly, with the example used in Chapter 5 (Analytics) of counting cells: pathologists are masters at ascertaining what to count, but computers are the masters of counting.

The image analysis tools for PD-L1 IHC interpretation as explored in Chapter 5 provides evidence that the assistance of machine-learning based algorithms can decrease intra-pathologist and inter-pathologist discordance when scoring PD-L1, and may therefore increase overall accuracy of PD-L1

analysis. The most important metric when considering scoring PD-L1 IHC is its ability to predict response to PD-1/PD-L1 IMs. If the division of specimens into negative, weak and strong PD-L1 expressers (or indeed other categories) sees variation between original manual assessment of the specimens and image analysis algorithm assisted scores, the true measure of the value of the algorithm would be to see which sets of scores most accurately predicted response to IM therapy. The limiting factor for the use of this algorithm is that it was designed specifically to use with the SP263 clone. However, the PD-1/PD-L1 IM treated CCC cohort utilised the 22C3 clone to stain for PD-L1 (in keeping with our laboratories clinical practice at the time of these patients' diagnoses). Therefore tissue from this cohort will need to be re-stained with SP263, scored manually for a PD-L1 TPS, and then re-scored with the assistance of the Roche-PD-L1-algorithm to assess the value of using image analysis. In so doing, an important objective measurement of the value of using the image analysis to augment pathologists scoring by predicting response to treatment would be provided.

9.3.0 Image analysis for different PD-L1 clones

This future study would also provide the opportunity to further compare 22C3 and SP263. As has been explored in detail already, 22C3 and SP263 produce generally very similar TPS results, but are not identical, and when tumours are close to critical cut-off thresholds the difference between the clones is particularly susceptible to grouping patients differently. Therefore, as with the image analysis algorithm, the predictive power of 22C3 compared to SP263 would be a valuable area to explore. Despite the similarity in staining, there are fundamental differences between the two clones: 22C3 is a Dako manufactured clone with a cytoplasmic epitope binding region on the PD-L1 protein and was used in the Merck sponsored Keynote trials which tested the efficacy of the anti-PD-1 monoclonal antibody pembrolizumab. SP263 is Roche-Ventana manufactured clone with an extracellular epitope binding region and was used in the Astrazeneca sponsored 'geography' trials which tested the efficacy of the anti-PD-1 mAb durvalumab. 22C3 was originally assessed using the ≥1% and ≥50% TPS cut-offs to determine patient groupings,[48, 255] and is a companion diagnostic for pembrolizumab for NSCLC (and many other cancer types) which continues to use these cut-offs. SP263 was originally assessed using a ≥25% TPS cut-off as 'PD-L1 positive'[642, 644] but is now used within NSCLC (and other cancers) as a complementary diagnostic with varying cut-offs on the specific IM agent.[251] Whilst many studies have compared the quantitative differences between 22C3 and SP263, few have had the data to compare their predictive power. Furthermore, to date there have been no clinical trials that involve a head-to-head comparison of PD-1/PD-L1 IMs, nor a trial that has used more than 1 PD-L1 IHC clone to guide clinical groupings. The differences between these two

clones maybe subtle, but they have potential clinical ramifications, and thus even this modestly sized cohort would provide an ideal opportunity to compare the 22C3 and SP263 clones' predictive power.

The use of digital image analysis in conjunction with IHC, both monoplex and multiplex, may therefore be a powerful tool in better predicting response to IM therapies, but as with mRNA and gene expression, IHC is not a perfect technique with many limitations, not least that its use in this manner to quantify proteins is overstretching its original function as a qualitative assay, as well as the fact IHC cannot reliably detect PD-L1 with certain post-translational modifications or on tissue sections that have had poor pre-analytical conditions. Alternative technologies may well be able to overcome some of these limitations, and has been explored in brief in Chapter 4 (pre-analytics), mass spectrometry is one such technique.

9.4 Mass-spectrometry (MS) and the proteome

MS is an established technique for studying substrates of unknown composition, typically by ionising the particles within the substrate, accelerating the charged particles to the same speed, and then using a magnetic or electric field to divide the particles by mass to charge ratio (m:z), on the principle that similarly sized particles will be deflected by the same degree, and so can be compared to known libraries of data. It is outside the scope of this thesis for an in depth review of the principles and various techniques of MS, but as per Chapter 4, collaboration with Protypia has allowed for the application of MS to FFPE tissue samples via liquid chromatography-tandem mass spectrometry (LC-MS), in which a sample mixture is separated by liquid chromatography before being ionised and characterised using two mass spectrometers in tandem, using m:z and relative abundance to qualify and quantify the peptides.

Different MS technologies have advantages and limitations; for example, the bottom-up approach to mass spectrometry (BU-MS) analyses peptides rather than whole proteins, requires extensive comparison to known libraries, is confounded by isoforms and does not technically sequence the protein. Top-down MS (TD-MS) on the other hand, can detect intact proteins, including sequence variations, but is much less sensitive and struggles to cope with proteins of larger molecular mass (>50kDa).[750, 751] The use of LC-MS allows for the analysis of more complex structures, a broader array of molecular weight proteins and improved multiplex abilities, and is widely used in clinical biochemistry and proteomic work[751-753] and is combined with BU-MS by Protypia for the analysis of FFPE tissue proteomes.[360] The use of MS on FFPE tissue has been explored by other groups, although concerns include uncertainties around mapping FFPE proteomes to libraries acquired from fresh tissue in that they may not be concordant, and that aldehyde induced changes may alter protein quality considerably.[754, 755] However, recent work has shown that LC-MS is potentially superior to

predicting immune checkpoint inhibitor expression, including PD-L1, than mRNA levels are, and has the potential ability to more accurately and reliably quantify PD-L1 and other immune markers than IHC.[756]

LC-MS can be utilised to provide a global proteome, allowing for the generation of 'deep proteotypes', as well as a targeted proteome that provides greater coverage of a smaller number of specific proteins. Therefore in a similar manner to multiplex IHC, immune TME signatures based on the presence of proteins may be defined by detection via MS.

By utilising this approach, the LLP cohort can be analysed to provide data on specific protein markers and test the feasibility of this technology on FFPE tissue by comparing the presence of proteins that have also been assessed by multiplex IHC in Chapter 7. Global analyses of the specimens will allow for the generation of prototypes based on the immune TME, to provide further data on the mechanisms that underpin PD-L1 expression and variation in immune TMEs by other methods.

9.5 Circulating biomarkers to predict response

Another area that has received much interest in the context of predicting response to IM therapy is the use of circulating biomarkers, with blood, serum, plasma or other extracted components used as the substrate for these assays. These are particularly attractive as blood is much easier and cheaper to acquire than tissue samples, allows for repeat sampling over time, and maybe less susceptible to issues pertaining to tissue heterogeneity.

These so called 'liquid biopsies' have seen success in NSCLC in other contexts; in particular the repeat sampling of patients to detect resistance associated mutations of the *EGFR* gene (e.g. T790M, C797S) and have been explored for predicting response to IM therapies.[757]

9.5.0 Circulating tumour cells and PD-L1 expression

PD-L1 in particular has been studied extensively as a circulating biomarker. Its presence can be detected on circulating tumour cells (CTCs), circulating exosomes, or even as a soluble form of PD-L1.

PD-L1 expression on CTCs has variable correlation with matched tissue samples, at best reaching 80% but in most studies achieving much poorer or no correlation at all,[758-763] and favours a poorer prognosis or response IM therapy in NSCLC.[762-767]

Perhaps unsurprisingly, but disappointingly, PD-L1 expression on CTCs suffers from intra-patient heterogenous expression, ranging from 3% to 100% of tested CTCs expressing PD-L1 in various patients.[761] Many studies considering the feasibility of PD-L1 on CTCs as a surrogate marker for PD-L1 expression on tissue samples conclude that the lack of concordance between these specimen

types renders it a biomarker of limited value. However, it can also be viewed as a positive in that PD-L1 on CTCs can potentially function as a distinct biomarker to PD-L1 expression on conventional tissue samples, and may therefore act as an additional predictive test, with more recent studies taking this approach.[768]

Another challenge of CTC PD-L1 expression is the varying techniques and assays that can be used to detect it, ranging across techniques such as EpCAM-based,[764, 767] size-based e.g. CellSieve Microfiltration Assay[769] or ISET (isolation by size of epithelial tumour cells)[763], Immunomagnetic depletion[770] and Microfluidic graphene oxide chip[765] approaches, the pros and cons of which are outside the scope of this thesis, but suffice to say they are not entirely concordant with each other.[771] CTCs are a route of metastases for many tumours, and their increased prevalence is typically associated with increasing overall tumour burden,[772, 773] and so it is perhaps unsurprising their presence is associated with worse survival and response to IMs regardless of PD-L1 expression.[772, 774] PD-L1 expression may be a remnant of their primary site dominant clone, although the disparity between tissue samples and CTC expression of PD-L1 may indicate a further, as yet unknown role of PD-L1 expressed on CTCs. Its more immediate application as a biomarker, however, is likely to yield success, albeit in patients with already poorer prognoses.

9.5.1 Exosomal and soluble PD-L1

Exosomal PD-L1 is thought to arise by endocytosis of plasma membrane in cells that express PD-L1 including tumour cells, macrophages, dendritic cells and so forth.[775, 776] Exosomal PD-L1 appears to play a direct role in immunosuppression that benefits cancerous cells with various mechanisms described that include the transportation of PD-L1 from tumour cells that express it to those that do not,[777] the inhibition of cytotoxic T-cells that can result in their dysfunction and eventual apoptosis, including circulating T-cells, via multiple mechanisms including the suppression of the MAPK and NF-κB mediated pathways of activating T-cells, the blockade of T-cell receptors and the suppression of Granzyme B, TNF-α and IL-2 production.[776, 778-781] As with PD-L1 expression on CTCs, exosomal PD-L1 expression is also associated with more advanced disease and generally poorer prognosis[779, 782, 783] and does not correlate particularly well with tissue PD-L1 expression,[782] although it may be associated with improved response to PD-1/PD-L1 IM therapy.[784] There are also multiple methods of detecting exosomal PD-L1, including ELISA (Enzyme linked immunosorbent assay), flow cytometry, and ddPCR (Digital Droplet PCR), each with advantages and disadvantages,[775] with the ideal approach yet to be agreed upon.

Soluble PD-L1 is another potential circulating biomarker that appears to be produced from exosomal PD-L1, splice variants of PD-L1, proteolytic cleavage of membrane bound PD-L1 or by other sources

such as cell injury or cell death.[172, 174, 785] As with other forms of circulating PD-L1, its presence is associated with poorer clinical outcomes in a variety of tumours, including NSCLC[172, 173, 368, 372, 786] and does not correlate with tissue expression of PD-L1.[172] There are various approaches to detecting and measuring soluble PD-L1, although these have largely been concerned with different platforms that use ELISA applied to either serum or plasma samples,[581, 787, 788] but the optimised approach to detecting all forms of soluble PD-L1 is yet to be elucidated.

There is some potential overlap of circulating forms of PD-L1 and multiple methods of detecting them, but they all individually appear to play functional roles that are outwith the conventional activity of PD-L1 in primary tissue specimens, and as a result of their lack of concordance and independent clinical associations, are prime targets to be utilised as additional biomarkers alongside PD-L1 IHC on conventional specimens.

9.3.2 Other circulating biomarkers

A large number of other circulating biomarkers to use for predicting response to IM therapy are also in active development. These include the study of circulating nucleic acids such as ctDNA (circulating DNA). Increasing levels of ctDNA are an early indicator of response to IM therapy[789-792] and whilst ctDNA is inherently well suited to detecting specific mutations (e.g. *EGFR* mutants) it can also be used to measure TMB[793-796] and MSI-H (Micro-satellite instability – High).[797, 798] The value of measuring MSI-H is that these tumours tend to be highly immunogenic, often expressing high levels of PD-L1 and associated with large quantities of TILs[799, 800] and therefore respond very favourably to IM therapy, which has seen these tumours studied in several IM clinical trials.[801, 802] The prevalence of MSI-H in NSCLC varies between studies, ranging from 1% to 25%,[803-806] but is a potential alternative to TMB for at least some patients.

Other circulating nucleic acids that could act as predictive biomarkers for IM therapy include RNA splice variants that produce circular RNAs (circRNA). The functional role of circRNAs is poorly understood, but thousands of circRNAs have been shown to be up or downregulated in NSCLC, including circRNAs related to genes that control tumour suppression, tumour promotion, apoptosis and other key genes relating to tumour development[807-812] and have shown in cell lines that some specific circRNAs may be involved in resistance to TKIs.[813, 814] Their role in predicting immunotherapy has yet to be fully understood, but they do appear to play a key role in modulating the TME including the inhibition of apoptosis in T-cells, enhancing cytotoxic T-cell activity, mediating the proliferation and differentiation of T-cells and B-cells and having activity on macrophages, fibroblasts, MDSCs, NK-cells and many other aspects of the TME.[812, 815] Measuring circRNA has several advantages to measuring ctDNA in that it is a highly dynamic biomarker that represents a wider variation of tumour

cell stages and can be highly specific for tumour cell activity, although it is less stable than ctDNA and is prone to much more non-cancer related variation between patients including age, gender and changes due to technical variation.[814] miRNA has received attention as a potential screening tool for NSCLC[816] or a predictor of survival in NSCLC,[817] and recent studies have showed the feasibility of using panels of miRNAs to predict response to PD-1/PD-L1 IM therapy in NSCLC.[818, 819]

The challenges with using nucleic acid based circulating biomarkers includes ascertaining which component of a blood sample is most appropriate for their detection, which variants of the thousands that can be detected are genuinely useful predictors, what assays and techniques to use, what should be used as a 'cut-off', individual patient variation and so forth, but the potential advantages of having easy access to repeat specimens will ensure continuing development in this area.

The use of routine haematology assays to predict response to IM therapy has also been considered. The LIPI (lung immune prognostic index) combines LDH (Lactate-dehydrogenase) and lymphocyte to neutrophil ratio,[699, 820] with others using the neutrophil-to-lymphocyte ratio (NLR) and platelet-to-lymphocyte ratio (PLR) in NSCLC patients, with a recent meta-analysis showing increased levels of NLR and PLR are generally associated with worse outcomes for NSCLC patients treated with IM therapy.[821] The modified Glasgow prognostic score, that uses CRP and albumin is considered a marker of systemic inflammation and has been recently shown to have potentially predictive power for response to PD-1/PD-L1 IM therapy in HNSCC.[822] The measurement of circulating cytokines has also seen some success as predictive biomarkers, with increased levels of IL-6,IL-10 and IFN-γ and reduced levels of IL-8 associated with superior response to IM-therapy[823, 824] with at least one attempt made to integrate measurement of serum cytokines into a 'cytoscore' made,[825] although there is little recent data on this area.

These daunting array of potential circulating biomarkers is far from exhaustive, with tumour educated platelets, metabolites and many others all potentially having a role.[826, 827] Some of these will undoubtedly play a major role in the future, but the example of the various forms of circulating PD-L1 exemplifies the difficulty in using circulating biomarkers: what specimen to use, which component to measure, what assays to apply, what definition/categories should be integrated are all critical questions with no easy answers. Even the notion that circulating biomarkers will bypass the challenge of PD-L1 expression heterogeneity does not appear to be true, and it seems quite apparent that they will not be able to be used as surrogate or replacement biomarkers for good quality tissue assessment. However, using circulating biomarkers to augment current and future biomarkers is a valid approach, and the fact that many of these biomarkers are more highly

expressed in advanced tumours, means the association of findings in patients in the CCC cohort whom have serum/plasma/blood already in storage, or for patients able to consent for additional blood testing, would be an excellent opportunity to validate at least some of these biomarkers in a robust PD-1/PD-L1 treated cohort of NSCLC patients.

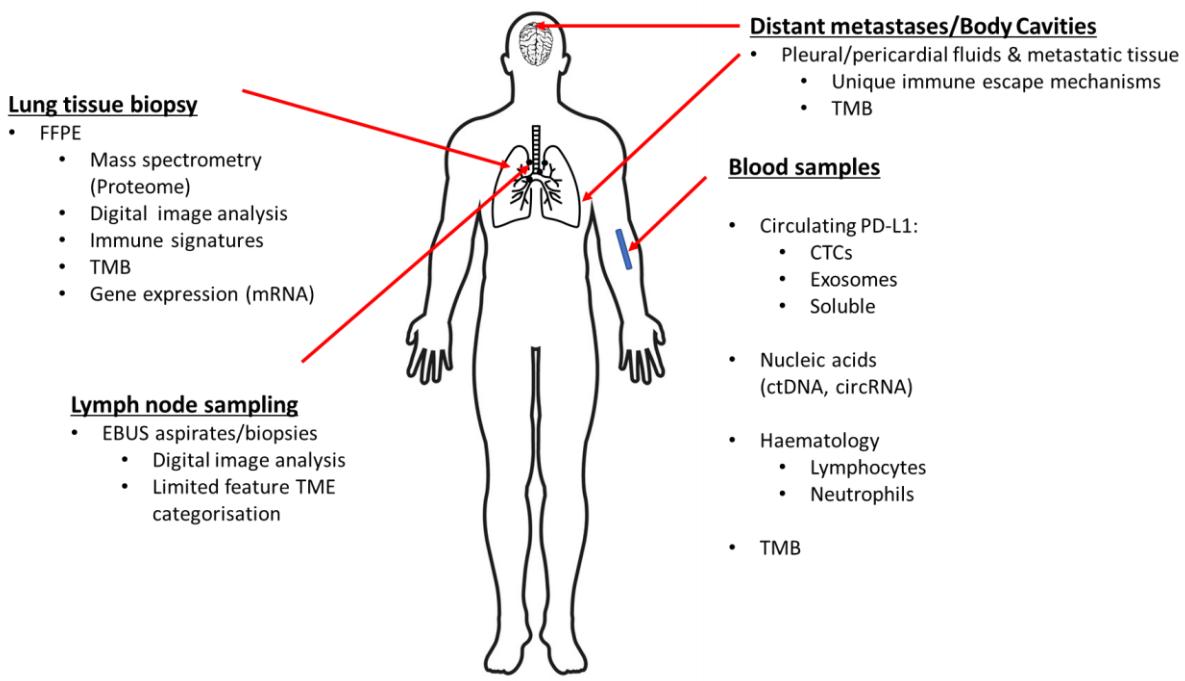


Fig 9.1.1 Different areas for sampling NSCLC specimens, the biological material and possible techniques that each could provide for better predicting response to PD-1/PD-L1 IM therapy. Many of these are future prospects as well as areas for further work following on from data from this thesis.

9.6 Predicting response in the near future

The future of predicting response to IM therapy is an exciting and important one. Unlike treatments with XCT, radiotherapy or even targeted therapies by TKIs, IM therapy has seen sustained, durable, long-term benefit in advanced NSCLC with a small cohort of long-term survivors tumour free even several years after treatment; a notion previously unthinkable for these patients.

However, an important point to note is that many clinical trials looking at IM therapy have an early group of cross-over in which patients treated with IM therapy rapidly progress and perform worse

than those with standard treatment of care; for these patients IM therapy is profoundly worse than other treatment options. (Fig 1.6.2)

Our ability to predict response to IM therapy is far short of where we would want it to be. PD-L1 IHC is the only clinically validated test for predicting response to IM therapy in NSCLC patients at present, but other biomarkers are sure to follow in the near future. The combination of new technologies, new targets and applying established biomarkers to new clinical situations has been reviewed briefly in this chapter (Fig 9.1.0 and Fig 9.1.1). The two main areas of interest likely to yield new biomarkers for predicting IM therapy is TMB and the features of the TME. Challenges for the use of TMB include which targets should be included to measure it and the types of mutations to include, the platforms to measure TMB, where to define 'high' and most importantly the conflicting evidence that TMB does or does not predict response to IM therapy in NSCLC.[267, 740, 742, 743] Using the TME has many limitations that include several similar problems in regards to what targets to measure, how to measure them, defining the levels of clinical relevance and so forth.

However, an important point to make is that these novel biomarkers involve the integration of measuring tumours by their genome (e.g. TMB), proteome (TME by IHC or MS) as well as the intermediary step of gene expression and splice variants (e.g. mRNA, circMRA) and therefore should not be seen as conflicting technologies but as complementary approaches that measure multiple aspects of the complex biology that is the interaction of tumours and the immune response.

Furthermore, as our ability to measure this complex biology improves, the need to integrate the data together to provide increasing personalised information for each patient is increasingly reliant on the use of digital pathology, machine learning, image analysis, AI and so forth in order to ensure the accurate and consistent application of these technologies. The opportunity to apply several of these technologies on both the LLP and CCC cohorts means comparisons, similarities and key differences between these platforms can be identified.

In an ideal near future, a patient with a new diagnosis of NSCLC may have, in addition to the standard clinical and radiological work-up, multiple tissue biopsies of the primary site, plus specimens from metastatic tissue if present, that will be tested for a TME immune signature and the presence of specific immune checkpoint proteins that includes the use of digital image analysis, submitted for a NGS-panel that incorporates actionable mutations/translocations etc. as well as a measure of TMB and a blood sample that will be used to look for circulating biomarkers such as CTCs expressing PD-L1. A patient will therefore have at the time of their first oncologist meeting data on their actionable targets, and in the context of IM therapy, a stratified profile that provides personalised data on their chances of response to IM therapy and their risk of irAEs, with monitoring

of response to treatment and follow-up by blood samples that can be used to adjust their patient profile of response and clinical outcomes.

This style of approach in the future must consider multiple biomarkers in the context of each other, and the use of more flexible groupings, that are prone to change overtime. Alongside new treatments being developed in the IM field such as small molecule inhibitors of the PD-1/PD-L1 axis,[828] new checkpoint inhibitors such as IDO, TIM3, LAG3, VISTA,[829] CAR-T cells and antibody conjugates.[238, 830] This thesis has highlighted just some of the fundamental challenges of using PD-L1 IHC, and the lessons that have been learned in the wider community in order to ensure its accurate use. It is vitally important the issues and errors that have limited PD-L1 IHC as a predictive tool are not repeated each time a new biomarker or assay is developed. There is no easy solution to this, but a move away from 'blockbuster' style drugs to stratified or even personalised medicine is a welcome acceptance that as each patient and each tumour is unique, so should be the approach to subtyping tumours and deciding treatments.

Chapter 10 - Discussion

The main aim of this thesis was to achieve a more comprehensive understanding of the factors that influence PD-L1 expression so as to ultimately provide knowledge to further improve the prediction of NSCLC patient response to PD-1/PD-L1 IM therapy. The combined outcomes and achievements of this project are grouped into two main categories: optimising the use of PD-L1 IHC in patient management, and the analysis of additional biomarkers to use alongside PD-L1 IHC to improve its clinical use in immunotherapy of lung cancer patients

10.0 Strengths and weaknesses of PD-L1 expression by IHC

This thesis has comprehensively reviewed the current body of knowledge pertaining to the strengths and limitations of using PD-L1 expression by IHC as a biomarker. Combining the literature review with findings from Chapter 4, 'The effect of pre-analytical conditions on PD-L1 expression' , and Chapter 5, 'Analytics and post-analytics: Digital pathology and PD-L1 interpretation', it can be inferred that some conditions have virtually no negative impact on PD-L1 expression, such as different sampling methodology[282, 300, 302] or the use of minimally aged tissue. [272, 283-285] A number of other factors require optimisation and careful regulation in order to minimise aberrant PD-L1 expression; fixation of specimens and storage conditions are critical but relatively straightforward areas where this can be achieved.[272, 274, 275, 290] The choice of PD-L1 IHC clones is an extensively studied area, and there is wide acknowledgement that inter-observer variation of PD-L1 expression interpretation is a significant challenge in achieving consistent results.[304, 335] However, data from this thesis reinforce the argument that an appreciation and awareness of the specific benefits and limitations of each clone, together with quality assurance schemes and good communication with clinical colleagues seem the most sensible approaches in choosing and optimising a specific PD-L1 IHC assay. The value of digital pathology and machine-learning image analysis algorithms to assist with interpretation of PD-L1 expression is a currently unproven area,[527, 531, 533] but a promising forefront to modern pathology, with data described within this thesis providing promising initial results.

Conversely, certain aspects of PD-L1 expression that undoubtedly impact on its predictive power remain a challenging area with no simple solution, most notably that of its heterogeneity. In Chapter 6, 'Heterogeneity of PD-L1 expression', a comprehensive analysis of both intra-tumoural and inter-tumoural heterogeneity was performed which, in keeping with many previous studies, demonstrates it to be a significant and common problem.[380, 387, 390, 831] Heterogeneity represents a biological and clinical challenge. Biological considerations, the reason for heterogeneity, underlying mechanisms and how this informs us concerning other aspects of the TME are questions to which our knowledge

is currently limited,[832] but Chapter 7, 'The tumour microenvironment', provides data to explain the inter-relationship between PD-L1 and other immune markers within the TME and suggests how this may be practically used by the clinical pathologist. The clinical challenges of PD-L1 expression heterogeneity are no less problematic.[266, 394] Questions such as whether different patterns of heterogeneity impact predictive power, if these patterns can yield information about toxicity and irAEs, or conundrums such as the correct therapeutic decision for a patient with a strongly PD-L1 positive primary tumour and entirely PD-L1 negative metastatic disease remains unsatisfactorily answered.

Ultimately, the current best solution to the issue of PD-L1 heterogeneity is to sample as widely and 'generously' as possible using the most contemporary tissue available. As molecular pathology continues to develop towards more comprehensive NGS-based panel assessments, the requirement for increasing quantities and qualities of tissue increases,[678] and this will provide an excellent opportunity to further study tumoural heterogeneity, and in time provide us with solutions to further minimise the impact of this phenomenon.

10.0.1 The predictive power of PD-L1 expression

The fact that, broadly speaking, PD-L1 expression by IHC works as a predictive biomarker is not in doubt.[46, 47, 833] In Chapter 8, 'Predicting response to PD-1/PD-L1 immunomodulatory therapy', it is clear that PD-L1 expression is usually associated with improved response to PD-1 IM therapy in NSCLC patients. This is particularly notable in the 1st-line setting, but the relationship with PD-L1 expression becomes more complex in the 2nd-line or 3rd-line settings, with the very highest expressors paradoxically representing a trend towards decreased response to therapy. Indeed, regardless of the specific treatment setting, it is consistently found that some patients with tumours that have high levels of PD-L1 expression do not respond favourably to treatment, whereas some patients with PD-L1 negative tumours do respond well.[46, 47, 286, 834] Whilst it is appreciated that some of this uncertainty may be explained by heterogeneity or even technical factors, it is clear that even the most widely sampled tumour, with the most optimised tissue pathway and careful interpretation of PD-L1 IHC cannot guarantee an accurate prediction of response to IM therapy. Additional biomarkers are almost certainly critical to enhance the predictive power of PD-L1 IHC.

10.1 Alternative biomarkers

Potential additional biomarkers to assist in predicting response to IM therapy are legion, with many studies and many groups focusing on this forefront area of research.[603, 674, 818, 835] These biomarkers can be broken down into three main categories: alternative measurements of the PD-L1 protein, the

presence and quantity of other immune-checkpoints, and the systemic and localised activity of the host immune response.

10.1.0 Alternative biomarkers: alternative methods of measuring PD-L1

PD-L1 IHC is but one way of measuring the PD-L1 protein and, whilst it is the only clinically validated technique, it has been demonstrated in previous studies that IHC may not be able to detect all forms of PD-L1; for example, PD-L1 protein with extensive post-translational modifications.[360] In Chapter 4, (Pre-analytics) and in Chapter 7 'The tumour microenvironment', PD-L1 protein was assessed not only by conventional IHC, but also by mass-spectrometry and multiplexed IHC and IF. Reassuringly, the correlation between these techniques and conventional IHC is, broadly speaking, good. This suggests that when looking for the membrane bound form of PD-L1 within tumours, the technique is less important than the quality and handling of the specimen. However, when considering future directions in Chapter 9 'The Next Frontier: Improving the prediction of response to PD-1/PD-L1 IMs', it is evident from multiple studies that circulating forms of PD-L1 to do not seem to correlate with tissue PD-L1 IHC expression levels.[172, 768, 782] This implies that PD-L1 in these alternative locations can act as additional biomarkers rather than as a replacement or surrogate for conventional PD-L1 IHC. Given the power that PD-L1 IHC has as a predictive biomarker, these alternative measurements of the PD-L1 protein seem like sensible areas to pursue.

10.1.1 Alternative biomarkers: lessons from single driver mutations

An important consideration is the inter-relationship of different immune-checkpoints within the TME, and much regarding these can be learned from the extensively studied area of single-driver mutations. Single-driver mutation status in NSCLC are generally mutually exclusive. For example, a tumour harbouring mutant *EGFR* will typically be wild-type for *KRAS* or *BRAF* and vice versa.[25] A shared feature of driver mutations with PD-L1 expression is clonal and sub-clonal heterogeneity within tumours, with even complex heterogeneous metastatic tumours able to be 'mapped' by their genetic aberrations.[544, 546] Furthermore, certain mutants, such as *EGFR* T790M and C797S mutations are rarely encountered *de novo*, but are a relatively common occurrence in the setting of patients treated with *EGFR* TKIs.[836, 837]

Whilst the fundamental mechanisms of immune-checkpoints and single driver mutations are distinct, shared features, particularly from a clinical perspective, are worth considering. For example the question of when to give a third generation *EGFR* TKI is complex. Osimertinib given 2nd-line to patients who have received an earlier generation TKI and developed a T790M mutation has been shown to improve outcome, but recent data from the FLAURA study showed osimertinib to be highly

effective in the 1st-line setting.[837, 838] Whilst the underlying mechanisms indicate no rationale for dual prescribing of an *EGFR* TKI with a *BRAF* TKI, the same is not certain for IMs, nor is the best approach to multiple immune checkpoint treatment yet established.

Several studies have combined PD-1/PD-L1 and CTLA4 IMs, on the basis that CTLA4 inhibition might increase PD-L1 expression, and that dual blockade may upregulate T-cell recruitment to the TME.[231, 243, 690, 839, 840] Whilst many studies have shown some promise in this area, others have not, probably in part due to the differing combinations: the tremelimumab and durvalumab combination has shown underwhelming outcomes, whereas ipilimumab and nivolumab appears to show very promising results.[840, 841] In addition, whilst both PD-L1 expression by IHC and high TMB are used as biomarkers of response, there are mixed results on how effective their predictive power might be.[232, 236, 743, 841, 842] It is of note the high frequencies of severe irAEs encountered in these trials[686, 687] serves as a warning to future potential combination therapies; multiple immune-checkpoint blockade might be achievable, but the cost in toxicity may prove to be too high.

An outstanding question is the relationship of differing immune-checkpoints to each other, and how blockade of one may induce another, or even induce as yet unknown iatrogenic resistance mechanisms similar to the emergence of the T790M mutation post-*EGFR* TKI treatment.[21] Work that has looked at the concurrence of IDO1 and PD-L1, for example, has shown that most NSCLCs do not co-express these immune-checkpoints.[756, 843, 844] For those that do, it may be an indicator of a more aggressive phenotype,[845] and might also be a marker of a tumour that shows particular response to IM therapy.[844, 846] Furthermore, the blockade of PD-1/PD-L1 has been shown to be associated with the upregulation of alternative immune checkpoints, such as TIM3.[244] It is, therefore, possible that a sequence of IM therapies, rather than combination therapies, may prove to be the best approach.

To further complicate matters, high levels of many immune-checkpoint markers, including PD-1, TIM3 and VISTA, are associated with T-cell exhaustion[225, 226, 640, 847] and are therefore indicative of a TME that will not respond favourably to IM therapy. Combined with the findings from Chapter 8 'Predicting response to PD-1/PD-L1 immunomodulatory therapy', in that the expression of PD-L1 at very high levels may lose predictive power in the 2nd-line setting, this infers both timing and previous treatment might be important factors in predicting response to IMs. It is likely that the presence and relationship of different immune checkpoints and the ascertainment of methods to measure their change over time will be a key step in optimising treatment pathways.

To borrow again from single-driver mutations, the value of circulating biomarkers may prove to be key. The FDA has approved the measurement of circulating *EGFR* mutations since 2016[848] and their presence in this capacity is a precursor to radiological and clinical progression. It is a tantalising

prospect that a similar achievement can be attained in the IM area by considering PD-L1 and other circulating immune checkpoints.

10.2 Measuring host immunity

The final piece in the puzzle in optimising predictive power is utilising knowledge of the systemic and localised host immune response. Unfortunately, there is no single established method for measuring generalised host immune ability as this is an incredibly complex, polyfactorial area with wide variation even between healthy individuals.[849]. Most attempts to measure systemic immune function in the context of cancer treatment have focused on circulating leukocytes, but these approaches have yet to be established in the routine clinical setting.[689, 850-852]

In vitro methods for studying immune response has had more success, and it is with these in mind that in Chapter 7, 'The tumour microenvironment', the project attempted to find practical ways in which to analyse and categorise TME. Many different approaches across many different tumour types have been previously suggested, from T-cell scores to 'immune hot' and 'immune cold', and the integration of digital pathology, image analysis and AI.[594, 597, 598, 600, 601, 672, 673] The major limitation of most of these studies is that they have been performed on large specimens, notably almost always with resected surgical specimens of NSCLC. The work described in Chapter 7 is confounded by the same potential problem; a technique that can be applied to sections of tissue that may be several centimetres squared may be wholly impractical to apply to the scanty specimens often received to make a diagnosis of NSCLC. In Chapter 8, 'Predicting response to PD-1/PD-L1 immunomodulatory therapy', this question was directly addressed with the sobering conclusion that perhaps only a third of these diagnostic specimens might be suitable for the more detailed analyses. None the less, the appeal of subtyping TMEs is a powerful one. Previous studies describe some success in their ability as predictive biomarkers[594, 598] and they might also prove to be a 'pan-tumour' biomarker of response to IM therapy. If the specific conditions for an IM conducive TME can be ascertained, this, combined with knowledge of the immune checkpoint(s) present may yield powerful predictive data. It may also provide more granularity in predicting irAEs. Whilst irAEs have been associated with increased response to PD-1/PD-L1 IMs in previous studies and in the CCC cohort studied in this thesis, the same is not generally true of CTLA4 inhibitors. It therefore follows that other immune-checkpoint IM therapy response rates will be variably associated with irAEs. [114, 680, 688] As these are a frequent and often life-threatening complication of IM therapy, this is an equally important aspect for prediction by analysing the TME.

10.3 Limitations

A major aim of this thesis was to provide definitive data to improve the predictive power of PD-L1 IHC and a significant limitation is that the majority of it utilised cohorts of patients that do not have clinical data pertaining to response to PD-1/PD-L1 IM therapy. In particular, the LLP cohort, which has been used extensively, whilst having robust clinical data and extensive quantities of good quality tissue, is limited in that none of the patients received IM therapy, as they were recruited prior to the IM era. Furthermore, although the LLP cohort was selected due to their being composed entirely of patients with metastatic nodal disease, the findings of this thesis may differ for a study of tumours with no metastases or a different pattern of metastases. This is less likely to be an issue when studying the technical aspects of PD-L1 IHC, but maybe more relevant for biological changes. However, as all lymph nodes were sampled at the time of resection of the primary tumour, and no patient received neoadjuvant therapy, the impact of temporal and iatrogenic heterogeneity can at least be disregarded. The large quantity of tissue in the LLP cohort has also allowed for a diverse range of studies to be performed and future work will allow for a comprehensive study of the genomic, proteomic and immune profile of these tumours. It has also enabled this project to minimise tissue usage from patient specimens where the quantity of tissue is limited; a recurring issue for diagnostic NSCLC specimens.

Another potential limitation is the use of novel approaches that have yielded interesting data, but have not been validated in multiple cohorts. In Chapter 4 (Pre-analytics) an acceleration chamber was used to study the effects of antigen degradation by varying the atmospheric conditions within the chamber. This reproduced loss of DAB staining equivalent to several months' or years' worth of storage time in a matter of weeks. It would be an interesting future project to trial out various other tissue types and IHC assays to ascertain whether consistency of loss can be achieved through the use of this chamber. In Chapter 6 (PD-L1 heterogeneity), the novel concept of 'simulated digital core biopsies' and 'digital FNAs' was developed. These provided a fast and effective way of simulating the sampling of tumours tens of thousands of times to provide robust data on the accuracy of small samples to overcome tumoural heterogeneity of PD-L1 expression. This preserves tissues and allows for 'sampling' of the same tumour many times over, which is impossible to achieve by 'real' sampling. It is not a perfect simulation, however, and falls prey to being limited to the analysis of 2D sections of tissue from a 3D tumour, and assumes a quantity and quality of tissue from biopsies and cytology specimens that is not always seen in the routine clinical setting.

In Chapter 7 (Tumour Microenvironment) novel TME subtypes were used to help rationalise the use of multiple immune markers alongside PD-L1 expression. In addition to the above limitations, the

definition of certain immune TMEs could likely be improved and refined with more biomarkers. Furthermore, the definition of any immune marker as 'high' or 'low' was entirely self-relative; 'high' in this cohort may be 'moderate' in another. It is probably reasonable to make some assumptions about certain findings however, for example, a tumour heavily infiltrated by lymphocytes with a 'high' CD3+ve phenotype is probably quite reasonably categorised as a tumour with significant TIL presence. The use of novel immune TME signatures is a sensible approach to translate complex multiple immune markers into a clinically useful tool, but more work is required to refine this approach.

10.4 Conclusion

The expression of PD-L1 in NSCLCs by IHC is a powerful, useful and, overall, a clinically accurate biomarker for predicting response to PD-1/PD-L1 IM therapy. Despite the many different ways this thesis has highlighted how the predictive power of PD-L1 IHC may be impaired, it is, nonetheless, a biomarker with a reassuringly robust nature. As the only clinically validated tool for this purpose to date, it is important that the optimisation and improvement of PD-L1 IHC continues. It is likely to remain at the forefront of predicting response to IM therapy for some time.

In spite of these general positivises, a significant proportion of patients will not have their response to IM treatment accurately predicted. Thus some patients receive treatment from which they derive no benefit, exposing them to a high risk of potential toxicity in the process, and others will be denied it inappropriately. Despite the successes of PD-L1 expression as a biomarker, improved predictive power is an important area of clinical need. This thesis has identified many key aspects in this regard and has provided data to help mitigate several of them. A combination of highly optimised PD-L1 IHC combined with the novel approaches outlined throughout it might assist in stratifying patients for PD-1/PD-L1 IM therapy. Future in this field, based on the results of this thesis, are an exciting area of development in the future.

References

1 Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin* 2020;70(1):7-30.

2 (CRUK) CRU. Lung Cancer Statistics. Vol. 2019: CRUK, 2019.

3 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. *J Thorac Oncol* 2015;10(9):1243-1260.

4 Scagliotti GV, Parikh P, von Pawel J*, et al.* Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naive patients with advanced-stage non-small-cell lung cancer. *J Clin Oncol* 2008;26(21):3543-3551.

5 Fergenbaum JH, Garcia-Closas M, Hewitt SM*, et al.* Loss of antigenicity in stored sections of breast cancer tissue microarrays. *Cancer Epidemiol Biomarkers Prev* 2004;13(4):667-672.

6 Gadgeel SM. Role of Chemotherapy and Targeted Therapy in Early-Stage Non-Small Cell Lung Cancer. *Am Soc Clin Oncol Educ Book* 2017;37:630-639.

7 Johnson DH, Fehrenbacher L, Novotny WF*, et al.* Randomized phase II trial comparing bevacizumab plus carboplatin and paclitaxel with carboplatin and paclitaxel alone in previously untreated locally advanced or metastatic non-small-cell lung cancer. *J Clin Oncol* 2004;22(11):2184-2191.

8 Sandler A, Gray R, Perry MC*, et al.* Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer. *N Engl J Med* 2006;355(24):2542-2550.

9 Reck M, von Pawel J, Zatloukal P*, et al.* Overall survival with cisplatin-gemcitabine and bevacizumab or placebo as first-line therapy for nonsquamous non-small-cell lung cancer: results from a randomised phase III trial (AVAiL). *Annals of oncology : official journal of the European Society for Medical Oncology* 2010;21(9):1804-1809.

10 Liu TC, Jin X, Wang Y*, et al.* Role of epidermal growth factor receptor in lung cancer and targeted therapies. *Am J Cancer Res* 2017;7(2):187-202.

11 FDA. List of Cleared or Approved Companion Diagnostic Devices (In Vitro and Imaging Tools). Vol. 2020, 2020.

12 Fukuoka M, Yano S, Giaccone G*, et al.* Multi-institutional randomized phase II trial of gefitinib for previously treated patients with advanced non-small-cell lung cancer (The IDEAL 1 Trial) [corrected]. *J Clin Oncol* 2003;21(12):2237-2246.

13 Kris MG, Natale RB, Herbst RS*, et al.* Efficacy of gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase, in symptomatic patients with non-small cell lung cancer: a randomized trial. *JAMA* 2003;290(16):2149-2158.

14 Thatcher N, Chang A, Parikh P*, et al.* Gefitinib plus best supportive care in previously treated patients with refractory advanced non-small-cell lung cancer: results from a randomised, placebo-controlled, multicentre study (Iressa Survival Evaluation in Lung Cancer). *Lancet* 2005;366(9496):1527-1537.

15 Herbst RS, Giaccone G, Schiller JH*, et al.* Gefitinib in combination with paclitaxel and carboplatin in advanced non-small-cell lung cancer: a phase III trial--INTACT 2. *J Clin Oncol* 2004;22(5):785-794.

16 Gatzemeier U, Pluzanska A, Szczesna A*, et al.* Phase III study of erlotinib in combination with cisplatin and gemcitabine in advanced non-small-cell lung cancer: the Tarceva Lung Cancer Investigation Trial. *J Clin Oncol* 2007;25(12):1545-1552.

17 Mitsudomi T, Morita S, Yatabe Y*, et al.* Gefitinib versus cisplatin plus docetaxel in patients with non-small-cell lung cancer harbouring mutations of the epidermal growth factor receptor (WJTOG3405): an open label, randomised phase 3 trial. *Lancet Oncol* 2010;11(2):121-128.

18 Maemondo M, Inoue A, Kobayashi K*, et al.* Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. *N Engl J Med* 2010;362(25):2380-2388.

19 Pirker R, Herth FJ, Kerr KM*, et al.* Consensus for EGFR mutation testing in non-small cell lung cancer: results from a European workshop. *J Thorac Oncol* 2010;5(10):1706-1713.
20 Harrison PT, Vyse S, Huang PH. Rare epidermal growth factor receptor (EGFR) mutations in non-small cell lung cancer. *Semin Cancer Biol* 2020;61:167-179.
21 Morgillo F, Della Corte CM, Fasano M*, et al.* Mechanisms of resistance to EGFR-targeted drugs: lung cancer. *ESMO Open* 2016;1(3):e000060.
22 Tan AC, Teh YL, Lai GGY*, et al.* Third generation EGFR TKI landscape for metastatic EGFR mutant non-small cell lung cancer (NSCLC). *Expert Rev Anticancer Ther* 2019;19(6):431-435.
23 Shi Y, Au JS, Thongprasert S*, et al.* A prospective, molecular epidemiology study of EGFR mutations in Asian patients with advanced non-small-cell lung cancer of adenocarcinoma histology (PIONEER). *J Thorac Oncol* 2014;9(2):154-162.
24 Zhang Y-L, Yuan J-Q, Wang K-F*, et al.* The prevalence of EGFR mutation in patients with non-small cell lung cancer: a systematic review and meta-analysis. *Oncotarget* 2016;7(48):78985-78993.
25 Skoulidis F, Heymach JV. Co-occurring genomic alterations in non-small-cell lung cancer biology and therapy. *Nat Rev Cancer* 2019;19(9):495-509.
26 Baseri B, Samra B, Tam E*, et al.* An Exceptional Responder to Nivolumab in Metastatic Non-Small-Cell Lung Cancer: A Case Report and Literature Review of Long-Term Survivors. *Case Rep Oncol Med* 2019;2019:1816472.
27 Ito K, Oguri T, Takeda N*, et al.* A case of non-small cell lung cancer with long-term response after re-challenge with nivolumab. *Respir Med Case Rep* 2020;29:100979.
28 Garon EB, Hellmann MD, Rizvi NA*, et al.* Five-Year Overall Survival for Patients With Advanced NonSmall-Cell Lung Cancer Treated With Pembrolizumab: Results From the Phase I KEYNOTE-001 Study. *J Clin Oncol* 2019;37(28):2518-2527.
29 Oiseth SJ, Aziz MS. Cancer immunotherapy: a brief review of the history, possibilities, and challenges ahead. *Journal of Cancer Metastasis and Treatment* 2017;3:250-261.
30 Busch W. Aus der Sitzung der medicinischen Section vom 13 November 1867. *Berl Klin Wochenschr* 1868;5:137.
31 Fehleisen F. Ueber die Züchtung der Erysipelkokken auf künstlichem Nährboden und ihre Übertragbarkeit auf den Menschen. *Dtsch Med Wochenschr* 1882;8(31):553-554.
32 Coley WB. II. Contribution to the Knowledge of Sarcoma. *Ann Surg* 1891;14(3):199-220.
33 Coley WB. The treatment of malignant tumors by repeated inoculations of erysipelas. With a report of ten original cases. 1893. *Clin Orthop Relat Res* 1991(262):3-11.
34 Coley WB. The Treatment of Inoperable Sarcoma by Bacterial Toxins (the Mixed Toxins of the Streptococcus erysipelas and the Bacillus prodigiosus). *Proc R Soc Med* 1910;3(Surg Sect):1-48.
35 Parish CR. Cancer immunotherapy: the past, the present and the future. *Immunol Cell Biol* 2003;81(2):106-113.
36 Fleming A. On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of B. influenzæ. *Br J Exp Pathol* 1929;10(3):226-236.
37 Morales A, Eidinger D, Bruce AW. Intracavitary Bacillus Calmette-Guerin in the treatment of superficial bladder tumors. *J Urol* 1976;116(2):180-183.
38 Alhunaidi O, Zlotta AR. The use of intravesical BCG in urothelial carcinoma of the bladder. *Ecancermedicalscience* 2019;13:905.
39 Leach DR, Krummel MF, Allison JP. Enhancement of antitumor immunity by CTLA-4 blockade. *Science* 1996;271(5256):1734-1736.
40 Traynor K. Ipilimumab approved for metastatic melanoma. *Am J Health Syst Pharm* 2011;68(9):768.
41 Raedler LA. Keytruda (Pembrolizumab): First PD-1 Inhibitor Approved for Previously Treated Unresectable or Metastatic Melanoma. *Am Health Drug Benefits* 2015;8(Spec Feature):96-100.

42 Raedler LA. Opdivo (Nivolumab): Second PD-1 Inhibitor Receives FDA Approval for Unresectable or Metastatic Melanoma. *Am Health Drug Benefits* 2015;8(Spec Feature):180-183.
43 Kazandjian D, Suzman DL, Blumenthal G*, et al.* FDA Approval Summary: Nivolumab for the Treatment of Metastatic Non-Small Cell Lung Cancer With Progression On or After Platinum-Based Chemotherapy. *Oncologist* 2016;21(5):634-642.
44 Pai-Scherf L, Blumenthal GM, Li H*, et al.* FDA Approval Summary: Pembrolizumab for Treatment of Metastatic Non-Small Cell Lung Cancer: First-Line Therapy and Beyond. *The oncologist* 2017;22(11):1392-1399.
45 Weinstock C, Khozin S, Suzman D*, et al.* U.S. Food and Drug Administration Approval Summary: Atezolizumab for Metastatic Non–Small Cell Lung Cancer. *Clinical Cancer Research* 2017;23(16):4534-4539.
46 Brahmer J, Reckamp KL, Baas P*, et al.* Nivolumab versus Docetaxel in Advanced Squamous-Cell Non-Small-Cell Lung Cancer. *N Engl J Med* 2015;373(2):123-135.
47 Borghaei H, Paz-Ares L, Horn L*, et al.* Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer. *N Engl J Med* 2015;373(17):1627-1639.
48 Reck M, Rodríguez-Abreu D, Robinson AG*, et al.* Pembrolizumab versus Chemotherapy for PD-L1–Positive Non–Small-Cell Lung Cancer. *New England Journal of Medicine* 2016;375(19):1823-1833.
49 Rittmeyer A, Barlesi F, Waterkamp D*, et al.* Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. *Lancet* 2017;389(10066):255-265.
50 Antonia SJ, Villegas A, Daniel D*, et al.* Overall Survival with Durvalumab after Chemoradiotherapy in Stage III NSCLC. *New England Journal of Medicine* 2018;379(24):2342-2350.
51 Fuchs CS, Doi T, Jang RW*, et al.* Safety and Efficacy of Pembrolizumab Monotherapy in Patients With Previously Treated Advanced Gastric and Gastroesophageal Junction Cancer: Phase 2 Clinical KEYNOTE-059 Trial. *JAMA oncology* 2018;4(5):e180013-e180013.
52 Ferris RL, Blumenschein G, Fayette J*, et al.* Nivolumab for Recurrent Squamous-Cell Carcinoma of the Head and Neck. *New England Journal of Medicine* 2016;375(19):1856-1867.
53 Schmid P, Adams S, Rugo HS*, et al.* Atezolizumab and Nab-Paclitaxel in Advanced Triple-Negative Breast Cancer. *New England Journal of Medicine* 2018;379(22):2108-2121.
54 Patel MR, Ellerton J, Infante JR*, et al.* Avelumab in metastatic urothelial carcinoma after platinum failure (JAVELIN Solid Tumor): pooled results from two expansion cohorts of an open-label, phase 1 trial. *Lancet Oncol* 2018;19(1):51-64.
55 Forde PM, Chaft JE, Smith KN*, et al.* Neoadjuvant PD-1 Blockade in Resectable Lung Cancer. *N Engl J Med* 2018;378(21):1976-1986.
56 Gao S, Li N, Gao S*, et al.* Neoadjuvant PD-1 inhibitor (Sintilimab) in NSCLC. *J Thorac Oncol* 2020;15(5):816-826.
57 Ahmadzadeh M, Johnson LA, Heemskerk B*, et al.* Tumor antigen-specific CD8 T cells infiltrating the tumor express high levels of PD-1 and are functionally impaired. *Blood* 2009;114(8):1537-1544.
58 Ishida Y, Agata Y, Shibahara K*, et al.* Induced expression of PD-1, a novel member of the immunoglobulin gene superfamily, upon programmed cell death. *EMBO J* 1992;11(11):3887-3895.
59 Chen L. Co-inhibitory molecules of the B7-CD28 family in the control of T-cell immunity. *Nat Rev Immunol* 2004;4(5):336-347.
60 Jiang X, Wang J, Deng X*, et al.* Role of the tumor microenvironment in PD-L1/PD-1-mediated tumor immune escape. *Mol Cancer* 2019;18(1):10.

61 Sharpe AH, Pauken KE. The diverse functions of the PD1 inhibitory pathway. *Nat Rev Immunol* 2018;18(3):153-167.
62 Patsoukis N, Li L, Sari D*, et al.* PD-1 increases PTEN phosphatase activity while decreasing PTEN protein stability by inhibiting casein kinase 2. *Mol Cell Biol* 2013;33(16):3091-3098.
63 Sun C, Mezzadra R, Schumacher TN. Regulation and Function of the PD-L1 Checkpoint. *Immunity* 2018;48(3):434-452.
64 Zou W, Wolchok JD, Chen L. PD-L1 (B7-H1) and PD-1 pathway blockade for cancer therapy: Mechanisms, response biomarkers, and combinations. *Sci Transl Med* 2016;8(328):328rv324.
65 Dong Y, Sun Q, Zhang X. PD-1 and its ligands are important immune checkpoints in cancer. *Oncotarget* 2017;8(2):2171-2186.
66 Lesterhuis WJ, Steer H, Lake RA. PD-L2 is predominantly expressed by Th2 cells. *Mol Immunol* 2011;49(1-2):1-3.
67 Yang H, Zhou X, Sun L*, et al.* Correlation Between PD-L2 Expression and Clinical Outcome in Solid Cancer Patients: A Meta-Analysis. *Front Oncol* 2019;9:47.
68 Yearley JH, Gibson C, Yu N*, et al.* PD-L2 Expression in Human Tumors: Relevance to Anti-PD-1 Therapy in Cancer. *Clinical Cancer Research* 2017;23(12):3158-3167.
69 Schmid P, Hegde PS, Zou W*, et al.* Association of PD-L2 expression in human tumors with atezolizumab activity. *Journal of Clinical Oncology* 2016;34(15_suppl):11506-11506.
70 Brown JA, Dorfman DM, Ma F-R*, et al.* Blockade of Programmed Death-1 Ligands on Dendritic Cells Enhances T Cell Activation and Cytokine Production. *The Journal of Immunology* 2003;170(3):1257-1266.
71 Dong H, Zhu G, Tamada K*, et al.* B7-H1, a third member of the B7 family, co-stimulates T-cell proliferation and interleukin-10 secretion. *Nat Med* 1999;5(12):1365-1369.
72 Freeman GJ, Long AJ, Iwai Y*, et al.* Engagement of the PD-1 immunoinhibitory receptor by a novel B7 family member leads to negative regulation of lymphocyte activation. *The Journal of experimental medicine* 2000;192(7):1027-1034.
73 Wang X, Teng F, Kong L*, et al.* PD-L1 expression in human cancers and its association with clinical outcomes. *Onco Targets Ther* 2016;9:5023-5039.
74 Qin W, Hu L, Zhang X*, et al.* The Diverse Function of PD-1/PD-L Pathway Beyond Cancer. *Front Immunol* 2019;10:2298.
75 Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer* 2012;12(4):252-264.
76 Mazanet MM, Hughes CC. B7-H1 is expressed by human endothelial cells and suppresses T cell cytokine synthesis. *J Immunol* 2002;169(7):3581-3588.
77 Liang SC, Latchman YE, Buhlmann JE*, et al.* Regulation of PD-1, PD-L1, and PD-L2 expression during normal and autoimmune responses. *Eur J Immunol* 2003;33(10):2706-2716.
78 Petroff MG, Chen L, Phillips TA*, et al.* B7 family molecules are favorably positioned at the human maternal-fetal interface. *Biol Reprod* 2003;68(5):1496-1504.
79 Ruocco E, Ruocco V, Tornesello ML*, et al.* Kaposi's sarcoma: etiology and pathogenesis, inducing factors, causal associations, and treatments: facts and controversies. *Clin Dermatol* 2013;31(4):413-422.
80 de Oliveria Andrade LJ, D'Oliveira A, Melo RC*, et al.* Association between hepatitis C and hepatocellular carcinoma. *J Glob Infect Dis* 2009;1(1):33-37.
81 Jiang J, Li J, Yunxia Z*, et al.* The role of prostatitis in prostate cancer: meta-analysis. *PLoS One* 2013;8(12):e85179.
82 Ehrlich P. Über den jetzigen Stand der Chemotherapie. *Berichte der deutschen chemischen Gesellschaft* 1909;42(1):17-47.
83 Burnet M. Cancer—A Biological Approach. *III Viruses Associated with Neoplastic Conditions IV Practical Applications* 1957;1(5023):841-847.
84 Thomas L. On immunosurveillance in human cancer. *Yale J Biol Med* 1982;55(3-4):329-333.
85 Stutman O. Immunodepression and malignancy. *Adv Cancer Res* 1975;22:261-422.

86 Engel AM, Svane IM, Rygaard J*, et al.* MCA sarcomas induced in scid mice are more immunogenic than MCA sarcomas induced in congenic, immunocompetent mice. *Scand J Immunol* 1997;45(5):463-470.
87 Kaplan DH, Shankaran V, Dighe AS*, et al.* Demonstration of an interferon gamma-dependent tumor surveillance system in immunocompetent mice. *Proc Natl Acad Sci U S A* 1998;95(13):7556-7561.
88 Dunn GP, Bruce AT, Ikeda H*, et al.* Cancer immunoediting: from immunosurveillance to tumor escape. *Nat Immunol* 2002;3(11):991-998.
89 Schreiber RD, Old LJ, Smyth MJ. Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. *Science* 2011;331(6024):1565-1570.
90 Quezada SA, Peggs KS, Simpson TR*, et al.* Shifting the equilibrium in cancer immunoediting: from tumor tolerance to eradication. *Immunol Rev* 2011;241(1):104-118.
91 Shankaran V, Ikeda H, Bruce AT*, et al.* IFNgamma and lymphocytes prevent primary tumour development and shape tumour immunogenicity. *Nature* 2001;410(6832):1107-1111.
92 Beatty GL, Gladney WL. Immune escape mechanisms as a guide for cancer immunotherapy. *Clin Cancer Res* 2015;21(4):687-692.
93 Aris M, Barrio MM, Mordoh J. Lessons from cancer immunoediting in cutaneous melanoma. *Clin Dev Immunol* 2012;2012:192719-192719.
94 Koebel CM, Vermi W, Swann JB*, et al.* Adaptive immunity maintains occult cancer in an equilibrium state. *Nature* 2007;450(7171):903-907.
95 Teng MW, Swann JB, Koebel CM*, et al.* Immune-mediated dormancy: an equilibrium with cancer. *J Leukoc Biol* 2008;84(4):988-993.
96 Bhatia A, Kumar Y. Cancer-immune equilibrium: questions unanswered. *Cancer Microenviron* 2011;4(2):209-217.
97 Alexandrov LB, Nik-Zainal S, Wedge DC*, et al.* Signatures of mutational processes in human cancer. *Nature* 2013;500(7463):415-421.
98 Garrido F, Ruiz-Cabello F, Aptsiauri N. Rejection versus escape: the tumor MHC dilemma. *Cancer Immunol Immunother* 2017;66(2):259-271.
99 Anagnostou V, Smith KN, Forde PM*, et al.* Evolution of Neoantigen Landscape During Immune Checkpoint Blockade in Non-Small Cell Lung Cancer. *Cancer Discovery* 2017:CD-16-0828.
100 Vinay DS, Ryan EP, Pawelec G*, et al.* Immune evasion in cancer: Mechanistic basis and therapeutic strategies. *Semin Cancer Biol* 2015;35 Suppl:S185-s198.
101 Kong X. Discovery of New Immune Checkpoints: Family Grows Up. *Adv Exp Med Biol* 2020;1248:61-82.
102 Alkasalias T, Moyano-Galceran L, Arsenian-Henriksson M*, et al.* Fibroblasts in the Tumor Microenvironment: Shield or Spear? *Int J Mol Sci* 2018;19(5).
103 Roma-Rodrigues C, Mendes R, Baptista PV*, et al.* Targeting Tumor Microenvironment for Cancer Therapy. *Int J Mol Sci* 2019;20(4).
104 Liu T, Han C, Wang S*, et al.* Cancer-associated fibroblasts: an emerging target of anti-cancer immunotherapy. *J Hematol Oncol* 2019;12(1):86.
105 Ribeiro AL, Okamoto OK. Combined effects of pericytes in the tumor microenvironment. *Stem Cells Int* 2015;2015:868475.
106 Wu Q, Li B, Li Z*, et al.* Cancer-associated adipocytes: key players in breast cancer progression. *Journal of hematology & oncology* 2019;12(1):95-95.
107 Zhu J, Paul WE. CD4 T cells: fates, functions, and faults. *Blood* 2008;112(5):1557-1569.
108 Togashi Y, Shitara K, Nishikawa H. Regulatory T cells in cancer immunosuppression - implications for anticancer therapy. *Nat Rev Clin Oncol* 2019;16(6):356-371.
109 Ostroumov D, Fekete-Drimusz N, Saborowski M*, et al.* CD4 and CD8 T lymphocyte interplay in controlling tumor growth. *Cell Mol Life Sci* 2018;75(4):689-713.

110 Goding SR, Wilson KA, Xie Y*, et al.* Restoring Immune Function of Tumor-Specific CD4⁺ T Cells during Recurrence of Melanoma. *The Journal of Immunology* 2013;190(9):4899-4909.
111 Thornton AM, Shevach EM. CD4+CD25+ immunoregulatory T cells suppress polyclonal T cell activation in vitro by inhibiting interleukin 2 production. *J Exp Med* 1998;188(2):287-296.
112 Wing K, Onishi Y, Prieto-Martin P*, et al.* CTLA-4 control over Foxp3+ regulatory T cell function. *Science* 2008;322(5899):271-275.
113 Jarnicki AG, Lysaght J, Todryk S*, et al.* Suppression of antitumor immunity by IL-10 and TGF-beta-producing T cells infiltrating the growing tumor: influence of tumor environment on the induction of CD4+ and CD8+ regulatory T cells. *J Immunol* 2006;177(2):896-904.
114 Deaglio S, Dwyer KM, Gao W*, et al.* Adenosine generation catalyzed by CD39 and CD73 expressed on regulatory T cells mediates immune suppression. *J Exp Med* 2007;204(6):1257-1265.
115 Grossman WJ, Verbsky JW, Barchet W*, et al.* Human T regulatory cells can use the perforin pathway to cause autologous target cell death. *Immunity* 2004;21(4):589-601.
116 Togashi Y, Nishikawa H. Regulatory T Cells: Molecular and Cellular Basis for Immunoregulation. *Curr Top Microbiol Immunol* 2017;410:3-27.
117 M; JCJTPW. T cell-mediated cytotoxicity. In: Charles A Janeway; Jr PTMWMJS, editor. Immunobiology: The Immune System in Health and Disease. 5th ed. New York: Garland Science, 2001.
118 Zhang N, Bevan MJ. CD8(+) T cells: foot soldiers of the immune system. *Immunity* 2011;35(2):161-168.
119 Okazaki T, Chikuma S, Iwai Y*, et al.* A rheostat for immune responses: the unique properties of PD-1 and their advantages for clinical application. *Nature Immunology* 2013;14(12):1212-1218.
120 Joller N, Peters A, Anderson AC*, et al.* Immune checkpoints in central nervous system autoimmunity. *Immunological reviews* 2012;248(1):122-139.
121 Riudavets M, Barba A, Maroto P*, et al.* Correlation between immune-related adverse events (irAEs) and efficacy in patients with solid tumors treated with immune-checkpoints inhibitors (ICIs). *Journal of Clinical Oncology* 2018;36(15_suppl):3064-3064.
122 Ma Y, Shurin GV, Peiyuan Z*, et al.* Dendritic cells in the cancer microenvironment. *J Cancer* 2013;4(1):36-44.
123 Fu C, Jiang A. Dendritic Cells and CD8 T Cell Immunity in Tumor Microenvironment. *Front Immunol* 2018;9:3059.
124 Hargadon KM. Tumor-altered dendritic cell function: implications for anti-tumor immunity. *Front Immunol* 2013;4:192.
125 Steinbrink K, Jonuleit H, Müller G*, et al.* Interleukin-10-treated human dendritic cells induce a melanoma-antigen-specific anergy in CD8(+) T cells resulting in a failure to lyse tumor cells. *Blood* 1999;93(5):1634-1642.
126 Sharma MD, Baban B, Chandler P*, et al.* Plasmacytoid dendritic cells from mouse tumor-draining lymph nodes directly activate mature Tregs via indoleamine 2,3-dioxygenase. *J Clin Invest* 2007;117(9):2570-2582.
127 Larmonier N, Fraszczak J, Lakomy D*, et al.* Killer dendritic cells and their potential for cancer immunotherapy. *Cancer Immunology, Immunotherapy* 2009;59(1):1.
128 Sun NY, Chen YL, Wu WY*, et al.* Blockade of PD-L1 Enhances Cancer Immunotherapy by Regulating Dendritic Cell Maturation and Macrophage Polarization. *Cancers (Basel)* 2019;11(9).
129 Orecchioni M, Ghosheh Y, Pramod AB*, et al.* Macrophage Polarization: Different Gene Signatures in M1(LPS+) vs. Classically and M2(LPS-) vs. Alternatively Activated Macrophages. *Front Immunol* 2019;10:1084.

130 Yang L, Zhang Y. Tumor-associated macrophages, potential targets for cancer treatment. *Biomark Res* 2017;5:25.
131 Murray PJ, Allen JE, Biswas SK*, et al.* Macrophage activation and polarization: nomenclature and experimental guidelines. *Immunity* 2014;41(1):14-20.
132 Grivennikov SI, Greten FR, Karin M. Immunity, inflammation, and cancer. *Cell* 2010;140(6):883-899.
133 Lu D, Ni Z, Liu X*, et al.* Beyond T Cells: Understanding the Role of PD-1/PD-L1 in Tumor-Associated Macrophages. *J Immunol Res* 2019;2019:1919082.
134 Singhal S, Stadanlick J, Annunziata MJ*, et al.* Human tumor-associated monocytes/macrophages and their regulation of T cell responses in early-stage lung cancer. *Sci Transl Med* 2019;11(479).
135 Hartley GP, Chow L, Ammons DT*, et al.* Programmed Cell Death Ligand 1 (PD-L1) Signaling Regulates Macrophage Proliferation and Activation. *Cancer Immunol Res* 2018;6(10):1260-1273.
136 Zhang X, Zeng Y, Qu Q*, et al.* PD-L1 induced by IFN-γ from tumor-associated macrophages via the JAK/STAT3 and PI3K/AKT signaling pathways promoted progression of lung cancer. *Int J Clin Oncol* 2017;22(6):1026-1033.
137 Lund FE, Randall TD. Effector and regulatory B cells: modulators of CD4+ T cell immunity. *Nat Rev Immunol* 2010;10(4):236-247.
138 Wennhold K, Shimabukuro-Vornhagen A, von Bergwelt-Baildon M. B Cell-Based Cancer Immunotherapy. *Transfus Med Hemother* 2019;46(1):36-46.
139 Ding T, Yan F, Cao S*, et al.* Regulatory B cell: New member of immunosuppressive cell club. *Hum Immunol* 2015;76(9):615-621.
140 Kessel A, Haj T, Peri R*, et al.* Human CD19(+)CD25(high) B regulatory cells suppress proliferation of CD4(+) T cells and enhance Foxp3 and CTLA-4 expression in T-regulatory cells. *Autoimmun Rev* 2012;11(9):670-677.
141 Guan H, Wan Y, Lan J*, et al.* PD-L1 is a critical mediator of regulatory B cells and T cells in invasive breast cancer. *Scientific reports* 2016;6:35651-35651.
142 Khan AR, Hams E, Floudas A*, et al.* PD-L1hi B cells are critical regulators of humoral immunity. *Nat Commun* 2015;6:5997.
143 Lecot P, Sarabi M, Pereira Abrantes M*, et al.* Neutrophil Heterogeneity in Cancer: From Biology to Therapies. *Front Immunol* 2019;10:2155.
144 Fridlender ZG, Sun J, Kim S*, et al.* Polarization of tumor-associated neutrophil phenotype by TGF-beta: "N1" versus "N2" TAN. *Cancer Cell* 2009;16(3):183-194.
145 Liou G-Y, Storz P. Reactive oxygen species in cancer. *Free Radic Res* 2010;44(5):479-496.
146 Wang TT, Zhao YL, Peng LS*, et al.* Tumour-activated neutrophils in gastric cancer foster immune suppression and disease progression through GM-CSF-PD-L1 pathway. *Gut* 2017;66(11):1900-1911.
147 Gershkovitz M, Yajuk O, Fainsod-Levi T*, et al.* The pd-l1/pd-1 axis blocks neutrophil cytotoxicity in cancer. *bioRxiv* 2020:2020.2002.2028.969410.
148 Paul S, Lal G. The Molecular Mechanism of Natural Killer Cells Function and Its Importance in Cancer Immunotherapy. *Front Immunol* 2017;8:1124.
149 Pesce S, Greppi M, Grossi F*, et al.* PD/1-PD-Ls Checkpoint: Insight on the Potential Role of NK Cells. *Front Immunol* 2019;10:1242.
150 Hsu JM, Li CW, Lai YJ*, et al.* Posttranslational Modifications of PD-L1 and Their Applications in Cancer Therapy. *Cancer Res* 2018;78(22):6349-6353.
151 André P, Denis C, Soulas C*, et al.* Anti-NKG2A mAb Is a Checkpoint Inhibitor that Promotes Anti-tumor Immunity by Unleashing Both T and NK Cells. *Cell* 2018;175(7):1731-1743.e1713.
152 Yang Z, Guo J, Weng L*, et al.* Myeloid-derived suppressor cells-new and exciting players in lung cancer. *J Hematol Oncol* 2020;13(1):10.

153 Law AMK, Valdes-Mora F, Gallego-Ortega D. Myeloid-Derived Suppressor Cells as a Therapeutic Target for Cancer. *Cells* 2020;9(3):561.
154 Noman MZ, Desantis G, Janji B*, et al.* PD-L1 is a novel direct target of HIF-1α, and its blockade under hypoxia enhanced MDSC-mediated T cell activation. *The Journal of experimental medicine* 2014;211(5):781-790.
155 Yamauchi Y, Safi S, Blattner C*, et al.* Circulating and Tumor Myeloid-derived Suppressor Cells in Resectable Non-Small Cell Lung Cancer. *Am J Respir Crit Care Med* 2018;198(6):777-787.
156 Limagne E, Richard C, Thibaudin M*, et al.* Tim-3/galectin-9 pathway and mMDSC control primary and secondary resistances to PD-1 blockade in lung cancer patients. *Oncoimmunology* 2019;8(4):e1564505.
157 Kim HR, Park SM, Seo SU*, et al.* The Ratio of Peripheral Regulatory T Cells to Lox-1(+) Polymorphonuclear Myeloid-derived Suppressor Cells Predicts the Early Response to Anti-PD-1 Therapy in Patients with Non-Small Cell Lung Cancer. *Am J Respir Crit Care Med* 2019;199(2):243-246.
158 Ni L, Lu J. Interferon gamma in cancer immunotherapy. *Cancer Med* 2018;7(9):4509-4516.
159 Castro F, Cardoso AP, Gonçalves RM*, et al.* Interferon-Gamma at the Crossroads of Tumor Immune Surveillance or Evasion. *Front Immunol* 2018;9:847.
160 Mojic M, Takeda K, Hayakawa Y. The Dark Side of IFN-gamma: Its Role in Promoting Cancer Immunoevasion. *Int J Mol Sci* 2017;19(1).
161 Wang X, Lin Y. Tumor necrosis factor and cancer, buddies or foes? *Acta Pharmacol Sin* 2008;29(11):1275-1288.
162 Garon EB, Chih-Hsin Yang J, Dubinett SM. The Role of Interleukin 1β in the Pathogenesis of Lung Cancer. *JTO Clinical and Research Reports* 2020;1(1):100001.
163 Millares L, Barreiro E, Cortes R*, et al.* Tumor-associated metabolic and inflammatory responses in early stage non-small cell lung cancer: Local patterns and prognostic significance. *Lung Cancer* 2018;122:124-130.
164 Lopez-Castejon G, Brough D. Understanding the mechanism of IL-1β secretion. *Cytokine & Growth Factor Reviews* 2011;22(4):189-195.
165 Sun N, Taguchi A, Hanash S. Switching Roles of TGF-beta in Cancer Development: Implications for Therapeutic Target and Biomarker Studies. *J Clin Med* 2016;5(12).
166 Azuma T, Yao S, Zhu G*, et al.* B7-H1 is a ubiquitous antiapoptotic receptor on cancer cells. *Blood* 2008;111(7):3635-3643.
167 Gato-Cañas M, Zuazo M, Arasanz H*, et al.* PDL1 Signals through Conserved Sequence Motifs to Overcome Interferon-Mediated Cytotoxicity. *Cell Rep* 2017;20(8):1818-1829.
168 Kim S, Jang J-Y, Koh J*, et al.* Programmed cell death ligand-1-mediated enhancement of hexokinase 2 expression is inversely related to T-cell effector gene expression in non-small-cell lung cancer. *Journal of Experimental & Clinical Cancer Research* 2019;38(1):462.
169 Clark CA, Gupta HB, Sareddy G*, et al.* Tumor-Intrinsic PD-L1 Signals Regulate Cell Growth, Pathogenesis, and Autophagy in Ovarian Cancer and Melanoma. *Cancer Res* 2016;76(23):6964-6974.
170 Chang CH, Qiu J, O'Sullivan D*, et al.* Metabolic Competition in the Tumor Microenvironment Is a Driver of Cancer Progression. *Cell* 2015;162(6):1229-1241.
171 Theivanthiran B, Evans KS, DeVito NC*, et al.* A tumor-intrinsic PD-L1/NLRP3 inflammasome signaling pathway drives resistance to anti-PD-1 immunotherapy. *J Clin Invest* 2020;130(5):2570-2586.
172 Asanuma K, Nakamura T, Hayashi A*, et al.* Soluble programmed death-ligand 1 rather than PD-L1 on tumor cells effectively predicts metastasis and prognosis in soft tissue sarcomas. *Sci Rep* 2020;10(1):9077.
173 Sun J, Qiu M-Z, Mei T*, et al.* Dynamic monitoring of serum soluble programmed cell death ligand 1 as a response predictor to chemotherapy in metastatic or recurrent gastrointestinal cancer. *Translational Cancer Research* 2020;9(4):2434-2448.

174 Mahoney KM, Shukla SA, Patsoukis N*, et al.* A secreted PD-L1 splice variant that covalently dimerizes and mediates immunosuppression. *Cancer Immunol Immunother* 2019;68(3):421-432.
175 Wang SF, Fouquet S, Chapon M*, et al.* Early T cell signalling is reversibly altered in PD-1+ T lymphocytes infiltrating human tumors. *PLoS One* 2011;6(3):e17621.
176 Rekik R, Belhadj Hmida N, Ben Hmid A*, et al.* PD-1 induction through TCR activation is partially regulated by endogenous TGF-β. *Cell Mol Immunol* 2015;12(5):648-649.
177 Austin JW, Lu P, Majumder P*, et al.* STAT3, STAT4, NFATc1, and CTCF regulate PD-1 through multiple novel regulatory regions in murine T cells. *J Immunol* 2014;192(10):4876-4886.
178 Kinter AL, Godbout EJ, McNally JP*, et al.* The common gamma-chain cytokines IL-2, IL-7, IL-15, and IL-21 induce the expression of programmed death-1 and its ligands. *J Immunol* 2008;181(10):6738-6746.
179 Yu X, Gao R, Li Y*, et al.* Regulation of PD-1 in T cells for cancer immunotherapy. *Eur J Pharmacol* 2020;881:173240.
180 Jiang X, Zhou J, Giobbie-Hurder A*, et al.* The activation of MAPK in melanoma cells resistant to BRAF inhibition promotes PD-L1 expression that is reversible by MEK and PI3K inhibition. *Clin Cancer Res* 2013;19(3):598-609.
181 Stutvoet TS, Kol A, de Vries EG*, et al.* MAPK pathway activity plays a key role in PD-L1 expression of lung adenocarcinoma cells. *J Pathol* 2019;249(1):52-64.
182 Ota K, Azuma K, Kawahara A*, et al.* Induction of PD-L1 Expression by the EML4–ALK Oncoprotein and Downstream Signaling Pathways in Non–Small Cell Lung Cancer. *Clinical Cancer Research* 2015;21(17):4014-4021.
183 Chen J, Jiang CC, Jin L*, et al.* Regulation of PD-L1: a novel role of pro-survival signalling in cancer. *Ann Oncol* 2016;27(3):409-416.
184 Okita R, Maeda A, Shimizu K*, et al.* PD-L1 overexpression is partially regulated by EGFR/HER2 signaling and associated with poor prognosis in patients with non-small-cell lung cancer. *Cancer Immunol Immunother* 2017;66(7):865-876.
185 Han Y, Liu D, Li L. PD-1/PD-L1 pathway: current researches in cancer. *Am J Cancer Res* 2020;10(3):727-742.
186 Lisberg A, Cummings A, Goldman JW*, et al.* A Phase II Study of Pembrolizumab in EGFR-Mutant, PD-L1+, Tyrosine Kinase Inhibitor Naive Patients With Advanced NSCLC. *J Thorac Oncol* 2018;13(8):1138-1145.
187 Lee JM, Lee SH, Chang B*, et al.* Positive PD-L1 expression is associated with unfavorable clinical outcome in EGFR-mutated lung adenocarcinomas treated with EGFR-TKIs. *European Respiratory Journal* 2019;54(suppl 63):PA367.
188 Fabian MR, Sonenberg N. The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. *Nat Struct Mol Biol* 2012;19(6):586-593.
189 Chen L, Gibbons DL, Goswami S*, et al.* Metastasis is regulated via microRNA-200/ZEB1 axis control of tumour cell PD-L1 expression and intratumoral immunosuppression. *Nat Commun* 2014;5:5241.
190 Jia L, Xi Q, Wang H*, et al.* miR-142-5p regulates tumor cell PD-L1 expression and enhances anti-tumor immunity. *Biochem Biophys Res Commun* 2017;488(2):425-431.
191 Zhu J, Chen L, Zou L*, et al.* MiR-20b, -21, and -130b inhibit PTEN expression resulting in B7-H1 over-expression in advanced colorectal cancer. *Hum Immunol* 2014;75(4):348-353.
192 Fujita Y, Yagishita S, Hagiwara K*, et al.* The clinical relevance of the miR-197/CKS1B/STAT3-mediated PD-L1 network in chemoresistant non-small-cell lung cancer. *Mol Ther* 2015;23(4):717-727.
193 Gong AY, Zhou R, Hu G*, et al.* MicroRNA-513 regulates B7-H1 translation and is involved in IFN-gamma-induced B7-H1 expression in cholangiocytes. *J Immunol* 2009;182(3):1325-1333.

194 Yee D, Shah KM, Coles MC*, et al.* MicroRNA-155 induction via TNF-alpha and IFN-gamma suppresses expression of programmed death ligand-1 (PD-L1) in human primary cells. *J Biol Chem* 2017;292(50):20683-20693.
195 Mimura K, Teh JL, Okayama H*, et al.* PD-L1 expression is mainly regulated by interferon gamma associated with JAK-STAT pathway in gastric cancer. *Cancer Sci* 2018;109(1):43-53.
196 Liu J, Hamrouni A, Wolowiec D*, et al.* Plasma cells from multiple myeloma patients express B7-H1 (PD-L1) and increase expression after stimulation with IFN-{gamma} and TLR ligands via a MyD88-, TRAF6-, and MEK-dependent pathway. *Blood* 2007;110(1):296-304.
197 Alshaker HA, Matalka KZ. IFN-gamma, IL-17 and TGF-beta involvement in shaping the tumor microenvironment: The significance of modulating such cytokines in treating malignant solid tumors. *Cancer Cell Int* 2011;11:33.
198 Meyskens FL, Jr., Kopecky KJ, Taylor CW*, et al.* Randomized trial of adjuvant human interferon gamma versus observation in high-risk cutaneous melanoma: a Southwest Oncology Group study. *J Natl Cancer Inst* 1995;87(22):1710-1713.
199 Alberts DS, Marth C, Alvarez RD*, et al.* Randomized phase 3 trial of interferon gamma-1b plus standard carboplatin/paclitaxel versus carboplatin/paclitaxel alone for first-line treatment of advanced ovarian and primary peritoneal carcinomas: results from a prospectively designed analysis of progression-free survival. *Gynecol Oncol* 2008;109(2):174-181.
200 Prior C, Oroszy S, Oberaigner W*, et al.* Advanced non-small-cell lung cancer: adjunctive interferon gamma in induction and maintenance therapy. *J Cancer Res Clin Oncol* 1999;125(1):42-46.
201 Mattson K, Niiranen A, Pyrhonen S*, et al.* Recombinant interferon gamma treatment in non-small cell lung cancer. Antitumour effect and cardiotoxicity. *Acta Oncol* 1991;30(5):607-610.
202 Shang GS, Liu L, Qin YW. IL-6 and TNF-alpha promote metastasis of lung cancer by inducing epithelial-mesenchymal transition. *Oncol Lett* 2017;13(6):4657-4660.
203 Scott KA, Moore RJ, Arnott CH*, et al.* An Anti-Tumor Necrosis Factor-α Antibody Inhibits the Development of Experimental Skin Tumors. *Molecular Cancer Therapeutics* 2003;2(5):445-451.
204 Hartley G, Regan D, Guth A*, et al.* Regulation of PD-L1 expression on murine tumor-associated monocytes and macrophages by locally produced TNF-alpha. *Cancer Immunol Immunother* 2017;66(4):523-535.
205 Akdis M, Aab A, Altunbulakli C*, et al.* Interleukins (from IL-1 to IL-38), interferons, transforming growth factor beta, and TNF-alpha: Receptors, functions, and roles in diseases. *J Allergy Clin Immunol* 2016;138(4):984-1010.
206 Hirahara K, Ghoreschi K, Yang XP*, et al.* Interleukin-27 priming of T cells controls IL-17 production in trans via induction of the ligand PD-L1. *Immunity* 2012;36(6):1017-1030.
207 Carbotti G, Barisione G, Airoldi I*, et al.* IL-27 induces the expression of IDO and PD-L1 in human cancer cells. *Oncotarget* 2015;6(41):43267-43280.
208 Wang X, Yang L, Huang F*, et al.* Inflammatory cytokines IL-17 and TNF-α up-regulate PD-L1 expression in human prostate and colon cancer cells. *Immunol Lett* 2017;184:7-14.
209 Xiong HY, Ma TT, Wu BT*, et al.* IL-12 regulates B7-H1 expression in ovarian cancer-associated macrophages by effects on NF-κB signalling. *Asian Pac J Cancer Prev* 2014;15(14):5767-5772.
210 Zhang N, Zeng Y, Du W*, et al.* The EGFR pathway is involved in the regulation of PD-L1 expression via the IL-6/JAK/STAT3 signaling pathway in EGFR-mutated non-small cell lung cancer. *Int J Oncol* 2016;49(4):1360-1368.
211 Krzywinska E, Stockmann C. Hypoxia, Metabolism and Immune Cell Function. *Biomedicines* 2018;6(2).
212 Barsoum IB, Smallwood CA, Siemens DR*, et al.* A mechanism of hypoxia-mediated escape from adaptive immunity in cancer cells. *Cancer Res* 2014;74(3):665-674.
213 Socinski MA, Jotte RM, Cappuzzo F*, et al.* Atezolizumab for First-Line Treatment of Metastatic Nonsquamous NSCLC. *N Engl J Med* 2018;378(24):2288-2301.

214 Zarour HM. Reversing T-cell Dysfunction and Exhaustion in Cancer. *Clin Cancer Res* 2016;22(8):1856-1864.
215 Schwarz F, Aebi M. Mechanisms and principles of N-linked protein glycosylation. *Curr Opin Struct Biol* 2011;21(5):576-582.
216 Fathman CG, Lineberry NB. Molecular mechanisms of CD4+ T-cell anergy. *Nat Rev Immunol* 2007;7(8):599-609.
217 Staveley-O'Carroll K, Sotomayor E, Montgomery J*, et al.* Induction of antigen-specific T cell anergy: An early event in the course of tumor progression. *Proc Natl Acad Sci U S A* 1998;95(3):1178-1183.
218 Abe BT, Macian F. Uncovering the mechanisms that regulate tumor-induced T-cell anergy. *Oncoimmunology* 2013;2(2):e22679-e22679.
219 Zha Y, Marks R, Ho AW*, et al.* T cell anergy is reversed by active Ras and is regulated by diacylglycerol kinase-alpha. *Nat Immunol* 2006;7(11):1166-1173.
220 Beverly B, Kang S-M, Lenardo MJ*, et al.* Reversal of in vitro T cell clonal anergy by IL-2 stimulation. *International Immunology* 1992;4(6):661-671.
221 Zheng Y, Zha Y, Gajewski TF. Molecular regulation of T-cell anergy. *EMBO Rep* 2008;9(1):50-55.
222 Brooks DG, Ha SJ, Elsaesser H*, et al.* IL-10 and PD-L1 operate through distinct pathways to suppress T-cell activity during persistent viral infection. *Proc Natl Acad Sci U S A* 2008;105(51):20428-20433.
223 Wherry EJ, Ha S-J, Kaech SM*, et al.* Molecular Signature of CD8⁺ T Cell Exhaustion during Chronic Viral Infection. *Immunity* 2007;27(4):670-684.
224 Sharma P, Hu-Lieskovan S, Wargo JA*, et al.* Primary, Adaptive, and Acquired Resistance to Cancer Immunotherapy. *Cell* 2017;168(4):707-723.
225 Blackburn SD, Shin H, Haining WN*, et al.* Coregulation of CD8+ T cell exhaustion by multiple inhibitory receptors during chronic viral infection. *Nat Immunol* 2009;10(1):29-37.
226 Jin HT, Anderson AC, Tan WG*, et al.* Cooperation of Tim-3 and PD-1 in CD8 T-cell exhaustion during chronic viral infection. *Proc Natl Acad Sci U S A* 2010;107(33):14733-14738.
227 Kamphorst AO, Wieland A, Nasti T*, et al.* Rescue of exhausted CD8 T cells by PD-1-targeted therapies is CD28-dependent. *Science* 2017;355(6332):1423-1427.
228 Hui E, Cheung J, Zhu J*, et al.* T cell costimulatory receptor CD28 is a primary target for PD-1-mediated inhibition. *Science* 2017;355(6332):1428-1433.
229 Chauvin JM, Pagliano O, Fourcade J*, et al.* TIGIT and PD-1 impair tumor antigen-specific CD8(+) T cells in melanoma patients. *J Clin Invest* 2015;125(5):2046-2058.
230 Zahavi DJ, Weiner LM. Targeting Multiple Receptors to Increase Checkpoint Blockade Efficacy. *International journal of molecular sciences* 2019;20(1):158.
231 Curran MA, Montalvo W, Yagita H*, et al.* PD-1 and CTLA-4 combination blockade expands infiltrating T cells and reduces regulatory T and myeloid cells within B16 melanoma tumors. *Proc Natl Acad Sci U S A* 2010;107(9):4275-4280.
232 Büttner R, Longshore JW, López-Ríos F*, et al.* Implementing TMB measurement in clinical practice: considerations on assay requirements. *ESMO Open* 2019;4(1):e000442.
233 Jiang T, Shi J, Dong Z*, et al.* Genomic landscape and its correlations with tumor mutational burden, PD-L1 expression, and immune cells infiltration in Chinese lung squamous cell carcinoma. *Journal of hematology & oncology* 2019;12(1):75-75.
234 Rizvi NA, Hellmann MD, Snyder A*, et al.* Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science (New York, NY)* 2015;348(6230):124-128.
235 Peters S, Gettinger S, Johnson ML*, et al.* Phase II Trial of Atezolizumab As First-Line or Subsequent Therapy for Patients With Programmed Death-Ligand 1-Selected Advanced Non-Small-Cell Lung Cancer (BIRCH). *J Clin Oncol* 2017;35(24):2781-2789.

236 Herbst RS, Lopes G, Kowalski DM*, et al.* Association between tissue TMB (tTMB) and clinical outcomes with pembrolizumab monotherapy (pembro) in PD-L1-positive advanced NSCLC in the KEYNOTE-010 and -042 trials. *Annals of Oncology* 2019;30:v916-v917.
237 Paz-Ares L, Langer CJ, Novello S*, et al.* Pembrolizumab (pembro) plus platinum-based chemotherapy (chemo) for metastatic NSCLC: Tissue TMB (tTMB) and outcomes in KEYNOTE-021, 189, and 407. *Annals of Oncology* 2019;30:v917-v918.
238 Waldman AD, Fritz JM, Lenardo MJ. A guide to cancer immunotherapy: from T cell basic science to clinical practice. *Nature Reviews Immunology* 2020.
239 Peng D, Kryczek I, Nagarsheth N*, et al.* Epigenetic silencing of TH1-type chemokines shapes tumour immunity and immunotherapy. *Nature* 2015;527(7577):249-253.
240 Peng W, Chen JQ, Liu C*, et al.* Loss of PTEN Promotes Resistance to T Cell–Mediated Immunotherapy. *Cancer Discovery* 2016;6(2):202-216.
241 George S, Miao D, Demetri GD*, et al.* Loss of PTEN Is Associated with Resistance to Anti-PD-1 Checkpoint Blockade Therapy in Metastatic Uterine Leiomyosarcoma. *Immunity* 2017;46(2):197-204.
242 Nirschl CJ, Drake CG. Molecular pathways: coexpression of immune checkpoint molecules: signaling pathways and implications for cancer immunotherapy. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2013;19(18):4917-4924.
243 Antonia S, Goldberg SB, Balmanoukian A*, et al.* Safety and antitumour activity of durvalumab plus tremelimumab in non-small cell lung cancer: a multicentre, phase 1b study. *Lancet Oncol* 2016;17(3):299-308.
244 Koyama S, Akbay EA, Li YY*, et al.* Adaptive resistance to therapeutic PD-1 blockade is associated with upregulation of alternative immune checkpoints. *Nature Communications* 2016;7(1):10501.
245 Zaretsky JM, Garcia-Diaz A, Shin DS*, et al.* Mutations Associated with Acquired Resistance to PD-1 Blockade in Melanoma. *New England Journal of Medicine* 2016;375(9):819-829.
246 Pauken KE, Sammons MA, Odorizzi PM*, et al.* Epigenetic stability of exhausted T cells limits durability of reinvigoration by PD-1 blockade. *Science* 2016;354(6316):1160-1165.
247 Matos LLd, Trufelli DC, de Matos MGL*, et al.* Immunohistochemistry as an important tool in biomarkers detection and clinical practice. *Biomark Insights* 2010;5:9-20.
248 Haines DM, West KH. Immunohistochemistry: forging the links between immunology and pathology. *Vet Immunol Immunopathol* 2005;108(1-2):151-156.
249 Hao J, Crawford W, Murillo A*, et al.* Development of an IHC Detection System Comparable to OptiView DAB IHC Detection for Rat Primary Antibodies. *American Journal of Clinical Pathology* 2012;138(suppl_2):A101-A101.
250 Roche V. VENTANA PD-L1 (SP142) Assay. Vol. 2019, 2016.
251 Roche V. VENTANA PD-L1 (SP263) Assay. Vol. 2019, 2017.
252 Dako A. PD-L1 IHC 22C3 pharmDx. Vol. 2019, 2015.
253 Dako A. PD-L1 IHC 28-8 pharmDx. Vol. 2019, 2015.
254 Garon EB, Rizvi NA, Hui R*, et al.* Pembrolizumab for the treatment of non-small-cell lung cancer. *N Engl J Med* 2015;372(21):2018-2028.
255 Herbst RS, Baas P, Kim DW*, et al.* Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. *Lancet* 2016;387(10027):1540-1550.
256 Gandhi L, Rodríguez-Abreu D, Gadgeel S*, et al.* Pembrolizumab plus Chemotherapy in Metastatic Non–Small-Cell Lung Cancer. *New England Journal of Medicine* 2018;378(22):2078-2092.
257 Paz-Ares L, Luft A, Vicente D*, et al.* Pembrolizumab plus Chemotherapy for Squamous Non–Small-Cell Lung Cancer. *New England Journal of Medicine* 2018;379(21):2040-2051.
258 Hellmann MD, Paz-Ares L, Bernabe Caro R*, et al.* Nivolumab plus Ipilimumab in Advanced Non–Small-Cell Lung Cancer. *New England Journal of Medicine* 2019;381(21):2020-2031.

259 West H, McCleod M, Hussein M*, et al.* Atezolizumab in combination with carboplatin plus nab-paclitaxel chemotherapy compared with chemotherapy alone as first-line treatment for metastatic non-squamous non-small-cell lung cancer (IMpower130): a multicentre, randomised, open-label, phase 3 trial. *Lancet Oncol* 2019;20(7):924-937.
260 Fehrenbacher L, von Pawel J, Park K*, et al.* Updated Efficacy Analysis Including Secondary Population Results for OAK: A Randomized Phase III Study of Atezolizumab versus Docetaxel in Patients with Previously Treated Advanced Non-Small Cell Lung Cancer. *J Thorac Oncol* 2018;13(8):1156-1170.
261 Vogel CL, Cobleigh MA, Tripathy D*, et al.* Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. *J Clin Oncol* 2002;20(3):719-726.
262 Gutierrez C, Schiff R. HER2: biology, detection, and clinical implications. *Archives of pathology & laboratory medicine* 2011;135(1):55-62.
263 Brahmer JR, Drake CG, Wollner I*, et al.* Phase I study of single-agent anti-programmed death-1 (MDX-1106) in refractory solid tumors: safety, clinical activity, pharmacodynamics, and immunologic correlates. *J Clin Oncol* 2010;28(19):3167-3175.
264 Gettinger SN, Horn L, Gandhi L*, et al.* Overall Survival and Long-Term Safety of Nivolumab (Anti-Programmed Death 1 Antibody, BMS-936558, ONO-4538) in Patients With Previously Treated Advanced Non-Small-Cell Lung Cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2015;33(18):2004-2012.
265 Horn L, Mansfield AS, Szczęsna A*, et al.* First-Line Atezolizumab plus Chemotherapy in Extensive-Stage Small-Cell Lung Cancer. *New England Journal of Medicine* 2018;379(23):2220-2229.
266 Buttner R, Gosney JR, Skov BG*, et al.* Programmed Death-Ligand 1 Immunohistochemistry Testing: A Review of Analytical Assays and Clinical Implementation in Non-Small-Cell Lung Cancer. *J Clin Oncol* 2017;35(34):3867-3876.
267 Carbone DP, Reck M, Paz-Ares L*, et al.* First-Line Nivolumab in Stage IV or Recurrent Non–Small-Cell Lung Cancer. *New England Journal of Medicine* 2017;376(25):2415-2426.
268 Remon J, Besse B, Soria JC. Successes and failures: what did we learn from recent first-line treatment immunotherapy trials in non-small cell lung cancer? *BMC Med* 2017;15(1):55.
269 Scattone A, Catino A, Schirosi L*, et al.* Discordance between FISH, IHC, and NGS Analysis of ALK Status in Advanced Non-Small Cell Lung Cancer (NSCLC): a Brief Report of 7 Cases. *Transl Oncol* 2019;12(2):389-395.
270 Mok TSK, Wu YL, Kudaba I*, et al.* Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): a randomised, open-label, controlled, phase 3 trial. *Lancet* 2019;393(10183):1819-1830.
271 Thompson RH, Gillett MD, Cheville JC*, et al.* Costimulatory molecule B7-H1 in primary and metastatic clear cell renal cell carcinoma. *Cancer* 2005;104(10):2084-2091.
272 Rebelatto MC, Midha A, Mistry A*, et al.* Development of a programmed cell death ligand-1 immunohistochemical assay validated for analysis of non-small cell lung cancer and head and neck squamous cell carcinoma. *Diagn Pathol* 2016;11(1):95.
273 Wang C, Hahn E, Slodkowska E*, et al.* Reproducibility of PD-L1 immunohistochemistry interpretation across various types of genitourinary and head/neck carcinomas, antibody clones, and tissue types. *Human Pathology* 2018;82:131-139.
274 Jain D, Sukumar S, Mohan A*, et al.* Programmed death-ligand 1 immunoexpression in matched biopsy and liquid-based cytology samples of advanced stage non-small cell lung carcinomas. *Cytopathology* 2018;29(6):550-557.
275 Lloyd IE, Zhou W, Witt BL*, et al.* Characterization of PD-L1 Immunohistochemical Expression in Cell Blocks With Different Specimen Fixation and Processing Methods. *Appl Immunohistochem Mol Morphol* 2019;27(2):107-113.

276 Forest F, Cote G, Laville D*, et al.* Impact of delayed fixation and decalcification on PD-L1 expression: a comparison of two clones. *Virchows Arch* 2019;475(6):693-699.
277 Kawachi H, Fujimoto D, Yamashita D*, et al.* Association Between Formalin Fixation Time and Programmed Cell Death Ligand 1 Expression in Patients With Non-Small Cell Lung Cancer. *Anticancer Res* 2019;39(5):2561-2567.
278 Kai K, Yoda Y, Kawaguchi A*, et al.* Formalin fixation on HER-2 and PD-L1 expression in gastric cancer: A pilot analysis using the same surgical specimens with different fixation times. *World J Clin Cases* 2019;7(4):419-430.
279 Webster JD, Miller MA, Dusold D*, et al.* Effects of prolonged formalin fixation on diagnostic immunohistochemistry in domestic animals. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* 2009;57(8):753-761.
280 Rizvi NA, Mazières J, Planchard D*, et al.* Activity and safety of nivolumab, an anti-PD-1 immune checkpoint inhibitor, for patients with advanced, refractory squamous non-small-cell lung cancer (CheckMate 063): a phase 2, single-arm trial. *The Lancet Oncology* 2015;16(3):257-265.
281 Herbst RS, Baas P, Perez-Gracia JL*, et al.* PD1.06 (also presented as P2.41): Pembrolizumab vs Docetaxel for Previously Treated NSCLC (KEYNOTE-010): Archival vs New Tumor Samples for PD-L1 Assessment. *Journal of Thoracic Oncology* 2016;11(10):S174-S175.
282 Tsao MS, Kerr K, Dacic S*, et al.* IASLC ATLAS OF PD-L1 IMMUNOHISTOCHEMISTRY TESTING IN LUNG CANCER. 1st edn: First IASLC Press Printing, 2017, p. 132.
283 Calles A, Liao X, Sholl LM*, et al.* Expression of PD-1 and Its Ligands, PD-L1 and PD-L2, in Smokers and Never Smokers with KRAS-Mutant Lung Cancer. *J Thorac Oncol* 2015;10(12):1726-1735.
284 Midha A, Sharpe A, Scott M*, et al.* PD-L1 expression in advanced NSCLC: Primary lesions versus metastatic sites and impact of sample age. *Journal of Clinical Oncology* 2016;34(15_suppl):3025-3025.
285 Hata A, Katakami N, Nanjo S*, et al.* Programmed death-ligand 1 expression according to epidermal growth factor receptor mutation status in pretreated non-small cell lung cancer. *Oncotarget* 2017;8(69):113807-113816.
286 Boothman AM, Scott M, Ratcliffe M*, et al.* Impact of Patient Characteristics, Prior Therapy, and Sample Type on Tumor Cell Programmed Cell Death Ligand 1 Expression in Patients with Advanced NSCLC Screened for the ATLANTIC Study. *J Thorac Oncol* 2019;14(8):1390-1399.
287 Giunchi F, Degiovanni A, Daddi N*, et al.* Fading With Time of PD-L1 Immunoreactivity in Non-Small Cells Lung Cancer Tissues: A Methodological Study. *Appl Immunohistochem Mol Morphol* 2018;26(7):489-494.
288 Kim H, Kwon HJ, Park SY*, et al.* PD-L1 immunohistochemical assays for assessment of therapeutic strategies involving immune checkpoint inhibitors in non-small cell lung cancer: a comparative study. *Oncotarget* 2017;8(58):98524-98532.
289 Nakamura Y, Kobayashi T, Nishii Y*, et al.* Comparable immunoreactivity rates of PD-L1 in archival and recent specimens from non-small cell lung cancer. *Thorac Cancer* 2018;9(11):1476-1482.
290 Wang H, Agulnik J, Kasymjanova G*, et al.* Cytology cell blocks are suitable for immunohistochemical testing for PD-L1 in lung cancer. *Ann Oncol* 2018;29(6):1417-1422.
291 Gagné A, Wang E, Bastien N*, et al.* Impact of Specimen Characteristics on PD-L1 Testing in Non-Small Cell Lung Cancer: Validation of the IASLC PD-L1 Testing Recommendations. *J Thorac Oncol* 2019;14(12):2062-2070.
292 Xie R, Chung JY, Ylaya K*, et al.* Factors influencing the degradation of archival formalin-fixed paraffin-embedded tissue sections. *J Histochem Cytochem* 2011;59(4):356-365.

293 Economou M, Schoni L, Hammer C*, et al.* Proper paraffin slide storage is crucial for translational research projects involving immunohistochemistry stains. *Clin Transl Med* 2014;3(1):4.
294 Bertheau P, Cazals-Hatem D, Meignin V*, et al.* Variability of immunohistochemical reactivity on stored paraffin slides. *J Clin Pathol* 1998;51(5):370-374.
295 DiVito KA, Charette LA, Rimm DL*, et al.* Long-term preservation of antigenicity on tissue microarrays. *Lab Invest* 2004;84(8):1071-1078.
296 Sato Y, Fujimoto D, Uehara K*, et al.* Reduced Tumour Proportion Scores for Programmed Cell Death Ligand 1 in Stored Paraffin Tissue Sections. *Anticancer Res* 2018;38(3):1401-1405.
297 Hendry S, Byrne DJ, Wright GM*, et al.* Comparison of Four PD-L1 Immunohistochemical Assays in Lung Cancer. *J Thorac Oncol* 2018;13(3):367-376.
298 Levit LA, Peppercorn JM, Tam AL*, et al.* Ethical Framework for Including Research Biopsies in Oncology Clinical Trials: American Society of Clinical Oncology Research Statement. *Journal of Clinical Oncology* 2019;37(26):2368-2377.
299 Teixido C, Vilarino N, Reyes R*, et al.* PD-L1 expression testing in non-small cell lung cancer. *Ther Adv Med Oncol* 2018;10:1758835918763493.
300 Stoy SP, Rosen L, Mueller J*, et al.* Programmed death-ligand 1 testing of lung cancer cytology specimens obtained with bronchoscopy. *Cancer Cytopathol* 2018;126(2):122-128.
301 Biswas A, Leon ME, Drew P*, et al.* Clinical performance of endobronchial ultrasound-guided transbronchial needle aspiration for assessing programmed death ligand-1 expression in nonsmall cell lung cancer. *Diagn Cytopathol* 2018;46(5):378-383.
302 Arriola AGP, Bashover E, Joseph C*, et al.* The usefulness of various cytologic specimen preparations for PD-L1 immunostaining in non-small cell lung carcinoma. *J Am Soc Cytopathol* 2018;7(6):324-332.
303 Noll B, Wang WL, Gong Y*, et al.* Programmed death ligand 1 testing in non-small cell lung carcinoma cytology cell block and aspirate smear preparations. *Cancer Cytopathol* 2018;126(5):342-352.
304 Tsao MS, Kerr KM, Kockx M*, et al.* PD-L1 Immunohistochemistry Comparability Study in Real-Life Clinical Samples: Results of Blueprint Phase 2 Project. *J Thorac Oncol* 2018;13(9):1302-1311.
305 Villaruz LC, Ancevski Hunter K, Kurland BF*, et al.* Comparison of PD-L1 immunohistochemistry assays and response to PD-1/L1 inhibitors in advanced non-small-cell lung cancer in clinical practice. *Histopathology* 2019;74(2):269-275.
306 Kim I, Kim A, Lee CH*, et al.* Reliability of PD-L1 assays using small tissue samples compared with surgical specimens. *Medicine (Baltimore)* 2019;98(14):e14972.
307 Skov BG, Skov T. Paired Comparison of PD-L1 Expression on Cytologic and Histologic Specimens From Malignancies in the Lung Assessed With PD-L1 IHC 28-8pharmDx and PD-L1 IHC 22C3pharmDx. *Appl Immunohistochem Mol Morphol* 2017;25(7):453-459.
308 Heymann JJ, Bulman WA, Swinarski D*, et al.* PD-L1 expression in non-small cell lung carcinoma: Comparison among cytology, small biopsy, and surgical resection specimens. *Cancer Cytopathol* 2017;125(12):896-907.
309 Sakakibara R, Inamura K, Tambo Y*, et al.* EBUS-TBNA as a Promising Method for the Evaluation of Tumor PD-L1 Expression in Lung Cancer. *Clin Lung Cancer* 2017;18(5):527-534.e521.
310 Ilie M, Juco J, Huang L*, et al.* Use of the 22C3 anti-programmed death-ligand 1 antibody to determine programmed death-ligand 1 expression in cytology samples obtained from non-small cell lung cancer patients. *Cancer Cytopathol* 2018;126(4):264-274.
311 Russell-Goldman E, Kravets S, Dahlberg SE*, et al.* Cytologic-histologic correlation of programmed death-ligand 1 immunohistochemistry in lung carcinomas. *Cancer Cytopathol* 2018;126(4):253-263.

312 Hernandez A, Brandler TC, Zhou F*, et al.* 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.
313 Grosu HB, Arriola A, Stewart J*, et al.* PD-L1 detection in histology specimens and matched pleural fluid cell blocks of patients with NSCLC. *Respirology* 2019;24(12):1198-1203.
314 Sakata KK, Midthun DE, Mullon JJ*, et al.* Comparison of Programmed Death Ligand-1 Immunohistochemical Staining Between Endobronchial Ultrasound Transbronchial Needle Aspiration and Resected Lung Cancer Specimens. *Chest* 2018;154(4):827-837.
315 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.
316 Capizzi E, Ricci C, Giunchi F*, et al.* Validation of the immunohistochemical expression of programmed death ligand 1 (PD-L1) on cytological smears in advanced non small cell lung cancer. *Lung Cancer* 2018;126:9-14.
317 Lozano MD, Abengozar-Muela M, Echeveste JI*, et al.* Programmed death-ligand 1 expression on direct Pap-stained cytology smears from non-small cell lung cancer: Comparison with cell blocks and surgical resection specimens. *Cancer Cytopathol* 2019;127(7):470-480.
318 Evans M, O'Sullivan B, Hughes F*, et al.* The Clinicopathological and Molecular Associations of PD-L1 Expression in Non-small Cell Lung Cancer: Analysis of a Series of 10,005 Cases Tested with the 22C3 Assay. *Pathol Oncol Res* 2020;26(1):79-89.
319 Torous VF, Rangachari D, Gallant BP*, et al.* PD-L1 testing using the clone 22C3 pharmDx kit for selection of patients with non-small cell lung cancer to receive immune checkpoint inhibitor therapy: are cytology cell blocks a viable option? *J Am Soc Cytopathol* 2018;7(3):133-141.
320 Mei P, Shilo K, Wei L*, et al.* Programmed cell death ligand 1 expression in cytologic and surgical non-small cell lung carcinoma specimens from a single institution: Association with clinicopathologic features and molecular alterations. *Cancer Cytopathol* 2019;127(7):447-457.
321 Chavali P, Prayaga AK, Tandon A*, et al.* Utility of immunochemistry in cytology. *Journal of cytology* 2016;33(2):71-75.
322 Pareja F, Murray MP, Jean RD*, et al.* Cytologic assessment of estrogen receptor, progesterone receptor, and HER2 status in metastatic breast carcinoma. *Journal of the American Society of Cytopathology* 2017;6(1):33-40.
323 Lantuejoul S, Sound-Tsao M, Cooper WA*, et al.* PD-L1 Testing for Lung Cancer in 2019: Perspective From the IASLC Pathology Committee. *J Thorac Oncol* 2020;15(4):499-519.
324 Strickland AL, Blacketer S, Molberg K*, et al.* Effects of Decalcifying Agents of Variable Duration on PD-L1 Immunohistochemistry. *Am J Clin Pathol* 2020;153(2):258-265.
325 Pontarollo G, Confavreux CB, Pialat J-B*, et al.* Bone decalcification to assess programmed cell death ligand 1 expression in bone metastases of non-small cell lung cancers. *J Bone Oncol* 2020;21:100275-100275.
326 Schats K, Vre EAV, Schrijvers D*, et al.* Epitope mapping of PD-L1 primary antibodies (28-8, SP142, SP263, E1L3N). *Journal of Clinical Oncology* 2017;35(15_suppl):3028-3028.
327 Lawson NL, Dix CI, Scorer PW*, et al.* Mapping the binding sites of antibodies utilized in programmed cell death ligand-1 predictive immunohistochemical assays for use with immuno-oncology therapies. *Mod Pathol* 2019.
328 Adam J, Le Stang N, Rouquette I*, et al.* Multicenter harmonization study for PD-L1 IHC testing in non-small-cell lung cancer. *Ann Oncol* 2018;29(4):953-958.
329 Torlakovic E, Albadine R, Bigras G*, et al.* Canadian Multicenter Project on Standardization of Programmed Death-Ligand 1 Immunohistochemistry 22C3 Laboratory-Developed Tests for Pembrolizumab Therapy in NSCLC. *Journal of Thoracic Oncology*.

330 Scheel AH, Dietel M, Heukamp LC*, et al.* Harmonized PD-L1 immunohistochemistry for pulmonary squamous-cell and adenocarcinomas. *Mod Pathol* 2016;29(10):1165-1172.
331 Neuman T, London M, Kania-Almog J*, et al.* A Harmonization Study for the Use of 22C3 PD-L1 Immunohistochemical Staining on Ventana's Platform. *J Thorac Oncol* 2016;11(11):1863-1868.
332 Cogswell J, Inzunza HD, Wu Q*, et al.* An Analytical Comparison of Dako 28-8 PharmDx Assay and an E1L3N Laboratory-Developed Test in the Immunohistochemical Detection of Programmed Death-Ligand 1. *Mol Diagn Ther* 2017;21(1):85-93.
333 Smith J, Robida MD, Acosta K*, et al.* Quantitative and qualitative characterization of Two PD-L1 clones: SP263 and E1L3N. *Diagn Pathol* 2016;11(1):44.
334 Marchetti A, Barberis M, Franco R*, et al.* Multicenter Comparison of 22C3 PharmDx (Agilent) and SP263 (Ventana) Assays to Test PD-L1 Expression for NSCLC Patients to Be Treated with Immune Checkpoint Inhibitors. *Journal of Thoracic Oncology* 2017;12(11):1654-1663.
335 Brunnstrom H, Johansson A, Westbom-Fremer S*, et al.* PD-L1 immunohistochemistry in clinical diagnostics of lung cancer: inter-pathologist variability is higher than assay variability. *Mod Pathol* 2017;30(10):1411-1421.
336 Soo RA, Yun Lim JS, Asuncion BR*, et al.* Determinants of variability of five programmed death ligand-1 immunohistochemistry assays in non-small cell lung cancer samples. *Oncotarget* 2018;9(6):6841-6851.
337 Gaule P, Smithy JW, Toki M*, et al.* A Quantitative Comparison of Antibodies to Programmed Cell Death 1 Ligand 1. *JAMA Oncol* 2017;3(2):256-259.
338 Aperio. Positive Pixel Count Algorithm User's Guide Vol. 2019, 2009.
339 Parra ER, Uraoka N, Jiang M*, et al.* Validation of multiplex immunofluorescence panels using multispectral microscopy for immune-profiling of formalin-fixed and paraffin-embedded human tumor tissues. *Sci Rep* 2017;7(1):13380.
340 Fankhauser CD, Schüffler PJ, Gillessen S*, et al.* Comprehensive immunohistochemical analysis of PD-L1 shows scarce expression in castration-resistant prostate cancer. *Oncotarget* 2017;9(12):10284-10293.
341 Schats KA, Van Vré EA, De Schepper S*, et al.* Validated programmed cell death ligand 1 immunohistochemistry assays (E1L3N and SP142) reveal similar immune cell staining patterns in melanoma when using the same sensitive detection system. *Histopathology* 2017;70(2):253-263.
342 Rimm DL, Han G, Taube JM*, et al.* A Prospective, Multi-institutional, Pathologist-Based Assessment of 4 Immunohistochemistry Assays for PD-L1 Expression in Non-Small Cell Lung Cancer. *JAMA Oncol* 2017;3(8):1051-1058.
343 Ratcliffe MJ, Sharpe A, Midha A*, et al.* Agreement between Programmed Cell Death Ligand-1 Diagnostic Assays across Multiple Protein Expression Cutoffs in Non-Small Cell Lung Cancer. *Clin Cancer Res* 2017;23(14):3585-3591.
344 Batenchuk C, Albitar M, Zerba K*, et al.* A real-world, comparative study of FDA-approved diagnostic assays PD-L1 IHC 28-8 and 22C3 in lung cancer and other malignancies. *J Clin Pathol* 2018;71(12):1078-1083.
345 Velcheti V, Patwardhan PD, Liu FX*, et al.* Real-world PD-L1 testing and distribution of PD-L1 tumor expression by immunohistochemistry assay type among patients with metastatic non-small cell lung cancer in the United States. *PLoS One* 2018;13(11):e0206370.
346 Munari E, Zamboni G, Lunardi G*, et al.* PD-L1 expression comparison between primary and relapsed non-small cell lung carcinoma using whole sections and clone SP263. *Oncotarget* 2018;9(54):30465-30471.
347 Munari E, Zamboni G, Lunardi G*, et al.* PD-L1 expression in non-small cell lung cancer: evaluation of the diagnostic accuracy of a laboratory-developed test using clone E1L3N in comparison with 22C3 and SP263 assays. *Hum Pathol* 2019;90:54-59.

348 Tseng JS, Yang TY, Wu CY*, et al.* Characteristics and Predictive Value of PD-L1 Status in Real-World Non-Small Cell Lung Cancer Patients. *J Immunother* 2018;41(6):292-299.
349 Fujimoto D, Uehara K, Sato Y*, et al.* Alteration of PD-L1 expression and its prognostic impact after concurrent chemoradiation therapy in non-small cell lung cancer patients. *Scientific Reports* 2017;7(1):11373.
350 Krawczyk P, Jarosz B, Kucharczyk T*, et al.* Immunohistochemical assays incorporating SP142 and 22C3 monoclonal antibodies for detection of PD-L1 expression in NSCLC patients with known status of EGFR and ALK genes. *Oncotarget* 2017;8(38):64283-64293.
351 Xu H, Lin G, Huang C*, et al.* Assessment of Concordance between 22C3 and SP142 Immunohistochemistry Assays regarding PD-L1 Expression in Non-Small Cell Lung Cancer. *Scientific Reports* 2017;7(1):16956.
352 Pang C, Yin L, Zhou X*, et al.* Assessment of programmed cell death ligand-1 expression with multiple immunohistochemistry antibody clones in non-small cell lung cancer. *J Thorac Dis* 2018;10(2):816-824.
353 Scheel AH, Baenfer G, Baretton G*, et al.* Interlaboratory concordance of PD-L1 immunohistochemistry for non-small-cell lung cancer. *Histopathology* 2018;72(3):449-459.
354 Beck KS, Kim SJ, Kang JH*, et al.* CT-guided transthoracic needle biopsy for evaluation of PD-L1 expression: Comparison of 22C3 and SP263 assays. *Thorac Cancer* 2019;10(7):1612-1618.
355 Sheffield BS, Fulton R, Kalloger SE*, et al.* Investigation of PD-L1 Biomarker Testing Methods for PD-1 Axis Inhibition in Non-squamous Non-small Cell Lung Cancer. *J Histochem Cytochem* 2016;64(10):587-600.
356 Li C-W, Lim S-O, Xia W*, et al.* Glycosylation and stabilization of programmed death ligand-1 suppresses T-cell activity. *Nature Communications* 2016;7(1):12632.
357 Horita H, Law A, Hong S*, et al.* Identifying Regulatory Posttranslational Modifications of PD-L1: A Focus on Monoubiquitinaton. *Neoplasia* 2017;19(4):346-353.
358 Lim S-O, Li C-W, Xia W*, et al.* Deubiquitination and Stabilization of PD-L1 by CSN5. *Cancer cell* 2016;30(6):925-939.
359 Li CW, Lim SO, Chung EM*, et al.* Eradication of Triple-Negative Breast Cancer Cells by Targeting Glycosylated PD-L1. *Cancer Cell* 2018;33(2):187-201.e110.
360 Morales-Betanzos CA, Lee H, Gonzalez Ericsson PI*, et al.* Quantitative Mass Spectrometry Analysis of PD-L1 Protein Expression, N-glycosylation and Expression Stoichiometry with PD-1 and PD-L2 in Human Melanoma. *Mol Cell Proteomics* 2017;16(10):1705-1717.
361 López-Urrutia E, Campos-Parra A, Herrera LA*, et al.* Alternative splicing regulation in tumor necrosis factor-mediated inflammation. *Oncology letters* 2017;14(5):5114-5120.
362 Oaks MK, Hallett KM. Cutting Edge: A Soluble Form of CTLA-4 in Patients with Autoimmune Thyroid Disease. *The Journal of Immunology* 2000;164(10):5015-5018.
363 Takahashi N, Iwasa S, Sasaki Y*, et al.* Serum levels of soluble programmed cell death ligand 1 as a prognostic factor on the first-line treatment of metastatic or recurrent gastric cancer. *Journal of Cancer Research and Clinical Oncology* 2016;142(8):1727-1738.
364 Frigola X, Inman BA, Lohse CM*, et al.* Identification of a soluble form of B7-H1 that retains immunosuppressive activity and is associated with aggressive renal cell carcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2011;17(7):1915-1923.
365 Zhu X, Lang J. Soluble PD-1 and PD-L1: predictive and prognostic significance in cancer. *Oncotarget* 2017;8(57):97671-97682.
366 Takeuchi M, Doi T, Obayashi K*, et al.* Soluble PD-L1 with PD-1-binding capacity exists in the plasma of patients with non-small cell lung cancer. *Immunol Lett* 2018;196:155-160.
367 Tominaga T, Akiyoshi T, Yamamoto N*, et al.* Clinical significance of soluble programmed cell death-1 and soluble programmed cell death-ligand 1 in patients with locally advanced rectal cancer treated with neoadjuvant chemoradiotherapy. *PloS one* 2019;14(2):e0212978-e0212978.

368 Okuma Y, Hosomi Y, Nakahara Y*, et al.* High plasma levels of soluble programmed cell death ligand 1 are prognostic for reduced survival in advanced lung cancer. *Lung Cancer* 2017;104:1-6.
369 Finkelmeier F, Canli Ö, Tal A*, et al.* High levels of the soluble programmed death-ligand (sPD-L1) identify hepatocellular carcinoma patients with a poor prognosis. *Eur J Cancer* 2016;59:152-159.
370 Costantini A, Julie C, Dumenil C*, et al.* Predictive role of plasmatic biomarkers in advanced non-small cell lung cancer treated by nivolumab. *Oncoimmunology* 2018;7(8):e1452581.
371 Ruf M, Moch H, Schraml P. PD-L1 expression is regulated by hypoxia inducible factor in clear cell renal cell carcinoma. *Int J Cancer* 2016;139(2):396-403.
372 Shigemori T, Toiyama Y, Okugawa Y*, et al.* Soluble PD-L1 Expression in Circulation as a Predictive Marker for Recurrence and Prognosis in Gastric Cancer: Direct Comparison of the Clinical Burden Between Tissue and Serum PD-L1 Expression. *Ann Surg Oncol* 2019;26(3):876-883.
373 Koukourakis MI, Kontomanolis E, Sotiropoulou M*, et al.* Increased Soluble PD-L1 Levels in the Plasma of Patients with Epithelial Ovarian Cancer Correlate with Plasma Levels of miR34a and miR200. *Anticancer Res* 2018;38(10):5739-5745.
374 Chang B, Huang T, Wei H*, et al.* The correlation and prognostic value of serum levels of soluble programmed death protein 1 (sPD-1) and soluble programmed death-ligand 1 (sPD-L1) in patients with hepatocellular carcinoma. *Cancer Immunol Immunother* 2019;68(3):353-363.
375 Konishi J, Yamazaki K, Azuma M*, et al.* B7-H1 expression on non-small cell lung cancer cells and its relationship with tumor-infiltrating lymphocytes and their PD-1 expression. *Clin Cancer Res* 2004;10(15):5094-5100.
376 Velcheti V, Schalper KA, Carvajal DE*, et al.* Programmed death ligand-1 expression in non-small cell lung cancer. *Lab Invest* 2014;94(1):107-116.
377 McCabe A, Dolled-Filhart M, Camp RL*, et al.* Automated quantitative analysis (AQUA) of in situ protein expression, antibody concentration, and prognosis. *J Natl Cancer Inst* 2005;97(24):1808-1815.
378 McLaughlin J, Han G, Schalper KA*, et al.* Quantitative Assessment of the Heterogeneity of PD-L1 Expression in Non-Small-Cell Lung Cancer. *JAMA Oncol* 2016;2(1):46-54.
379 Rehman JA, Han G, Carvajal-Hausdorf DE*, et al.* Quantitative and pathologist-read comparison of the heterogeneity of programmed death-ligand 1 (PD-L1) expression in non-small cell lung cancer. *Mod Pathol* 2017;30(3):340-349.
380 Casadevall D, Clave S, Taus A*, et al.* Heterogeneity of Tumor and Immune Cell PD-L1 Expression and Lymphocyte Counts in Surgical NSCLC Samples. *Clin Lung Cancer* 2017;18(6):682-691 e685.
381 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. *Journal of Thoracic Oncology* 2017;12(10):1536-1543.
382 Gniadek TJ, Li QK, Tully E*, et al.* Heterogeneous expression of PD-L1 in pulmonary squamous cell carcinoma and adenocarcinoma: implications for assessment by small biopsy. *Mod Pathol* 2017;30(4):530-538.
383 Fukuoka J, Hofer MD, Hori T*, et al.* Spiral array: a new high-throughput technology covers tissue heterogeneity. *Arch Pathol Lab Med* 2012;136(11):1377-1384.
384 Nakamura S, Hayashi K, Imaoka Y*, et al.* Intratumoral heterogeneity of programmed cell death ligand-1 expression is common in lung cancer. *PLoS One* 2017;12(10):e0186192.
385 Koezuka S, Mikami T, Tochigi N*, et al.* Toward improving prognosis prediction in patients undergoing small lung adenocarcinoma resection: Radiological and pathological assessment of diversity and intratumor heterogeneity. *Lung Cancer* 2019;135:40-46.

386 Munari E, Zamboni G, Marconi M*, et al.* PD-L1 expression heterogeneity in non-small cell lung cancer: evaluation of small biopsies reliability. *Oncotarget* 2017;8(52):90123-90131.
387 Munari E, Zamboni G, Lunardi G*, et al.* PD-L1 Expression Heterogeneity in Non-Small Cell Lung Cancer: Defining Criteria for Harmonization between Biopsy Specimens and Whole Sections. *J Thorac Oncol* 2018;13(8):1113-1120.
388 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.
389 Ilie M, Hofman V, Dietel M*, et al.* Assessment of the PD-L1 status by immunohistochemistry: challenges and perspectives for therapeutic strategies in lung cancer patients. *Virchows Arch* 2016;468(5):511-525.
390 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.
391 Tsunoda A, Morikawa K, Inoue T*, et al.* A prospective observational study to assess PD-L1 expression in small biopsy samples for non-small-cell lung cancer. *BMC Cancer* 2019;19(1):546.
392 Elfving H, Mattsson JSM, Lindskog C*, et al.* Programmed Cell Death Ligand 1 Immunohistochemistry: A Concordance Study Between Surgical Specimen, Biopsy, and Tissue Microarray. *Clin Lung Cancer* 2019;20(4):258-262.e251.
393 Scorer P, Scott M, Lawson N*, et al.* Consistency of tumor and immune cell programmed cell death ligand-1 expression within and between tumor blocks using the VENTANA SP263 assay. *Diagn Pathol* 2018;13(1):47.
394 Pinato DJ, Shiner RJ, White SD*, et al.* Intra-tumoral heterogeneity in the expression of programmed-death (PD) ligands in isogeneic primary and metastatic lung cancer: Implications for immunotherapy. *Oncoimmunology* 2016;5(9):e1213934.
395 control Ufic. *TNM Classification of Malignant Tumours, 8th Edition*: Wiley-Blackwell, 2016.
396 Kim S, Koh J, Kwon D*, et al.* Comparative analysis of PD-L1 expression between primary and metastatic pulmonary adenocarcinomas. *Eur J Cancer* 2017;75:141-149.
397 Kim HR, Cha YJ, Hong MH*, et al.* Concordance of programmed death-ligand 1 expression between primary and metastatic non-small cell lung cancer by immunohistochemistry and RNA in situ hybridization. *Oncotarget* 2017;8(50):87234-87243.
398 Medical B. RISH. Vol. 2020, 2019.
399 Mansfield AS, Murphy SJ, Peikert T*, et al.* Heterogeneity of Programmed Cell Death Ligand 1 Expression in Multifocal Lung Cancer. *Clin Cancer Res* 2016;22(9):2177-2182.
400 Mansfield AS, Aubry MC, Moser JC*, et al.* Temporal and spatial discordance of programmed cell death-ligand 1 expression and lymphocyte tumor infiltration between paired primary lesions and brain metastases in lung cancer. *Ann Oncol* 2016;27(10):1953-1958.
401 Zhou J, Gong Z, Jia Q*, et al.* Programmed death ligand 1 expression and CD8(+) tumor-infiltrating lymphocyte density differences between paired primary and brain metastatic lesions in non-small cell lung cancer. *Biochem Biophys Res Commun* 2018;498(4):751-757.
402 Huang J, Xu B, Mo H*, et al.* Safety, Activity, and Biomarkers of SHR-1210, an Anti-PD-1 Antibody, for Patients with Advanced Esophageal Carcinoma. *Clin Cancer Res* 2018;24(6):1296-1304.
403 Téglási V, Pipek O, Lózsa R*, et al.* PD-L1 Expression of Lung Cancer Cells, Unlike Infiltrating Immune Cells, Is Stable and Unaffected by Therapy During Brain Metastasis. *Clin Lung Cancer* 2019;20(5):363-369.e362.
404 Wang H, Agulnik J, Kasymjanova G*, et al.* The metastatic site does not influence PD-L1 expression in advanced non-small cell lung carcinoma. *Lung Cancer* 2019;132:36-38.
405 Cho JH, Sorensen SF, Choi YL*, et al.* Programmed Death Ligand 1 Expression in Paired Non-Small Cell Lung Cancer Tumor Samples. *Clin Lung Cancer* 2017;18(6):e473-e479.

406 Dill EA, Gru AA, Atkins KA*, et al.* PD-L1 Expression and Intratumoral Heterogeneity Across Breast Cancer Subtypes and Stages: An Assessment of 245 Primary and 40 Metastatic Tumors. *Am J Surg Pathol* 2017;41(3):334-342.
407 Li M, Li A, Zhou S*, et al.* Heterogeneity of PD-L1 expression in primary tumors and paired lymph node metastases of triple negative breast cancer. *BMC Cancer* 2018;18(1):4.
408 Terra S, Mansfield AS, Dong H*, et al.* Temporal and spatial heterogeneity of programmed cell death 1-Ligand 1 expression in malignant mesothelioma. *Oncoimmunology* 2017;6(11):e1356146.
409 Rasmussen JH, Lelkaitis G, Hakansson K*, et al.* Intratumor heterogeneity of PD-L1 expression in head and neck squamous cell carcinoma. *Br J Cancer* 2019;120(10):1003-1006.
410 Wei XL, Wu QN, Chen DL*, et al.* The Clinical and Biomarker Association of Programmed Death Ligand 1 and its Spatial Heterogeneous Expression in Colorectal Cancer. *J Cancer* 2018;9(23):4325-4333.
411 Sakakibara A, Inagaki Y, Imaoka E*, et al.* Divergence and heterogeneity of neoplastic PD-L1 expression: Two autopsy case reports of intravascular large B-cell lymphoma. *Pathol Int* 2019;69(3):148-154.
412 Takamori S, Toyokawa G, Okamoto I*, et al.* Discrepancy in Programmed Cell Death-Ligand 1 Between Primary and Metastatic Non-small Cell Lung Cancer. *Anticancer Res* 2017;37(8):4223-4228.
413 Takamori S, Takada K, Tagawa T*, et al.* Differences in PD-L1 expression on tumor and immune cells between lung metastases and corresponding primary tumors. *Surg Oncol* 2018;27(4):637-641.
414 Shen MJ, Xu LJ, Yang L*, et al.* Radiation alters PD-L1/NKG2D ligand levels in lung cancer cells and leads to immune escape from NK cell cytotoxicity via IL-6-MEK/Erk signaling pathway. *Oncotarget* 2017;8(46):80506-80520.
415 Peng J, Hamanishi J, Matsumura N*, et al.* Chemotherapy Induces Programmed Cell Death-Ligand 1 Overexpression via the Nuclear Factor-κB to Foster an Immunosuppressive Tumor Microenvironment in Ovarian Cancer. *Cancer Res* 2015;75(23):5034-5045.
416 Ghebeh H, Lehe C, Barhoush E*, et al.* Doxorubicin downregulates cell surface B7-H1 expression and upregulates its nuclear expression in breast cancer cells: role of B7-H1 as an anti-apoptotic molecule. *Breast Cancer Res* 2010;12(4):R48.
417 Chacon JA, Schutsky K, Powell DJ. The Impact of Chemotherapy, Radiation and Epigenetic Modifiers in Cancer Cell Expression of Immune Inhibitory and Stimulatory Molecules and Anti-Tumor Efficacy. *Vaccines (Basel)* 2016;4(4).
418 Zhang P, Ma Y, Lv C*, et al.* Upregulation of programmed cell death ligand 1 promotes resistance response in non-small-cell lung cancer patients treated with neo-adjuvant chemotherapy. *Cancer Sci* 2016;107(11):1563-1571.
419 Sheng J, Fang W, Yu J*, et al.* Expression of programmed death ligand-1 on tumor cells varies pre and post chemotherapy in non-small cell lung cancer. *Scientific Reports* 2016;6(1):20090.
420 Song Z, Yu X, Zhang Y. Altered expression of programmed death-ligand 1 after neo-adjuvant chemotherapy in patients with lung squamous cell carcinoma. *Lung Cancer* 2016;99:166-171.
421 Omori S, Kenmotsu H, Abe M*, et al.* Changes in programmed death ligand 1 expression in non-small cell lung cancer patients who received anticancer treatments. *International Journal of Clinical Oncology* 2018;23(6):1052-1059.
422 Rojkó L, Reiniger L, Téglási V*, et al.* Chemotherapy treatment is associated with altered PD-L1 expression in lung cancer patients. *J Cancer Res Clin Oncol* 2018;144(7):1219-1226.
423 Haratake N, Toyokawa G, Tagawa T*, et al.* Positive Conversion of PD-L1 Expression After Treatments with Chemotherapy and Nivolumab. *Anticancer Res* 2017;37(10):5713-5717.

424 Sakai H, Takeda M, Sakai K*, et al.* Impact of cytotoxic chemotherapy on PD-L1 expression in patients with non-small cell lung cancer negative for EGFR mutation and ALK fusion. *Lung Cancer* 2019;127:59-65.
425 Shin J, Chung JH, Kim SH*, et al.* Effect of Platinum-Based Chemotherapy on PD-L1 Expression on Tumor Cells in Non-small Cell Lung Cancer. *Cancer Res Treat* 2019;51(3):1086-1097.
426 Guo L, Song P, Xue X*, et al.* Variation of Programmed Death Ligand 1 Expression After Platinum-based Neoadjuvant Chemotherapy in Lung Cancer. *J Immunother* 2019;42(6):215-220.
427 Arnold KM, Flynn NJ, Raben A*, et al.* The Impact of Radiation on the Tumor Microenvironment: Effect of Dose and Fractionation Schedules. *Cancer Growth and Metastasis* 2018;11:1179064418761639.
428 Badiyan SN, Roach MC, Chuong MD*, et al.* Combining immunotherapy with radiation therapy in thoracic oncology. *Journal of thoracic disease* 2018;10(Suppl 21):S2492-S2507.
429 Van Limbergen EJ, De Ruysscher DK, Olivo Pimentel V*, et al.* Combining radiotherapy with immunotherapy: the past, the present and the future. *Br J Radiol* 2017;90(1076):20170157-20170157.
430 Aperio. Aperio ImageScope - Pathology Slide Viewing Software. Vol. 2019, 2019.
431 Bankhead P, Loughrey MB, Fernández JA*, et al.* QuPath: Open source software for digital pathology image analysis. *Scientific Reports* 2017;7(1):16878.
432 Roche V. Roche uPath enterprise software. Vol. 2020, 2019.
433 control Ufic. *TNM Classification of Malignant Tumours, 7th Edition*: Wiley-Blackwell, 2011.
434 Vennapusa B, Baker B, Kowanetz M*, et al.* Development of a PD-L1 Complementary Diagnostic Immunohistochemistry Assay (SP142) for Atezolizumab. *Applied immunohistochemistry & molecular morphology : AIMM* 2019;27(2):92-100.
435 Zhang HF, Zeng XT, Xing F*, et al.* The diagnostic accuracy of CT-guided percutaneous core needle biopsy and fine needle aspiration in pulmonary lesions: a meta-analysis. *Clin Radiol* 2016;71(1):e1-10.
436 Minot DM, Gilman EA, Aubry MC*, et al.* An investigation into false-negative transthoracic fine needle aspiration and core biopsy specimens. *Diagn Cytopathol* 2014;42(12):1063-1068.
437 Roy S, Nandi A, Das I*, et al.* Comparative study of cytology and immunocytochemistry with trucut biopsy and immunohistochemistry in diagnosis of localized lung lesions: A prospective study. *J Cytol* 2015;32(2):90-95.
438 Schneider F, Smith MA, Lane MC*, et al.* Adequacy of core needle biopsy specimens and fine-needle aspirates for molecular testing of lung adenocarcinomas. *Am J Clin Pathol* 2015;143(2):193-200; quiz 306.
439 Sholl LM, Aisner DL, Varella-Garcia M*, et al.* Multi-institutional Oncogenic Driver Mutation Analysis in Lung Adenocarcinoma: The Lung Cancer Mutation Consortium Experience. *J Thorac Oncol* 2015;10(5):768-777.
440 VanderLaan PA. Fine-needle aspiration and core needle biopsy: An update on 2 common minimally invasive tissue sampling modalities. *Cancer Cytopathol* 2016;124(12):862-870.
441 Thavarajah R, Mudimbaimannar V, Elizabeth J*, et al.* Chemical and physical basics of routine formaldehyde fixation. *Journal of Oral and Maxillofacial Pathology* 2012;16(3):400-405.
442 Shi SR, Shi Y, Taylor CR. Antigen retrieval immunohistochemistry: review and future prospects in research and diagnosis over two decades. *J Histochem Cytochem* 2011;59(1):13-32.
443 Orlando V. Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. *Trends Biochem Sci* 2000;25(3):99-104.
444 Hoffman EA, Frey BL, Smith LM*, et al.* Formaldehyde crosslinking: a tool for the study of chromatin complexes. *J Biol Chem* 2015;290(44):26404-26411.

445 Troiano NW, Ciovacco WA, Kacena MA. The Effects of Fixation and Dehydration on the Histological Quality of Undecalcified Murine Bone Specimens Embedded in Methylmethacrylate. *Journal of Histotechnology* 2009;32(1):27-31.
446 Hoetelmans RW, Prins FA, Cornelese-ten Velde I*, et al.* Effects of acetone, methanol, or paraformaldehyde on cellular structure, visualized by reflection contrast microscopy and transmission and scanning electron microscopy. *Appl Immunohistochem Mol Morphol* 2001;9(4):346-351.
447 Hobro AJ, Smith NI. An evaluation of fixation methods: Spatial and compositional cellular changes observed by Raman imaging. *Vibrational Spectroscopy* 2017;91:31-45.
448 van Essen HF, Verdaasdonk MA, Elshof SM*, et al.* Alcohol based tissue fixation as an alternative for formaldehyde: influence on immunohistochemistry. *J Clin Pathol* 2010;63(12):1090-1094.
449 Milcheva R, Janega P, Celec P*, et al.* Alcohol based fixatives provide excellent tissue morphology, protein immunoreactivity and RNA integrity in paraffin embedded tissue specimens. *Acta Histochem* 2013;115(3):279-289.
450 Piskorz AM, Ennis D, Macintyre G*, et al.* Methanol-based fixation is superior to buffered formalin for next-generation sequencing of DNA from clinical cancer samples. *Ann Oncol* 2016;27(3):532-539.
451 Turashvili G, Yang W, McKinney S*, et al.* Nucleic acid quantity and quality from paraffin blocks: defining optimal fixation, processing and DNA/RNA extraction techniques. *Exp Mol Pathol* 2012;92(1):33-43.
452 Vincek V, Nassiri M, Nadji M*, et al.* A tissue fixative that protects macromolecules (DNA, RNA, and protein) and histomorphology in clinical samples. *Lab Invest* 2003;83(10):1427-1435.
453 Srinivasan M, Sedmak D, Jewell S. Effect of fixatives and tissue processing on the content and integrity of nucleic acids. *Am J Pathol* 2002;161(6):1961-1971.
454 Cox ML, Schray CL, Luster CN*, et al.* Assessment of fixatives, fixation, and tissue processing on morphology and RNA integrity. *Exp Mol Pathol* 2006;80(2):183-191.
455 Engel KB, Moore HM. Effects of preanalytical variables on the detection of proteins by immunohistochemistry in formalin-fixed, paraffin-embedded tissue. *Arch Pathol Lab Med* 2011;135(5):537-543.
456 van den Broek LJ, van de Vijver MJ. Assessment of problems in diagnostic and research immunohistochemistry associated with epitope instability in stored paraffin sections. *Appl Immunohistochem Mol Morphol* 2000;8(4):316-321.
457 Ramos-Vara JA, Webster JD, DuSold D*, et al.* Immunohistochemical evaluation of the effects of paraffin section storage on biomarker stability. *Vet Pathol* 2014;51(1):102-109.
458 van Seijen M, Brcic L, Gonzales AN*, et al.* Impact of delayed and prolonged fixation on the evaluation of immunohistochemical staining on lung carcinoma resection specimen. *Virchows Arch* 2019;475(2):191-199.
459 Blind C, Koepenik A, Pacyna-Gengelbach M*, et al.* Antigenicity testing by immunohistochemistry after tissue oxidation. *J Clin Pathol* 2008;61(1):79-83.
460 Kaye J, Briceño Moraia L, Curren L*, et al.* Consent for Biobanking: The Legal Frameworks of Countries in the BioSHaRE-EU Project. *Biopreserv Biobank* 2016;14(3):195-200.
461 Kulangara K, Zhang N, Corigliano E*, et al.* Clinical Utility of the Combined Positive Score for Programmed Death Ligand-1 Expression and the Approval of Pembrolizumab for Treatment of Gastric Cancer. *Arch Pathol Lab Med* 2019;143(3):330-337.
462 Haragan A, Liebler DC, Das DM*, et al.* Accelerated instability testing reveals quantitative mass spectrometry overcomes specimen storage limitations associated with PD-L1 immunohistochemistry. *Lab Invest* 2020;100(6):874-886.
463 Squier TC, Bigelow DJ. Protein oxidation and age-dependent alterations in calcium homeostasis. *Front Biosci* 2000;5:D504-526.

464 Schey KL, Finley EL. Identification of peptide oxidation by tandem mass spectrometry. *Acc Chem Res* 2000;33(5):299-306.
465 Li S, Yang X, Zhang Y*, et al.* Assessment accuracy of core needle biopsy for hormone receptors in breast cancer: a meta-analysis. *Breast Cancer Res Treat* 2012;135(2):325-334.
466 You K, Park S, Ryu JM*, et al.* Comparison of Core Needle Biopsy and Surgical Specimens in Determining Intrinsic Biological Subtypes of Breast Cancer with Immunohistochemistry. *J Breast Cancer* 2017;20(3):297-303.
467 Meric-Bernstam F, Akcakanat A, Chen H*, et al.* Influence of biospecimen variables on proteomic biomarkers in breast cancer. *Clin Cancer Res* 2014;20(14):3870-3883.
468 Ahn S, Lee J, Cho MS*, et al.* Evaluation of Ki-67 Index in Core Needle Biopsies and Matched Breast Cancer Surgical Specimens. *Arch Pathol Lab Med* 2018;142(3):364-368.
469 Robertson S, Ronnlund C, de Boniface J*, et al.* Re-testing of predictive biomarkers on surgical breast cancer specimens is clinically relevant. *Breast Cancer Res Treat* 2019;174(3):795-805.
470 Gruchy JR, Barnes PJ, Dakin Haché KA. CytoLyt® fixation and decalcification pretreatments alter antigenicity in normal tissues compared with standard formalin fixation. *Appl Immunohistochem Mol Morphol* 2015;23(4):297-302.
471 Ilié M, Ngo-Mai M, Long-Mira E*, et al.* Using 22C3 Anti-PD-L1 Antibody Concentrate on Biopsy and Cytology Samples from Non-small Cell Lung Cancer Patients. *J Vis Exp* 2018(139).
472 Xu H, Bratton L, Nead M*, et al.* Comparison of programmed death-ligand 1 (PD-L1) immunostain for nonsmall cell lung carcinoma between paired cytological and surgical specimens. *Cytojournal* 2018;15:29.
473 Fowler CB, O'Leary TJ, Mason JT. Modeling formalin fixation and histological processing with ribonuclease A: effects of ethanol dehydration on reversal of formaldehyde cross-links. *Lab Invest* 2008;88(7):785-791.
474 Combs SE, Han G, Mani N*, et al.* Loss of antigenicity with tissue age in breast cancer. *Lab Invest* 2016;96(3):264-269.
475 Ehinger A, Bendahl PO, Rydén L*, et al.* Stability of oestrogen and progesterone receptor antigenicity in formalin-fixed paraffin-embedded breast cancer tissue over time. *Apmis* 2018;126(9):746-754.
476 Camp RL, Charette LA, Rimm DL. Validation of tissue microarray technology in breast carcinoma. *Lab Invest* 2000;80(12):1943-1949.
477 Dowsett T, Verghese E, Pollock S*, et al.* The value of archival tissue blocks in understanding breast cancer biology. *J Clin Pathol* 2014;67(3):272-275.
478 Blows FM, Ali HR, Dawson SJ*, et al.* Decline in Antigenicity of Tumor Markers by Storage Time Using Pathology Sections Cut From Tissue Microarrays. *Appl Immunohistochem Mol Morphol* 2016;24(3):221-226.
479 Grillo F, Pigozzi S, Ceriolo P*, et al.* Factors affecting immunoreactivity in long-term storage of formalin-fixed paraffin-embedded tissue sections. *Histochem Cell Biol* 2015;144(1):93-99.
480 Jacobs TW, Prioleau JE, Stillman IE*, et al.* Loss of tumor marker-immunostaining intensity on stored paraffin slides of breast cancer. *J Natl Cancer Inst* 1996;88(15):1054-1059.
481 Wester K, Wahlund E, Sundstrom C*, et al.* Paraffin section storage and immunohistochemistry. Effects of time, temperature, fixation, and retrieval protocol with emphasis on p53 protein and MIB1 antigen. *Appl Immunohistochem Mol Morphol* 2000;8(1):61-70.
482 Gelb AB, Freeman VA, Astrow SH. Evaluation of methods for preserving PTEN antigenicity in stored paraffin sections. *Appl Immunohistochem Mol Morphol* 2011;19(6):569-573.
483 Vasilescu J, Guo X, Kast J. Identification of protein-protein interactions using in vivo cross-linking and mass spectrometry. *Proteomics* 2004;4(12):3845-3854.
484 Scalia CR, Boi G, Bolognesi MM*, et al.* Antigen Masking During Fixation and Embedding, Dissected. *J Histochem Cytochem* 2017;65(1):5-20.

485 Shi SR, Key ME, Kalra KL. Antigen retrieval in formalin-fixed, paraffin-embedded tissues: an enhancement method for immunohistochemical staining based on microwave oven heating of tissue sections. *J Histochem Cytochem* 1991;39(6):741-748.
486 Shi SR, Cote RJ, Taylor CR. Antigen retrieval techniques: current perspectives. *J Histochem Cytochem* 2001;49(8):931-937.
487 Leong TY-M, Leong AS-Y. How Does Antigen Retrieval Work? *Advances in Anatomic Pathology* 2007;14(2):129-131.
488 Quievryn G, Zhitkovich A. Loss of DNA-protein crosslinks from formaldehyde-exposed cells occurs through spontaneous hydrolysis and an active repair process linked to proteosome function. *Carcinogenesis* 2000;21(8):1573-1580.
489 Costa M, Zhitkovich A, Harris M*, et al.* DNA-protein cross-links produced by various chemicals in cultured human lymphoma cells. *J Toxicol Environ Health* 1997;50(5):433-449.
490 Mirlacher M, Kasper M, Storz M*, et al.* Influence of slide aging on results of translational research studies using immunohistochemistry. *Modern Pathology* 2004;17(11):1414-1420.
491 Forse CL, Pinnaduwage D, Bull SB*, et al.* Fresh Cut Versus Stored Cut Paraffin-embedded Tissue: Effect on Immunohistochemical Staining for Common Breast Cancer Markers. *Appl Immunohistochem Mol Morphol* 2019;27(3):231-237.
492 Ezaki T. Antigen retrieval on formaldehyde-fixed paraffin sections: its potential drawbacks and optimization for double immunostaining. *Micron* 2000;31(6):639-649.
493 Bogen SA, Vani K, Sompuram SR. Molecular mechanisms of antigen retrieval: antigen retrieval reverses steric interference caused by formalin-induced cross-links. *Biotech Histochem* 2009;84(5):207-215.
494 Sompuram SR, Vani K, Bogen SA. A molecular model of antigen retrieval using a peptide array. *Am J Clin Pathol* 2006;125(1):91-98.
495 Scalia CR, Gendusa R, Basciu M*, et al.* Epitope recognition in the human-pig comparison model on fixed and embedded material. *J Histochem Cytochem* 2015;63(10):805-822.
496 Hao G, Wesolowski JS, Jiang X*, et al.* Epitope characterization of an anti-PD-L1 antibody using orthogonal approaches. *J Mol Recognit* 2015;28(4):269-276.
497 Shin HJ, Kalapurakal SK, Lee JJ*, et al.* Comparison of p53 immunoreactivity in fresh-cut versus stored slides with and without microwave heating. *Mod Pathol* 1997;10(3):224-230.
498 Dodson A, Parry S, Lissenberg-Witte B*, et al.* External quality assessment demonstrates that PD-L1 22C3 and SP263 assays are systematically different. *The Journal of Pathology: Clinical Research* 2020;6(2):138-145.
499 Parra ER, Villalobos P, Mino B*, et al.* Comparison of Different Antibody Clones for Immunohistochemistry Detection of Programmed Cell Death Ligand 1 (PD-L1) on Non-Small Cell Lung Carcinoma. *Appl Immunohistochem Mol Morphol* 2018;26(2):83-93.
500 Grifa RA, Pozzoli G. Performance evaluation of in vitro diagnostic medical devices: Methodology and differences compared to studies on other medical devices. *Microchemical Journal* 2018;136:279-282.
501 Fujimoto D, Sato Y, Uehara K*, et al.* Predictive Performance of Four Programmed Cell Death Ligand 1 Assay Systems on Nivolumab Response in Previously Treated Patients with Non-Small Cell Lung Cancer. *J Thorac Oncol* 2018;13(3):377-386.
502 Faratian D, Kay C, Robson T*, et al.* Automated image analysis for high-throughput quantitative detection of ER and PR expression levels in large-scale clinical studies: the TEAM Trial Experience. *Histopathology* 2009;55(5):587-593.
503 Zarrella ER, Coulter M, Welsh AW*, et al.* Automated measurement of estrogen receptor in breast cancer: a comparison of fluorescent and chromogenic methods of measurement. *Lab Invest* 2016;96(9):1016-1025.
504 Nielsen PS, Lindebjerg J, Rasmussen J*, et al.* Virtual microscopy: an evaluation of its validity and diagnostic performance in routine histologic diagnosis of skin tumors. *Hum Pathol* 2010;41(12):1770-1776.

505 Houghton JP, Ervine AJ, Kenny SL*, et al.* Concordance between digital pathology and light microscopy in general surgical pathology: a pilot study of 100 cases. *J Clin Pathol* 2014;67(12):1052-1055.
506 Snead DR, Tsang YW, Meskiri A*, et al.* Validation of digital pathology imaging for primary histopathological diagnosis. *Histopathology* 2016;68(7):1063-1072.
507 Hartman DJ, Pantanowitz L, McHugh JS*, et al.* Enterprise Implementation of Digital Pathology: Feasibility, Challenges, and Opportunities. *J Digit Imaging* 2017;30(5):555-560.
508 Williams BJ, Jayewardene D, Treanor D. Digital immunohistochemistry implementation, training and validation: experience and technical notes from a large clinical laboratory. *J Clin Pathol* 2019;72(5):373-378.
509 Van Es SL, Kumar RK, Pryor WM*, et al.* Cytopathology whole slide images and adaptive tutorials for postgraduate pathology trainees: a randomized crossover trial. *Hum Pathol* 2015;46(9):1297-1305.
510 Dee FR, Donnelly A, Radio S*, et al.* Utility of 2-D and 3-D virtual microscopy in cervical cytology education and testing. *Acta Cytol* 2007;51(4):523-529.
511 Ross J, Greaves J, Earls P*, et al.* Digital vs traditional: Are diagnostic accuracy rates similar for glass slides vs whole slide images in a non-gynaecological external quality assurance setting? *Cytopathology* 2018;29(4):326-334.
512 Van Es SL. Digital pathology: semper ad meliora. *Pathology* 2019;51(1):1-10.
513 Pantanowitz L. Digital images and the future of digital pathology. *J Pathol Inform* 2010;1.
514 Madabhushi A, Lee G. Image analysis and machine learning in digital pathology: Challenges and opportunities. *Med Image Anal* 2016;33:170-175.
515 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.
516 FDA. Artificial Intelligence and Machine Learning in Software as a Medical Device. Vol. 2020, 2020.
517 Gandomkar Z, Brennan PC, Mello-Thoms C. Computer-based image analysis in breast pathology. *J Pathol Inform* 2016;7:43.
518 Ordi O, Bomb, #237*, et al.* Virtual microscopy in the undergraduate teaching of pathology. *Journal of Pathology Informatics* 2015;6(1):1-1.
519 Kumar A, Rao A, Bhavani S*, et al.* Automated analysis of immunohistochemistry images identifies candidate location biomarkers for cancers. *Proc Natl Acad Sci U S A* 2014;111(51):18249-18254.
520 Newberg J, Murphy RF. A framework for the automated analysis of subcellular patterns in human protein atlas images. *J Proteome Res* 2008;7(6):2300-2308.
521 Guirado R, Carceller H, Castillo-Gomez E*, et al.* Automated analysis of images for molecular quantification in immunohistochemistry. *Heliyon* 2018;4(6):e00669.
522 Su HK, Wenig BM, Haser GC*, et al.* Inter-Observer Variation in the Pathologic Identification of Minimal Extrathyroidal Extension in Papillary Thyroid Carcinoma. *Thyroid* 2016;26(4):512-517.
523 Vandenberghe ME, Scott MLJ, Scorer PW*, et al.* Relevance of deep learning to facilitate the diagnosis of HER2 status in breast cancer. *Scientific Reports* 2017;7(1):45938.
524 Nederlof I, De Bortoli D, Bareche Y*, et al.* Comprehensive evaluation of methods to assess overall and cell-specific immune infiltrates in breast cancer. *Breast Cancer Research* 2019;21(1):151.
525 Niazi MKK, Abas FS, Senaras C*, et al.* Nuclear IHC enumeration: A digital phantom to evaluate the performance of automated algorithms in digital pathology. *PLoS One* 2018;13(5):e0196547.
526 Ianni JD, Soans RE, Sankarapandian S*, et al.* Tailored for Real-World: A Whole Slide Image Classification System Validated on Uncurated Multi-Site Data Emulating the Prospective Pathology Workload. *Scientific Reports* 2020;10(1):3217.

527 Roche V. uPath PD-L1 (SP263) image analysis, NSCLC. Vol. 2020, 2020.
528 Fitzgibbons PL, Bradley LA, Fatheree LA*, et al.* Principles of Analytic Validation of Immunohistochemical Assays: Guideline From the College of American Pathologists Pathology and Laboratory Quality Center. *Archives of Pathology & Laboratory Medicine* 2014;138(11):1432-1443.
529 Ivell R, Teerds K, Hoffman GE. Proper application of antibodies for immunohistochemical detection: antibody crimes and how to prevent them. *Endocrinology* 2014;155(3):676-687.
530 Nambirajan A, Husain N, Shukla S*, et al.* Comparison of laboratory-developed test & validated assay of programmed death ligand-1 immunohistochemistry in non-small-cell lung carcinoma. *Indian Journal of Medical Research* 2019;150(4):376-384.
531 Koelzer VH, Gisler A, Hanhart JC*, et al.* Digital image analysis improves precision of PD-L1 scoring in cutaneous melanoma. *Histopathology* 2018;73(3):397-406.
532 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. *J Oncol* 2018;2018:2937012.
533 Taylor CR, Jadhav AP, Gholap A*, et al.* A Multi-Institutional Study to Evaluate Automated Whole Slide Scoring of Immunohistochemistry for Assessment of Programmed Death-Ligand 1 (PD-L1) Expression in Non-Small Cell Lung Cancer. *Appl Immunohistochem Mol Morphol* 2019;27(4):263-269.
534 Widmaier M, Wiestler T, Walker J*, et al.* Comparison of continuous measures across diagnostic PD-L1 assays in non-small cell lung cancer using automated image analysis. *Modern Pathology* 2020;33(3):380-390.
535 Cooper WA, Russell PA, Cherian M*, et al.* Intra- and Interobserver Reproducibility Assessment of PD-L1 Biomarker in Non-Small Cell Lung Cancer. *Clin Cancer Res* 2017;23(16):4569-4577.
536 Dimitriou N, Arandjelovic O, Caie PD. Deep Learning for Whole Slide Image Analysis: An Overview. *Front Med (Lausanne)* 2019;6:264.
537 Iizuka O, Kanavati F, Kato K*, et al.* Deep Learning Models for Histopathological Classification of Gastric and Colonic Epithelial Tumours. *Scientific Reports* 2020;10(1):1504.
538 Chen C, Qi XJ, Cao YW*, et al.* Bladder Tumor Heterogeneity: The Impact on Clinical Treatment. *Urol Int* 2015;95(1):1-8.
539 Saha A, Harowicz MR, Cain EH*, et al.* Intra-tumor molecular heterogeneity in breast cancer: definitions of measures and association with distant recurrence-free survival. *Breast Cancer Res Treat* 2018;172(1):123-132.
540 Song J, Dong D, Huang Y*, et al.* Association between tumor heterogeneity and overall survival in patients with non-small cell lung cancer. *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, 2016, p. 1249-1252.
541 Miyauchi T, Yaguchi T, Kawakami Y. Inter-patient and Intra-tumor Heterogeneity in the Sensitivity to Tumor-targeted Immunity in Colorectal Cancer. *Nihon Rinsho Meneki Gakkai Kaishi* 2017;40(1):54-59.
542 Uruga H, Bozkurtlar E, Huynh TG*, et al.* Programmed Cell Death Ligand (PD-L1) Expression in Stage II and III Lung Adenocarcinomas and Nodal Metastases. *J Thorac Oncol* 2017;12(3):458-466.
543 Travis WD, Brambilla E, Noguchi M*, et al.* International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer* 2011;6(2):244-285.
544 Jamal-Hanjani M, Hackshaw A, Ngai Y*, et al.* Tracking genomic cancer evolution for precision medicine: the lung TRACERx study. *PLoS Biol* 2014;12(7):e1001906.
545 Jamal-Hanjani M, Wilson GA, McGranahan N*, et al.* Tracking the Evolution of Non-Small-Cell Lung Cancer. *N Engl J Med* 2017;376(22):2109-2121.

546 Negrao MV, Quek K, Zhang J*, et al.* TRACERx: Tracking tumor evolution to impact the course of lung cancer. *J Thorac Cardiovasc Surg* 2018;155(3):1199-1202.
547 Gerlinger M, Rowan AJ, Horswell S*, et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* 2012;366(10):883-892.
548 Chen ZY, Zhong WZ, Zhang XC*, et al.* EGFR mutation heterogeneity and the mixed response to EGFR tyrosine kinase inhibitors of lung adenocarcinomas. *Oncologist* 2012;17(7):978-985.
549 Kim H, Xu X, Yoo SB*, et al.* Discordance between anaplastic lymphoma kinase status in primary non-small-cell lung cancers and their corresponding metastases. *Histopathology* 2013;62(2):305-314.
550 Jakobsen JN, Santoni-Rugiu E, Ravn J*, et al.* Intratumour variation of biomarker expression by immunohistochemistry in resectable non-small cell lung cancer. *Eur J Cancer* 2013;49(11):2494-2503.
551 Mansuet-Lupo A, Zouiti F, Alifano M*, et al.* Intratumoral distribution of EGFR mutations and copy number in metastatic lung cancer, what impact on the initial molecular diagnosis? *J Transl Med* 2014;12:131.
552 Kordiak J, Szemraj J, Grabska-Kobylecka I*, et al.* Intratumor heterogeneity and tissue distribution of KRAS mutation in non-small cell lung cancer: implications for detection of mutated KRAS oncogene in exhaled breath condensate. *J Cancer Res Clin Oncol* 2019;145(1):241-251.
553 Remon J, Majem M. EGFR mutation heterogeneity and mixed response to EGFR tyrosine kinase inhibitors of non small cell lung cancer: a clue to overcoming resistance. *Transl Lung Cancer Res* 2013;2(6):445-448.
554 Sakurada A, Lara-Guerra H, Liu N*, et al.* Tissue Heterogeneity of EGFR Mutation in Lung Adenocarcinoma. *Journal of Thoracic Oncology* 2008;3(5):527-529.
555 Jiang S-X, Yamashita K, Yamamoto M*, et al.* EGFR genetic heterogeneity of nonsmall cell lung cancers contributing to acquired gefitinib resistance. *International Journal of Cancer* 2008;123(11):2480-2486.
556 Majem M, Remon J. Tumor heterogeneity: evolution through space and time in EGFR mutant non small cell lung cancer patients. *Transl Lung Cancer Res* 2013;2(3):226-237.
557 Siravegna G, Marsoni S, Siena S*, et al.* Integrating liquid biopsies into the management of cancer. *Nat Rev Clin Oncol* 2017;14(9):531-548.
558 Painter JT, Clayton NP, Herbert RA. Useful immunohistochemical markers of tumor differentiation. *Toxicol Pathol* 2010;38(1):131-141.
559 Gown AM. Diagnostic Immunohistochemistry: What Can Go Wrong and How to Prevent It. *Arch Pathol Lab Med* 2016;140(9):893-898.
560 Potts SJ, Krueger JS, Landis ND*, et al.* Evaluating tumor heterogeneity in immunohistochemistry-stained breast cancer tissue. *Lab Invest* 2012;92(9):1342-1357.
561 Zlatian OM, Comanescu MV, Rosu AF*, et al.* Histochemical and immunohistochemical evidence of tumor heterogeneity in colorectal cancer. *Rom J Morphol Embryol* 2015;56(1):175-181.
562 Ock CY, Keam B, Kim S*, et al.* Pan-Cancer Immunogenomic Perspective on the Tumor Microenvironment Based on PD-L1 and CD8 T-Cell Infiltration. *Clin Cancer Res* 2016;22(9):2261-2270.
563 Chen S, Crabill GA, Pritchard TS*, et al.* Mechanisms regulating PD-L1 expression on tumor and immune cells. *J Immunother Cancer* 2019;7(1):305.
564 Hao D, Wang G, Yang W*, et al.* Reactive versus Constitutive: Reconcile the Controversial Results about the Prognostic Value of PD-L1 Expression in cancer. *Int J Biol Sci* 2019;15(9):1933-1941.
565 Ribas A, Hu-Lieskovan S. What does PD-L1 positive or negative mean? *J Exp Med* 2016;213(13):2835-2840.

566 Casey SC, Tong L, Li Y*, et al.* MYC regulates the antitumor immune response through CD47 and PD-L1. *Science* 2016;352(6282):227-231.
567 NICE. Lung cancer – everything NICE says in an interactive flowchart. Vol. 2020, 2020.
568 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.
569 Zito Marino F, Bianco R, Accardo M*, et al.* Molecular heterogeneity in lung cancer: from mechanisms of origin to clinical implications. *Int J Med Sci* 2019;16(7):981-989.
570 Keller MD, Neppl C, Irmak Y*, et al.* Adverse prognostic value of PD-L1 expression in primary resected pulmonary squamous cell carcinomas and paired mediastinal lymph node metastases. *Mod Pathol* 2018;31(1):101-110.
571 Forkman J. Estimator and Tests for Common Coefficients of Variation in Normal Distributions. *Communications in Statistics - Theory and Methods* 2009;38(2):233-251.
572 Gagné A, Enlow W, Pigeon MA*, et al.* Comprehensive Assessment of PD-L1 Staining Heterogeneity in Pulmonary Adenocarcinomas Using Tissue Microarrays: Impact of the Architecture Pattern and the Number of Cores. *Am J Surg Pathol* 2018;42(5):687-694.
573 Gosney JR, Boothman AM, Ratcliffe M*, et al.* Cytology for PD-L1 testing: A systematic review. *Lung Cancer* 2020;141:101-106.
574 Jain D, Nambirajan A, Borczuk A*, et al.* Immunocytochemistry for predictive biomarker testing in lung cancer cytology. *Cancer Cytopathol* 2019;127(5):325-339.
575 Wang Y, Wang H, Yao H*, et al.* Regulation of PD-L1: Emerging Routes for Targeting Tumor Immune Evasion. *Front Pharmacol* 2018;9:536.
576 Zieliński P, Dyszkiewicz W, Piwkowski CT*, et al.* Can the condition of the cell microenvironment of mediastinal lymph nodes help predict the risk of metastases in non-small cell lung cancer? *Cancer Epidemiol* 2009;33(5):387-390.
577 Stankovic B, Bjorhovde HAK, Skarshaug R*, et al.* Immune Cell Composition in Human Non-small Cell Lung Cancer. *Front Immunol* 2018;9:3101.
578 Turajlic S, Swanton C. Metastasis as an evolutionary process. *Science* 2016;352(6282):169-175.
579 Caswell DR, Chuang CH, Yang D*, et al.* Obligate progression precedes lung adenocarcinoma dissemination. *Cancer Discov* 2014;4(7):781-789.
580 Aramaki N, Ishii G, Yamada E*, et al.* Drastic morphological and molecular differences between lymph node micrometastatic tumors and macrometastatic tumors of lung adenocarcinoma. *J Cancer Res Clin Oncol* 2016;142(1):37-46.
581 Abu Hejleh T, Furqan M, Ballas Z*, et al.* The clinical significance of soluble PD-1 and PD-L1 in lung cancer. *Crit Rev Oncol Hematol* 2019;143:148-152.
582 Turajlic S, Xu H, Litchfield K*, et al.* Tracking Cancer Evolution Reveals Constrained Routes to Metastases: TRACERx Renal. *Cell* 2018;173(3):581-594 e512.
583 Ishaque N, Abba ML, Hauser C*, et al.* Whole genome sequencing puts forward hypotheses on metastasis evolution and therapy in colorectal cancer. *Nature Communications* 2018;9(1):4782.
584 Jalil BA, Yasufuku K, Khan AM. Uses, limitations, and complications of endobronchial ultrasound. *Proc (Bayl Univ Med Cent)* 2015;28(3):325-330.
585 Muthu V, Sehgal IS, Dhooria S*, et al.* Endobronchial Ultrasound-Guided Transbronchial Needle Aspiration: Techniques and Challenges. *J Cytol* 2019;36(1):65-70.
586 Janku F. Tumor heterogeneity in the clinic: is it a real problem? *Ther Adv Med Oncol* 2014;6(2):43-51.
587 Akinleye A, Rasool Z. Immune checkpoint inhibitors of PD-L1 as cancer therapeutics. *J Hematol Oncol* 2019;12(1):92.

588 Sun L, Zhang L, Yu J*, et al.* Clinical efficacy and safety of anti-PD-1/PD-L1 inhibitors for the treatment of advanced or metastatic cancer: a systematic review and meta-analysis. *Sci Rep* 2020;10(1):2083.
589 de Charette M, Marabelle A, Houot R. Turning tumour cells into antigen presenting cells: The next step to improve cancer immunotherapy? *Eur J Cancer* 2016;68:134-147.
590 Lee SK, Seo SH, Kim BS*, et al.* IFN-gamma regulates the expression of B7-H1 in dermal fibroblast cells. *J Dermatol Sci* 2005;40(2):95-103.
591 Schumacher TN, Schreiber RD. Neoantigens in cancer immunotherapy. *Science* 2015;348(6230):69-74.
592 Lawrence MS, Stojanov P, Polak P*, et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 2013;499(7457):214-218.
593 Scognamiglio T, Chen YT. Beyond the Percentages of PD-L1-Positive Tumor Cells: Induced Versus Constitutive PD-L1 Expression in Primary and Metastatic Head and Neck Squamous Cell Carcinoma. *Head Neck Pathol* 2018;12(2):221-229.
594 Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol* 2017;18(1):220.
595 Newman AM, Liu CL, Green MR*, et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* 2015;12(5):453-457.
596 Chen B, Khodadoust MS, Liu CL*, et al.* Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol Biol* 2018;1711:243-259.
597 Galon J, Mlecnik B, Bindea G*, et al.* Towards the introduction of the 'Immunoscore' in the classification of malignant tumours. *J Pathol* 2014;232(2):199-209.
598 Angell HK, Bruni D, Barrett JC*, et al.* The Immunoscore: Colon Cancer and Beyond. *Clin Cancer Res* 2020;26(2):332-339.
599 Galon J, Costes A, Sanchez-Cabo F*, et al.* Type, Density, and Location of Immune Cells Within Human Colorectal Tumors Predict Clinical Outcome. *Science* 2006;313(5795):1960-1964.
600 Binnewies M, Roberts EW, Kersten K*, et al.* Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat Med* 2018;24(5):541-550.
601 Bonaventura P, Shekarian T, Alcazer V*, et al.* Cold Tumors: A Therapeutic Challenge for Immunotherapy. *Front Immunol* 2019;10:168.
602 Seo JS, Kim A, Shin JY*, et al.* Comprehensive analysis of the tumor immune micro-environment in non-small cell lung cancer for efficacy of checkpoint inhibitor. *Sci Rep* 2018;8(1):14576.
603 Gennen K, Kasmann L, Taugner J*, et al.* Prognostic value of PD-L1 expression on tumor cells combined with CD8+ TIL density in patients with locally advanced non-small cell lung cancer treated with concurrent chemoradiotherapy. *Radiat Oncol* 2020;15(1):5.
604 Kmiecik J, Poli A, Brons NH*, et al.* Elevated CD3+ and CD8+ tumor-infiltrating immune cells correlate with prolonged survival in glioblastoma patients despite integrated immunosuppressive mechanisms in the tumor microenvironment and at the systemic level. *J Neuroimmunol* 2013;264(1-2):71-83.
605 Matsushita H, Vesely MD, Koboldt DC*, et al.* Cancer exome analysis reveals a T-cell-dependent mechanism of cancer immunoediting. *Nature* 2012;482(7385):400-404.
606 Hodge G, Barnawi J, Jurisevic C*, et al.* Lung cancer is associated with decreased expression of perforin, granzyme B and interferon (IFN)-gamma by infiltrating lung tissue T cells, natural killer (NK) T-like and NK cells. *Clin Exp Immunol* 2014;178(1):79-85.
607 Hanson HL, Donermeyer DL, Ikeda H*, et al.* Eradication of established tumors by CD8+ T cell adoptive immunotherapy. *Immunity* 2000;13(2):265-276.
608 Slaney CY, Kershaw MH, Darcy PK. Trafficking of T cells into tumors. *Cancer Res* 2014;74(24):7168-7174.
609 Maimela NR, Liu S, Zhang Y. Fates of CD8+ T cells in Tumor Microenvironment. *Comput Struct Biotechnol J* 2019;17:1-13.

610 Liu JY, Li F, Wang LP*, et al.* CTL- vs Treg lymphocyte-attracting chemokines, CCL4 and CCL20, are strong reciprocal predictive markers for survival of patients with oesophageal squamous cell carcinoma. *British Journal of Cancer* 2015;113(5):747-755.
611 Bellone M, Calcinotto A. Ways to enhance lymphocyte trafficking into tumors and fitness of tumor infiltrating lymphocytes. *Front Oncol* 2013;3:231.
612 Wherry EJ. T cell exhaustion. *Nature Immunology* 2011;12(6):492-499.
613 Okoye IS, Houghton M, Tyrrell L*, et al.* Coinhibitory Receptor Expression and Immune Checkpoint Blockade: Maintaining a Balance in CD8(+) T Cell Responses to Chronic Viral Infections and Cancer. *Front Immunol* 2017;8:1215.
614 Lu L, Barbi J, Pan F. The regulation of immune tolerance by FOXP3. *Nat Rev Immunol* 2017;17(11):703-717.
615 Ghiringhelli F, Puig PE, Roux S*, et al.* Tumor cells convert immature myeloid dendritic cells into TGF-beta-secreting cells inducing CD4+CD25+ regulatory T cell proliferation. *J Exp Med* 2005;202(7):919-929.
616 Shang B, Liu Y, Jiang SJ*, et al.* Prognostic value of tumor-infiltrating FoxP3+ regulatory T cells in cancers: a systematic review and meta-analysis. *Sci Rep* 2015;5:15179.
617 Hu G, Li Z, Wang S. Tumor-infiltrating FoxP3(+) Tregs predict favorable outcome in colorectal cancer patients: A meta-analysis. *Oncotarget* 2017;8(43):75361-75371.
618 Schoenhals JE, Seyedin SN, Anderson C*, et al.* Uncovering the immune tumor microenvironment in non-small cell lung cancer to understand response rates to checkpoint blockade and radiation. *Transl Lung Cancer Res* 2017;6(2):148-158.
619 Campana D, Thompson JS, Amlot P*, et al.* The cytoplasmic expression of CD3 antigens in normal and malignant cells of the T lymphoid lineage. *J Immunol* 1987;138(2):648-655.
620 Leong F, J., Leong AS-Y. *Leong's Manual of Diagnostic Antibodies for Immunohistology*: Cambridge University Press, 2016.
621 Hori S, Nomura T, Sakaguchi S. Control of regulatory T cell development by the transcription factor Foxp3. *Science* 2003;299(5609):1057-1061.
622 Mellor-Heineke S, Villanueva J, Jordan MB*, et al.* Elevated Granzyme B in Cytotoxic Lymphocytes is a Signature of Immune Activation in Hemophagocytic Lymphohistiocytosis. *Front Immunol* 2013;4:72.
623 Hofman P, Badoual C, Henderson F*, et al.* Multiplexed Immunohistochemistry for Molecular and Immune Profiling in Lung Cancer-Just About Ready for Prime-Time? *Cancers (Basel)* 2019;11(3).
624 Dixon AR, Bathany C, Tsuei M*, et al.* Recent developments in multiplexing techniques for immunohistochemistry. *Expert Rev Mol Diagn* 2015;15(9):1171-1186.
625 Laboratories M. Mosaic Laboratories Publications. Vol. 2020: MOSAIC Laboratories, 2020.
626 Polyak K, Marusyk A. Cancer: Clonal cooperation. *Nature* 2014;508(7494):52-53.
627 Caswell DR, Swanton C. The role of tumour heterogeneity and clonal cooperativity in metastasis, immune evasion and clinical outcome. *BMC Med* 2017;15(1):133.
628 Kitamura T, Qian BZ, Pollard JW. Immune cell promotion of metastasis. *Nat Rev Immunol* 2015;15(2):73-86.
629 Kitamura T, Qian BZ, Soong D*, et al.* CCL2-induced chemokine cascade promotes breast cancer metastasis by enhancing retention of metastasis-associated macrophages. *J Exp Med* 2015;212(7):1043-1059.
630 Headley MB, Bins A, Nip A*, et al.* Visualization of immediate immune responses to pioneer metastatic cells in the lung. *Nature* 2016;531(7595):513-517.
631 Lin Y, Xu J, Lan H. Tumor-associated macrophages in tumor metastasis: biological roles and clinical therapeutic applications. *Journal of Hematology & Oncology* 2019;12(1):76.
632 DeNardo DG, Andreu P, Coussens LM. Interactions between lymphocytes and myeloid cells regulate pro- versus anti-tumor immunity. *Cancer Metastasis Rev* 2010;29(2):309-316.

633 Wculek SK, Malanchi I. Neutrophils support lung colonization of metastasis-initiating breast cancer cells. *Nature* 2015;528(7582):413-417.
634 Nguyen DX, Bos PD, Massague J. Metastasis: from dissemination to organ-specific colonization. *Nat Rev Cancer* 2009;9(4):274-284.
635 Mroz EA, Rocco JW. The challenges of tumor genetic diversity. *Cancer* 2017;123(6):917-927.
636 Zhao J, Chen AX, Gartrell RD*, et al.* Immune and genomic correlates of response to anti-PD-1 immunotherapy in glioblastoma. *Nat Med* 2019;25(3):462-469.
637 Gauci ML, Lanoy E, Champiat S*, et al.* Long-Term Survival in Patients Responding to Anti-PD-1/PD-L1 Therapy and Disease Outcome upon Treatment Discontinuation. *Clin Cancer Res* 2019;25(3):946-956.
638 Jez M, Bas T, Veber M*, et al.* The hazards of DAPI photoconversion: effects of dye, mounting media and fixative, and how to minimize the problem. *Histochem Cell Biol* 2013;139(1):195-204.
639 Conroy JM, Pabla S, Nesline MK*, et al.* Next generation sequencing of PD-L1 for predicting response to immune checkpoint inhibitors. *J Immunother Cancer* 2019;7(1):18.
640 Ma J, Zheng B, Goswami S*, et al.* PD1(Hi) CD8(+) T cells correlate with exhausted signature and poor clinical outcome in hepatocellular carcinoma. *J Immunother Cancer* 2019;7(1):331.
641 Cancer AJCo. Lung Cancer Staging. Vol. 2020, 2018.
642 Planchard D, Popat S, Kerr K*, et al.* Metastatic non-small cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2018;29(Suppl 4):iv192-iv237.
643 Akamatsu H, Ninomiya K, Kenmotsu H*, et al.* The Japanese Lung Cancer Society Guideline for non-small cell lung cancer, stage IV. *Int J Clin Oncol* 2019;24(7):731-770.
644 Antonia SJ, Villegas A, Daniel D*, et al.* Durvalumab after Chemoradiotherapy in Stage III Non-Small-Cell Lung Cancer. *N Engl J Med* 2017;377(20):1919-1929.
645 Postmus PE, Kerr KM, Oudkerk M*, et al.* Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2017;28 Suppl 4:iv1-iv21.
646 ESMO. Metastatic Non-Small-Cell Lung Cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Vol. 2020, 2020.
647 Incorvaia L, Fanale D, Badalamenti G*, et al.* Programmed Death Ligand 1 (PD-L1) as a Predictive Biomarker for Pembrolizumab Therapy in Patients with Advanced Non-Small-Cell Lung Cancer (NSCLC). *Adv Ther* 2019;36(10):2600-2617.
648 Anagnostou V, Yarchoan M, Hansen AR*, et al.* Immuno-oncology Trial Endpoints: Capturing Clinically Meaningful Activity. *Clin Cancer Res* 2017;23(17):4959-4969.
649 Mushti SL, Mulkey F, Sridhara R. Evaluation of Overall Response Rate and Progression-Free Survival as Potential Surrogate Endpoints for Overall Survival in Immunotherapy Trials. *Clin Cancer Res* 2018;24(10):2268-2275.
650 Pilz LR, Manegold C, Schmid-Bindert G. Statistical considerations and endpoints for clinical lung cancer studies: Can progression free survival (PFS) substitute overall survival (OS) as a valid endpoint in clinical trials for advanced non-small-cell lung cancer? *Transl Lung Cancer Res* 2012;1(1):26-35.
651 Cheema PK, Burkes RL. Overall survival should be the primary endpoint in clinical trials for advanced non-small-cell lung cancer. *Curr Oncol* 2013;20(2):e150-160.
652 Eisenhauer EA, Therasse P, Bogaerts J*, et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;45(2):228-247.
653 Seymour L, Bogaerts J, Perrone A*, et al.* iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. *The Lancet Oncology* 2017;18(3):e143-e152.
654 Oken MM, Creech RH, Tormey DC*, et al.* Toxicity and response criteria of the Eastern Cooperative Oncology Group. *Am J Clin Oncol* 1982;5(6):649-655.

655 Tamiya M, Tamiya A, Hosoya K*, et al.* Efficacy and safety of pembrolizumab as first-line therapy in advanced non-small cell lung cancer with at least 50% PD-L1 positivity: a multicenter retrospective cohort study (HOPE-001). *Invest New Drugs* 2019;37(6):1266-1273.
656 Ksienski D, Wai ES, Croteau N*, et al.* Pembrolizumab for advanced nonsmall cell lung cancer: Efficacy and safety in everyday clinical practice. *Lung Cancer* 2019;133:110-116.
657 Juergens RA, Mariano C, Jolivet J*, et al.* Real-world benefit of nivolumab in a Canadian non-small-cell lung cancer cohort. *Current oncology (Toronto, Ont)* 2018;25(6):384-392.
658 Rogado J, Sánchez-Torres JM, Romero-Laorden N*, et al.* Immune-related adverse events predict the therapeutic efficacy of anti-PD-1 antibodies in cancer patients. *Eur J Cancer* 2019;109:21-27.
659 Toi Y, Sugawara S, Kawashima Y*, et al.* Association of Immune-Related Adverse Events with Clinical Benefit in Patients with Advanced Non-Small-Cell Lung Cancer Treated with Nivolumab. *Oncologist* 2018;23(11):1358-1365.
660 Grangeon M, Tomasini P, Chaleat S*, et al.* Association Between Immune-related Adverse Events and Efficacy of Immune Checkpoint Inhibitors in Non-small-cell Lung Cancer. *Clin Lung Cancer* 2019;20(3):201-207.
661 Sato K, Akamatsu H, Murakami E*, et al.* Correlation between immune-related adverse events and efficacy in non-small cell lung cancer treated with nivolumab. *Lung Cancer* 2018;115:71-74.
662 Maher VE, Fernandes LL, Weinstock C*, et al.* Analysis of the Association Between Adverse Events and Outcome in Patients Receiving a Programmed Death Protein 1 or Programmed Death Ligand 1 Antibody. *J Clin Oncol* 2019;37(30):2730-2737.
663 Paz-Ares L, Spira A, Raben D*, et al.* Outcomes with durvalumab by tumour PD-L1 expression in unresectable, stage III non-small-cell lung cancer in the PACIFIC trial. *Annals of Oncology* 2020;31(6):798-806.
664 Domagala-Kulawik J, Raniszewska A. How to evaluate the immune status of lung cancer patients before immunotherapy. *Breathe* 2017;13(4):291-296.
665 Nelson CE, Mills LJ, McCurtain JL*, et al.* Reprogramming responsiveness to checkpoint blockade in dysfunctional CD8 T cells. *Proc Natl Acad Sci U S A* 2019;116(7):2640-2645.
666 vom Berg J, Kobold S. The need for speed: how PD1-blockade only works if T cells are properly activated. *Translational Cancer Research* 2017:S1018-S1021.
667 Zhao Y, Shao Q, Peng G. Exhaustion and senescence: two crucial dysfunctional states of T cells in the tumor microenvironment. *Cell Mol Immunol* 2020;17(1):27-35.
668 Kang D-H, Weaver MT, Park N-J*, et al.* Significant impairment in immune recovery after cancer treatment. *Nurs Res* 2009;58(2):105-114.
669 van der Most RG, Currie A, Robinson BW*, et al.* Cranking the immunologic engine with chemotherapy: using context to drive tumor antigen cross-presentation towards useful antitumor immunity. *Cancer Res* 2006;66(2):601-604.
670 Gardner RV. Long term hematopoietic damage after chemotherapy and cytokine. *Front Biosci* 1999;4:e47-57.
671 Galluzzi L, Humeau J, Buqué A*, et al.* Immunostimulation with chemotherapy in the era of immune checkpoint inhibitors. *Nature Reviews Clinical Oncology* 2020;17(12):725-741.
672 van den Eynden GGGM. Evaluation of Tumor Infiltrating Lymphocytes (TILs) in Lung Adenocarcinoma. Vol. 2020: Breast Cancer Research Foundation, 2020.
673 Hendry S, Salgado R, Gevaert T*, et al.* Assessing Tumor-Infiltrating Lymphocytes in Solid Tumors: A Practical Review for Pathologists and Proposal for a Standardized Method from the International Immuno-Oncology Biomarkers Working Group: Part 2: TILs in Melanoma, Gastrointestinal Tract Carcinomas, Non-Small Cell Lung Carcinoma and Mesothelioma, Endometrial and Ovarian Carcinomas, Squamous Cell Carcinoma of the Head and Neck, Genitourinary Carcinomas, and Primary Brain Tumors. *Adv Anat Pathol* 2017;24(6):311-335.

674 Amgad M, Stovgaard ES, Balslev E*, et al.* Report on computational assessment of Tumor Infiltrating Lymphocytes from the International Immuno-Oncology Biomarker Working Group. *npj Breast Cancer* 2020;6(1):16.
675 Gregg JP, Li T, Yoneda KY. Molecular testing strategies in non-small cell lung cancer: optimizing the diagnostic journey. *Translational Lung Cancer Research* 2019;8(3):286-301.
676 Bubendorf L, Lantuejoul S, de Langen AJ*, et al.* Nonsmall cell lung carcinoma: diagnostic difficulties in small biopsies and cytological specimens. *Number 2 in the Series "Pathology for the clinician" Edited by Peter Dorfmüller and Alberto Cavazza* 2017;26(144):170007.
677 Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol* 2018;15(2):81-94.
678 Ascierto PA, Bifulco C, Palmieri G*, et al.* Preanalytic Variables and Tissue Stewardship for Reliable Next-Generation Sequencing (NGS) Clinical Analysis. *J Mol Diagn* 2019;21(5):756-767.
679 Nagahashi M, Shimada Y, Ichikawa H*, et al.* Next generation sequencing-based gene panel tests for the management of solid tumors. *Cancer Sci* 2019;110(1):6-15.
680 June CH, Warshauer JT, Bluestone JA. Is autoimmunity the Achilles' heel of cancer immunotherapy? *Nat Med* 2017;23(5):540-547.
681 Naidoo J, Page DB, Li BT*, et al.* Toxicities of the anti-PD-1 and anti-PD-L1 immune checkpoint antibodies. *Ann Oncol* 2016;27(7):1362.
682 Eun Y, Kim IY, Sun JM*, et al.* Risk factors for immune-related adverse events associated with anti-PD-1 pembrolizumab. *Sci Rep* 2019;9(1):14039.
683 Das S, Johnson DB. Immune-related adverse events and anti-tumor efficacy of immune checkpoint inhibitors. *J Immunother Cancer* 2019;7(1):306.
684 Faje A, Reynolds K, Zubiri L*, et al.* Hypophysitis secondary to nivolumab and pembrolizumab is a clinical entity distinct from ipilimumab-associated hypophysitis. *Eur J Endocrinol* 2019;181(3):211-219.
685 Weinmann SC, Pisetsky DS. Mechanisms of immune-related adverse events during the treatment of cancer with immune checkpoint inhibitors. *Rheumatology (Oxford)* 2019;58(Suppl 7):vii59-vii67.
686 Connolly C, Bambhania K, Naidoo J. Immune-Related Adverse Events: A Case-Based Approach. *Front Oncol* 2019;9:530.
687 Almutairi AR, McBride A, Slack M*, et al.* Potential Immune-Related Adverse Events Associated With Monotherapy and Combination Therapy of Ipilimumab, Nivolumab, and Pembrolizumab for Advanced Melanoma: A Systematic Review and Meta-Analysis. *Front Oncol* 2020;10:91.
688 Duan J, Cui L, Zhao X*, et al.* Use of Immunotherapy With Programmed Cell Death 1 vs Programmed Cell Death Ligand 1 Inhibitors in Patients With Cancer: A Systematic Review and Meta-analysis. *JAMA Oncol* 2019.
689 Chaplin DD. Overview of the immune response. *The Journal of allergy and clinical immunology* 2010;125(2 Suppl 2):S3-S23.
690 Cha E, Klinger M, Hou Y*, et al.* Improved survival with T cell clonotype stability after anti-CTLA-4 treatment in cancer patients. *Sci Transl Med* 2014;6(238):238ra270.
691 Oh DY, Cham J, Zhang L*, et al.* Immune Toxicities Elicted by CTLA-4 Blockade in Cancer Patients Are Associated with Early Diversification of the T-cell Repertoire. *Cancer Res* 2017;77(6):1322-1330.
692 Vanderlugt CL, Miller SD. Epitope spreading in immune-mediated diseases: implications for immunotherapy. *Nat Rev Immunol* 2002;2(2):85-95.
693 Memarnejadian A, Meilleur CE, Shaler CR*, et al.* PD-1 Blockade Promotes Epitope Spreading in Anticancer CD8(+) T Cell Responses by Preventing Fratricidal Death of Subdominant Clones To Relieve Immunodomination. *J Immunol* 2017;199(9):3348-3359.

694 Schindler K, Harmankaya K, Kuk D*, et al.* Correlation of absolute and relative eosinophil counts with immune-related adverse events in melanoma patients treated with ipilimumab. *Journal of Clinical Oncology* 2014;32(15_suppl):9096-9096.
695 Tarhini AA, Zahoor H, Lin Y*, et al.* Baseline circulating IL-17 predicts toxicity while TGF-β1 and IL-10 are prognostic of relapse in ipilimumab neoadjuvant therapy of melanoma. *Journal for immunotherapy of cancer* 2015;3:39-39.
696 Ke W, Zhang L, Dai Y. The role of IL-6 in immunotherapy of non-small cell lung cancer (NSCLC) with immune-related adverse events (irAEs). *Thorac Cancer* 2020;11(4):835-839.
697 Nakamura Y. Biomarkers for Immune Checkpoint Inhibitor-Mediated Tumor Response and Adverse Events. *Front Med (Lausanne)* 2019;6:119.
698 Hopkins AM, Rowland A, Kichenadasse G*, et al.* Predicting response and toxicity to immune checkpoint inhibitors using routinely available blood and clinical markers. *Br J Cancer* 2017;117(7):913-920.
699 Meyers DE, Stukalin I, Vallerand IA*, et al.* The Lung Immune Prognostic Index Discriminates Survival Outcomes in Patients with Solid Tumors Treated with Immune Checkpoint Inhibitors. *Cancers (Basel)* 2019;11(11).
700 Minami S, Ihara S, Ikuta S*, et al.* Gustave Roussy Immune Score and Royal Marsden Hospital Prognostic Score Are Biomarkers of Immune-Checkpoint Inhibitor for Non-Small Cell Lung Cancer. *World J Oncol* 2019;10(2):90-100.
701 Pavan A, Calvetti L, Dal Maso A*, et al.* Circulating biomarkers and risk of immune-related adverse events (irAEs) in patients (pts) with advanced non-small cell lung cancer (aNSCLC) and metastatic melanoma (mMel). *Annals of Oncology* 2019;30:v45.
702 Zou Y, Xu L, Tang Q*, et al.* Cytology cell blocks from malignant pleural effusion are good candidates for PD-L1 detection in advanced NSCLC compared with matched histology samples. *BMC Cancer* 2020;20(1):344.
703 Bulusu VR, Jeffs Y, Jessop S*, et al.* Clinical characteristics of programmed death ligand-1 (PD-L1) positive non-small cell lung cancer patient cohort from a cancer unit – first year experience. *Lung Cancer* 2018;115:S40.
704 Li H, Tang Z, Zhu H*, et al.* Proteomic study of benign and malignant pleural effusion. *Journal of Cancer Research and Clinical Oncology* 2016;142(6):1191-1200.
705 Colonna M. All roads lead to CARD9. *Nature Immunology* 2007;8(6):554-555.
706 LeibundGut-Landmann S, Gross O, Robinson MJ*, et al.* Syk- and CARD9-dependent coupling of innate immunity to the induction of T helper cells that produce interleukin 17. *Nat Immunol* 2007;8(6):630-638.
707 Ma Y, Kang S, Wu X*, et al.* Up-regulated HMGB1 in the pleural effusion of non-small cell lung cancer (NSCLC) patients reduces the chemosensitivity of NSCLC cells. *Tumori* 2018;104(5):338-343.
708 Jakubowska K, Naumnik W, Niklińska W*, et al.* Clinical Significance of HMGB-1 and TGF-β Level in Serum and BALF of Advanced Non-Small Cell Lung Cancer. *Adv Exp Med Biol* 2015;852:49-58.
709 Shang G-H, Jia C-Q, Tian H*, et al.* Serum high mobility group box protein 1 as a clinical marker for non-small cell lung cancer. *Respiratory Medicine* 2009;103(12):1949-1953.
710 Feng L, Xue D, Chen E*, et al.* HMGB1 promotes the secretion of multiple cytokines and potentiates the osteogenic differentiation of mesenchymal stem cells through the Ras/MAPK signaling pathway. *Exp Ther Med* 2016;12(6):3941-3947.
711 Zhu J, Luo J, Li Y*, et al.* HMGB1 induces human non-small cell lung cancer cell motility by activating integrin αvβ3/FAK through TLR4/NF-κB signaling pathway. *Biochem Biophys Res Commun* 2016;480(4):522-527.
712 Wu X-J, Chen Y-Y, Guo W-W*, et al.* HMGB1 regulates SNAI1 during NSCLC metastasis, both directly, through transcriptional activation, and indirectly, in a RSF1-IT2-dependent manner. *Molecular oncology* 2020;14(6):1348-1364.

713 Pilzweger C, Holdenrieder S. Circulating HMGB1 and RAGE as Clinical Biomarkers in Malignant and Autoimmune Diseases. *Diagnostics (Basel)* 2015;5(2):219-253.
714 Liu Y, Zhuang GB, Zhou XZ. HMBG1 as a Driver of Inflammatory and Immune Processes in the Pathogenesis of Ocular Diseases. *J Ophthalmol* 2018;2018:5195290.
715 Wang H, Yang H, Czura CJ*, et al.* HMGB1 as a late mediator of lethal systemic inflammation. *Am J Respir Crit Care Med* 2001;164(10 Pt 1):1768-1773.
716 Andersson U, Tracey KJ. HMGB1 is a therapeutic target for sterile inflammation and infection. *Annu Rev Immunol* 2011;29:139-162.
717 Tang Z, Zang N, Fu Y*, et al.* HMGB1 mediates HAdV-7 infection-induced pulmonary inflammation in mice. *Biochem Biophys Res Commun* 2018;501(1):1-8.
718 Chen L, Zhang G, Sheng S*, et al.* Upregulation of soluble B7-H3 in NSCLC-derived malignant pleural effusion: A potential diagnostic biomarker correlated with NSCLC staging. *Clin Chim Acta* 2016;457:81-85.
719 Schneider T, Hoffmann H, Dienemann H*, et al.* Non-small cell lung cancer induces an immunosuppressive phenotype of dendritic cells in tumor microenvironment by upregulating B7-H3. *J Thorac Oncol* 2011;6(7):1162-1168.
720 Zhang G, Xu Y, Lu X*, et al.* Diagnosis value of serum B7-H3 expression in non-small cell lung cancer. *Lung Cancer* 2009;66(2):245-249.
721 Castellanos JR, Purvis IJ, Labak CM*, et al.* B7-H3 role in the immune landscape of cancer. *Am J Clin Exp Immunol* 2017;6(4):66-75.
722 Yim J, Koh J, Kim S*, et al.* Effects of B7-H3 expression on tumour-infiltrating immune cells and clinicopathological characteristics in non-small-cell lung cancer. *Eur J Cancer* 2020;133:74-85.
723 Yang S, Wei W, Zhao Q. B7-H3, a checkpoint molecule, as a target for cancer immunotherapy. *International Journal of Biological Sciences* 2020;16(11):1767-1773.
724 Lowery FJ, Yu D. Brain metastasis: Unique challenges and open opportunities. *Biochim Biophys Acta Rev Cancer* 2017;1867(1):49-57.
725 Griesinger F, Roeper J, Pöttgen C*, et al.* Brain metastases in ALK-positive NSCLC - time to adjust current treatment algorithms. *Oncotarget* 2018;9(80):35181-35194.
726 Han CH, Brastianos PK. Genetic Characterization of Brain Metastases in the Era of Targeted Therapy. *Frontiers in oncology* 2017;7:230-230.
727 Brantley EC, Guo L, Zhang C*, et al.* Nitric oxide-mediated tumoricidal activity of murine microglial cells. *Transl Oncol* 2010;3(6):380-388.
728 Pukrop T, Dehghani F, Chuang HN*, et al.* Microglia promote colonization of brain tissue by breast cancer cells in a Wnt-dependent way. *Glia* 2010;58(12):1477-1489.
729 Fitzgerald DP, Palmieri D, Hua E*, et al.* Reactive glia are recruited by highly proliferative brain metastases of breast cancer and promote tumor cell colonization. *Clin Exp Metastasis* 2008;25(7):799-810.
730 Mansour MSI, Seidal T, Mager U*, et al.* Determination of PD-L1 expression in effusions from mesothelioma by immuno-cytochemical staining. *Cancer Cytopathol* 2017;125(12):908-917.
731 Valmary-Degano S, Colpart P, Villeneuve L*, et al.* Immunohistochemical evaluation of two antibodies against PD-L1 and prognostic significance of PD-L1 expression in epithelioid peritoneal malignant mesothelioma: A RENAPE study. *Eur J Surg Oncol* 2017;43(10):1915-1923.
732 Watanabe T, Okuda K, Murase T*, et al.* Four immunohistochemical assays to measure the PD-L1 expression in malignant pleural mesothelioma. *Oncotarget* 2018;9(29):20769-20780.
733 Chapel DB, Stewart R, Furtado LV*, et al.* Tumor PD-L1 expression in malignant pleural and peritoneal mesothelioma by Dako PD-L1 22C3 pharmDx and Dako PD-L1 28-8 pharmDx assays. *Hum Pathol* 2019;87:11-17.
734 Cedrés S, Ponce-Aix S, Zugazagoitia J*, et al.* Analysis of expression of programmed cell death 1 ligand 1 (PD-L1) in malignant pleural mesothelioma (MPM). *PLoS One* 2015;10(3):e0121071.

735 Nguyen BH, Montgomery R, Fadia M*, et al.* PD-L1 expression associated with worse survival outcome in malignant pleural mesothelioma. *Asia Pac J Clin Oncol* 2018;14(1):69-73.
736 Chu GJ, van Zandwijk N, Rasko JEJ. The Immune Microenvironment in Mesothelioma: Mechanisms of Resistance to Immunotherapy. *Front Oncol* 2019;9:1366.
737 Dozier J, Zheng H, Adusumilli PS. Immunotherapy for malignant pleural mesothelioma: current status and future directions. *Transl Lung Cancer Res* 2017;6(3):315-324.
738 Shibaki R, Murakami S, Shinno Y*, et al.* Malignant pleural effusion as a predictor of the efficacy of anti-PD-1 antibody in patients with non-small cell lung cancer. *Thorac Cancer* 2019;10(4):815-822.
739 HTG. HTG EdgeSeq Precision Immuno-Oncology Panel. Vol. 2020, 2020.
740 Fumet JD, Truntzer C, Yarchoan M*, et al.* Tumour mutational burden as a biomarker for immunotherapy: Current data and emerging concepts. *Eur J Cancer* 2020;131:40-50.
741 Garassino M, Rodriguez-Abreu D, Gadgeel S*, et al.* OA04.06 Evaluation of TMB in KEYNOTE-189: Pembrolizumab Plus Chemotherapy vs Placebo Plus Chemotherapy for Nonsquamous NSCLC. *Journal of Thoracic Oncology* 2019;14(10):S216-S217.
742 Langer C, Gadgeel S, Borghaei H*, et al.* OA04.05 KEYNOTE-021: TMB and Outcomes for Carboplatin and Pemetrexed With or Without Pembrolizumab for Nonsquamous NSCLC. *Journal of Thoracic Oncology* 2019;14(10):S216.
743 ESMO. TTMB IS NOT ESTABLISHED AS A MARKER FOR PEMBROLIZUMAB EFFICACY IN NSCLC. Vol. 2020, 2019.
744 Ayers M, Lunceford J, Nebozhyn M*, et al.* IFN-gamma-related mRNA profile predicts clinical response to PD-1 blockade. *J Clin Invest* 2017;127(8):2930-2940.
745 Hwang S, Kwon AY, Jeong JY*, et al.* Immune gene signatures for predicting durable clinical benefit of anti-PD-1 immunotherapy in patients with non-small cell lung cancer. *Sci Rep* 2020;10(1):643.
746 Prendergast GC, Mondal A, Dey S*, et al.* Inflammatory Reprogramming with IDO1 Inhibitors: Turning Immunologically Unresponsive 'Cold' Tumors 'Hot'. *Trends Cancer* 2018;4(1):38-58.
747 Nuti M, Zizzari IG, Botticelli A*, et al.* The ambitious role of anti angiogenesis molecules: Turning a cold tumor into a hot one. *Cancer Treat Rev* 2018;70:41-46.
748 Serag A, Ion-Margineanu A, Qureshi H*, et al.* Translational AI and Deep Learning in Diagnostic Pathology. *Front Med (Lausanne)* 2019;6:185.
749 Hwang KS, Park KS, Lee SH*, et al.* Autonomous Machine Learning Modeling using a Task Ontology. *2018 Joint 10th International Conference on Soft Computing and Intelligent Systems (SCIS) and 19th International Symposium on Advanced Intelligent Systems (ISIS)*, 2018, p. 244-248.
750 Timp W, Timp G. Beyond mass spectrometry, the next step in proteomics. *Sci Adv* 2020;6(2):eaax8978.
751 Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature* 2003;422(6928):198-207.
752 Jannetto P. Liquid Chromatography Tandem Mass Spectrometry. Vol. 2020, 2015.
753 Pitt JJ. Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. *Clin Biochem Rev* 2009;30(1):19-34.
754 Zhu Y, Weiss T, Zhang Q*, et al.* High-throughput proteomic analysis of FFPE tissue samples facilitates tumor stratification. *Mol Oncol* 2019;13(11):2305-2328.
755 Gaffney EF, Riegman PH, Grizzle WE*, et al.* Factors that drive the increasing use of FFPE tissue in basic and translational cancer research. *Biotechnic & Histochemistry* 2018;93(5):373-386.
756 Liebler DC, Holzer TR, Haragan A*, et al.* Analysis of Immune Checkpoint Drug Targets and Tumor Proteotypes in Non-Small Cell Lung Cancer. *Scientific Reports* 2020;10(1):9805.
757 Cabanero M, Tsao MS. Circulating tumour DNA in EGFR-mutant non-small-cell lung cancer. *Current oncology (Toronto, Ont)* 2018;25(Suppl 1):S38-S44.

758 He Y, Shi J, Schmidt B*, et al.* Circulating Tumor Cells as a Biomarker to Assist Molecular Diagnosis for Early Stage Non-Small Cell Lung Cancer. *Cancer Manag Res* 2020;12:841-854.
759 Chen YL, Huang WC, Lin FM*, et al.* Novel circulating tumor cell-based blood test for the assessment of PD-L1 protein expression in treatment-naive, newly diagnosed patients with non-small cell lung cancer. *Cancer Immunol Immunother* 2019;68(7):1087-1094.
760 Guibert N, Delaunay M, Lusque A*, et al.* PD-L1 expression in circulating tumor cells of advanced non-small cell lung cancer patients treated with nivolumab. *Lung Cancer* 2018;120:108-112.
761 Koh Y, Yagi S, Akamatsu H*, et al.* Heterogeneous Expression of Programmed Death Receptor-ligand 1 on Circulating Tumor Cells in Patients With Lung Cancer. *Clin Lung Cancer* 2019;20(4):270-277 e271.
762 Cheng Y, Wang T, Lv X*, et al.* Detection of PD-L1 Expression and Its Clinical Significance in Circulating Tumor Cells from Patients with Non-Small-Cell Lung Cancer. *Cancer Manag Res* 2020;12:2069-2078.
763 Ilie M, Szafer-Glusman E, Hofman V*, et al.* Detection of PD-L1 in circulating tumor cells and white blood cells from patients with advanced non-small-cell lung cancer. *Ann Oncol* 2018;29(1):193-199.
764 Nicolazzo C, Raimondi C, Mancini M*, et al.* Monitoring PD-L1 positive circulating tumor cells in non-small cell lung cancer patients treated with the PD-1 inhibitor Nivolumab. *Sci Rep* 2016;6:31726.
765 Wang Y, Kim TH, Fouladdel S*, et al.* PD-L1 Expression in Circulating Tumor Cells Increases during Radio(chemo)therapy and Indicates Poor Prognosis in Non-small Cell Lung Cancer. *Scientific Reports* 2019;9(1):566.
766 Boffa DJ, Graf RP, Salazar MC*, et al.* Cellular Expression of PD-L1 in the Peripheral Blood of Lung Cancer Patients is Associated with Worse Survival. *Cancer Epidemiol Biomarkers Prev* 2017;26(7):1139-1145.
767 Janning M, Kobus F, Babayan A*, et al.* Determination of PD-L1 Expression in Circulating Tumor Cells of NSCLC Patients and Correlation with Response to PD-1/PD-L1 Inhibitors. *Cancers (Basel)* 2019;11(6).
768 Khattak MA, Reid A, Freeman J*, et al.* PD-L1 Expression on Circulating Tumor Cells May Be Predictive of Response to Pembrolizumab in Advanced Melanoma: Results from a Pilot Study. *Oncologist* 2020;25(3):e520-e527.
769 Adams DL, Adams DK, He J*, et al.* Sequential Tracking of PD-L1 Expression and RAD50 Induction in Circulating Tumor and Stromal Cells of Lung Cancer Patients Undergoing Radiotherapy. *Clinical Cancer Research* 2017;23(19):5948-5958.
770 Schehr JL, Schultz ZD, Warrick JW*, et al.* High Specificity in Circulating Tumor Cell Identification Is Required for Accurate Evaluation of Programmed Death-Ligand 1. *PLoS One* 2016;11(7):e0159397.
771 Kloten V, Lampignano R, Krahn T*, et al.* Circulating Tumor Cell PD-L1 Expression as Biomarker for Therapeutic Efficacy of Immune Checkpoint Inhibition in NSCLC. *Cells* 2019;8(8).
772 Koyyala VPB, Saini SD, Doval DC*, et al.* Correlation of circulating tumour cells with PET-CT in metastatic breast cancer. *Annals of Oncology* 2019;30:ix16-ix17.
773 Giuliano M, Giordano A, Jackson S*, et al.* Circulating tumor cells as early predictors of metastatic spread in breast cancer patients with limited metastatic dissemination. *Breast Cancer Res* 2014;16(5):440.
774 Tamminga M, de Wit S, Hiltermann TJN*, et al.* Circulating tumor cells in advanced non-small cell lung cancer patients are associated with worse tumor response to checkpoint inhibitors. *J Immunother Cancer* 2019;7(1):173.
775 Tang Y, Zhang P, Wang Y*, et al.* The Biogenesis, Biology, and Clinical Significance of Exosomal PD-L1 in Cancer. *Front Immunol* 2020;11:604.

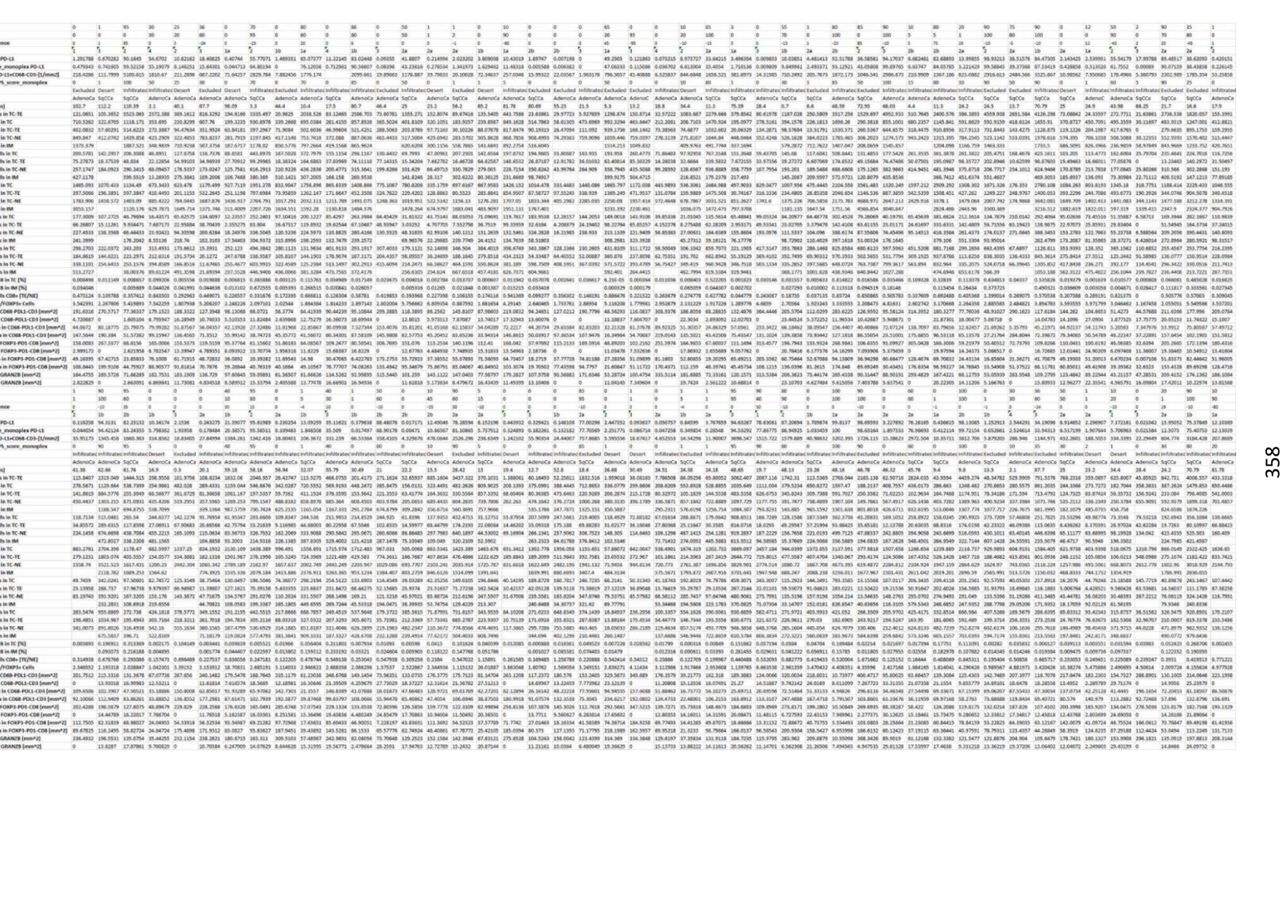
776 Poggio M, Hu T, Pai CC*, et al.* Suppression of Exosomal PD-L1 Induces Systemic Anti-tumor Immunity and Memory. *Cell* 2019;177(2):414-427.e413.
777 Yang Y, Li C-W, Chan L-C*, et al.* Exosomal PD-L1 harbors active defense function to suppress T cell killing of breast cancer cells and promote tumor growth. *Cell Res* 2018;28(8):862-864.
778 Chen G, Huang AC, Zhang W*, et al.* Exosomal PD-L1 contributes to immunosuppression and is associated with anti-PD-1 response. *Nature* 2018;560(7718):382-386.
779 Theodoraki MN, Yerneni SS, Hoffmann TK*, et al.* Clinical Significance of PD-L1(+) Exosomes in Plasma of Head and Neck Cancer Patients. *Clin Cancer Res* 2018;24(4):896-905.
780 Ricklefs FL, Alayo Q, Krenzlin H*, et al.* Immune evasion mediated by PD-L1 on glioblastoma-derived extracellular vesicles. *Science advances* 2018;4(3):eaar2766-eaar2766.
781 Lawler SE, Nowicki MO, Ricklefs FL*, et al.* Immune Escape Mediated by Exosomal PD-L1 in Cancer. *Advanced Biosystems*;n/a(n/a):2000017.
782 Cordonnier M, Nardin C, Chanteloup G*, et al.* Tracking the evolution of circulating exosomal-PD-L1 to monitor melanoma patients. *J Extracell Vesicles* 2020;9(1):1710899.
783 Li C, Li C, Zhi C*, et al.* Clinical significance of PD-L1 expression in serum-derived exosomes in NSCLC patients. *J Transl Med* 2019;17(1):355.
784 Del Re M, Marconcini R, Pasquini G*, et al.* PD-L1 mRNA expression in plasma-derived exosomes is associated with response to anti-PD-1 antibodies in melanoma and NSCLC. *Br J Cancer* 2018;118(6):820-824.
785 Zhou J, Mahoney KM, Giobbie-Hurder A*, et al.* Soluble PD-L1 as a Biomarker in Malignant Melanoma Treated with Checkpoint Blockade. *Cancer Immunol Res* 2017;5(6):480-492.
786 Yoshida J, Ishikawa T, Doi T*, et al.* Clinical significance of soluble forms of immune checkpoint molecules in advanced esophageal cancer. *Med Oncol* 2019;36(7):60.
787 Bonomi M, Ahmed T, Addo S*, et al.* Circulating immune biomarkers as predictors of the response to pembrolizumab and weekly low dose carboplatin and paclitaxel in NSCLC and poor PS: An interim analysis. *Oncology letters* 2019;17(1):1349-1356.
788 Jin J, Si J, Liu Y*, et al.* Elevated serum soluble programmed cell death ligand 1 concentration as a potential marker for poor prognosis in small cell lung cancer patients with chemotherapy. *Respiratory Research* 2018;19(1):197.
789 Hellmann MD, Nabet BY, Rizvi H*, et al.* Circulating Tumor DNA Analysis to Assess Risk of Progression after Long-term Response to PD-(L)1 Blockade in NSCLC. *Clin Cancer Res* 2020;26(12):2849-2858.
790 Goldberg SB, Narayan A, Kole AJ*, et al.* Early Assessment of Lung Cancer Immunotherapy Response via Circulating Tumor DNA. *Clin Cancer Res* 2018;24(8):1872-1880.
791 Anagnostou V, Forde PM, White JR*, et al.* Dynamics of Tumor and Immune Responses during Immune Checkpoint Blockade in Non–Small Cell Lung Cancer. *Cancer Research* 2019;79(6):1214-1225.
792 Moding EJ, Liu Y, Nabet BY*, et al.* Circulating tumor DNA dynamics predict benefit from consolidation immunotherapy in locally advanced non-small-cell lung cancer. *Nature Cancer* 2020;1(2):176-183.
793 Gandara DR, Paul SM, Kowanetz M*, et al.* Blood-based tumor mutational burden as a predictor of clinical benefit in non-small-cell lung cancer patients treated with atezolizumab. *Nat Med* 2018;24(9):1441-1448.
794 Wang Z, Duan J, Cai S*, et al.* Assessment of Blood Tumor Mutational Burden as a Potential Biomarker for Immunotherapy in Patients With Non–Small Cell Lung Cancer With Use of a Next-Generation Sequencing Cancer Gene Panel. *JAMA Oncology* 2019;5(5):696-702.
795 Peters S, Cho BC, Reinmuth N*, et al.* Abstract CT074: Tumor mutational burden (TMB) as a biomarker of survival in metastatic non-small cell lung cancer (mNSCLC): Blood and tissue TMB analysis from MYSTIC, a Phase III study of first-line durvalumab ± tremelimumab vs chemotherapy. *Cancer Research* 2019;79(13 Supplement):CT074-CT074.

796 Snyder A, Morrissey MP, Hellmann MD. Use of Circulating Tumor DNA for Cancer Immunotherapy. *Clinical Cancer Research* 2019:clincanres.2688.2019.
797 Georgiadis A, Durham JN, Keefer LA*, et al.* Noninvasive Detection of Microsatellite Instability and High Tumor Mutation Burden in Cancer Patients Treated with PD-1 Blockade. *Clinical Cancer Research* 2019.
798 Willis J, Lefterova MI, Artyomenko A*, et al.* Validation of Microsatellite Instability Detection Using a Comprehensive Plasma-Based Genotyping Panel. *Clinical Cancer Research* 2019.
799 Gargiulo P, Della Pepa C, Berardi S*, et al.* Tumor genotype and immune microenvironment in POLE-ultramutated and MSI-hypermutated Endometrial Cancers: New candidates for checkpoint blockade immunotherapy? *Cancer Treat Rev* 2016;48:61-68.
800 Le DT, Uram JN, Wang H*, et al.* PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med* 2015;372(26):2509-2520.
801 Nojadeh JN, Behrouz Sharif S, Sakhinia E. Microsatellite instability in colorectal cancer. *EXCLI J* 2018;17:159-168.
802 Oliveira AF, Bretes L, Furtado I. Review of PD-1/PD-L1 Inhibitors in Metastatic dMMR/MSI-H Colorectal Cancer. *Front Oncol* 2019;9:396.
803 Woenckhaus M, Stoehr R, Dietmaier W*, et al.* Microsatellite instability at chromosome 8p in non-small cell lung cancer is associated with lymph node metastasis and squamous differentiation. *Int J Oncol* 2003;23(5):1357-1363.
804 Sang Y, Mei J, Yu W*, et al.* The relationship of microsatellite instability with infiltration of T lymphocytes in tissues of non-small cell lung cancer. *Chinese Journal of Cancer Biotherapy* 2016;23(3):371-377.
805 Takamochi K, Takahashi F, Suehara Y*, et al.* DNA mismatch repair deficiency in surgically resected lung adenocarcinoma: Microsatellite instability analysis using the Promega panel. *Lung Cancer* 2017;110:26-31.
806 De Marchi P, Berardinelli GN, Cavagna R*, et al.* EP1.04-11 Frequency of Microsatellite Instability (MSI) in Brazilian TKI Non-Treatable Non-Small Cell Lung Cancer (NSCLC) Patients. *Journal of Thoracic Oncology* 2019;14(10):S973.
807 Panda AC, Grammatikakis I, Munk R*, et al.* Emerging roles and context of circular RNAs. *WIREs RNA* 2017;8(2):e1386.
808 Tian M, Chen R, Li T*, et al.* Reduced expression of circRNA hsa_circ_0003159 in gastric cancer and its clinical significance. *J Clin Lab Anal* 2018;32(3).
809 Wang S, Zhang Y, Cai Q*, et al.* Circular RNA FOXP1 promotes tumor progression and Warburg effect in gallbladder cancer by regulating PKLR expression. *Molecular Cancer* 2019;18(1):145.
810 Ding L, Zhao Y, Dang S*, et al.* Circular RNA circ-DONSON facilitates gastric cancer growth and invasion via NURF complex dependent activation of transcription factor SOX4. *Molecular Cancer* 2019;18(1):45.
811 Su M, Xiao Y, Ma J*, et al.* Circular RNAs in Cancer: emerging functions in hallmarks, stemness, resistance and roles as potential biomarkers. *Molecular Cancer* 2019;18(1):90.
812 Ma Z, Shuai Y, Gao X*, et al.* Circular RNAs in the tumour microenvironment. *Molecular Cancer* 2020;19(1):8.
813 Zhou Y, Zheng X, Xu B*, et al.* Circular RNA hsa_circ_0004015 regulates the proliferation, invasion, and TKI drug resistance of non-small cell lung cancer by miR-1183/PDPK1 signaling pathway. *Biochem Biophys Res Commun* 2019;508(2):527-535.
814 de Fraipont F, Gazzeri S, Cho WC*, et al.* Circular RNAs and RNA Splice Variants as Biomarkers for Prognosis and Therapeutic Response in the Liquid Biopsies of Lung Cancer Patients. *Front Genet* 2019;10:390-390.
815 Song H, Liu Q, Liao Q. Circular RNA and tumor microenvironment. *Cancer cell international* 2020;20:211-211.
816 Schwarzenbach H, Nishida N, Calin GA*, et al.* Clinical relevance of circulating cell-free microRNAs in cancer. *Nature Reviews Clinical Oncology* 2014;11(3):145-156.

817 Zhang L, Lin J, Ye Y*, et al.* Serum MicroRNA-150 Predicts Prognosis for Early-Stage Non-Small Cell Lung Cancer and Promotes Tumor Cell Proliferation by Targeting Tumor Suppressor Gene SRCIN1. *Clinical Pharmacology & Therapeutics* 2018;103(6):1061-1073.
818 Boeri M, Milione M, Proto C*, et al.* Circulating miRNAs and PD-L1 Tumor Expression Are Associated with Survival in Advanced NSCLC Patients Treated with Immunotherapy: a Prospective Study. *Clin Cancer Res* 2019;25(7):2166-2173.
819 Fan J, Yin Z, Xu J*, et al.* Circulating microRNAs predict the response to anti-PD-1 therapy in non-small cell lung cancer. *Genomics* 2020;112(2):2063-2071.
820 Mezquita L, Auclin E, Ferrara R*, et al.* Association of the Lung Immune Prognostic Index With Immune Checkpoint Inhibitor Outcomes in Patients With Advanced Non-Small Cell Lung Cancer. *JAMA Oncol* 2018;4(3):351-357.
821 Jin J, Yang L, Liu D*, et al.* Association of the neutrophil to lymphocyte ratio and clinical outcomes in patients with lung cancer receiving immunotherapy: a meta-analysis. *BMJ Open* 2020;10(6):e035031-e035031.
822 Matsuki T, Okamoto I, Fushimi C*, et al.* Hematological predictive markers for recurrent or metastatic squamous cell carcinomas of the head and neck treated with nivolumab: A multicenter study of 88 patients. *Cancer Medicine* 2020;9(14):5015-5024.
823 Yamazaki N, Kiyohara Y, Uhara H*, et al.* Cytokine biomarkers to predict antitumor responses to nivolumab suggested in a phase 2 study for advanced melanoma. *Cancer Sci* 2017;108(5):1022-1031.
824 Sanmamed MF, Perez-Gracia JL, Schalper KA*, et al.* Changes in serum interleukin-8 (IL-8) levels reflect and predict response to anti-PD-1 treatment in melanoma and non-small-cell lung cancer patients. *Ann Oncol* 2017;28(8):1988-1995.
825 Borghaei H, Brahmer JR, Horn L*, et al.* Nivolumab (nivo) vs docetaxel (doc) in patients (pts) with advanced NSCLC: CheckMate 017/057 2-y update and exploratory cytokine profile analyses. *Journal of Clinical Oncology* 2016;34(15_suppl):9025-9025.
826 Heitzer E, Haque IS, Roberts CES*, et al.* Current and future perspectives of liquid biopsies in genomics-driven oncology. *Nat Rev Genet* 2019;20(2):71-88.
827 Bellesoeur A, Torossian N, Amigorena S*, et al.* Advances in theranostic biomarkers for tumor immunotherapy. *Current Opinion in Chemical Biology* 2020;56:79-90.
828 Wu Q, Jiang L, Li S-c*, et al.* Small molecule inhibitors targeting the PD-1/PD-L1 signaling pathway. *Acta Pharmacologica Sinica* 2020.
829 Qin S, Xu L, Yi M*, et al.* Novel immune checkpoint targets: moving beyond PD-1 and CTLA-4. *Molecular Cancer* 2019;18(1):155.
830 Nejadmoghaddam M-R, Minai-Tehrani A, Ghahremanzadeh R*, et al.* Antibody-Drug Conjugates: Possibilities and Challenges. *Avicenna J Med Biotechnol* 2019;11(1):3-23.
831 Liu Y, Dong Z, Jiang T*, et al.* Heterogeneity of PD-L1 Expression Among the Different Histological Components and Metastatic Lymph Nodes in Patients With Resected Lung Adenosquamous Carcinoma. *Clin Lung Cancer* 2018;19(4):e421-e430.
832 Bassanelli M, Sioletic S, Martini M*, et al.* Heterogeneity of PD-L1 Expression and Relationship with Biology of NSCLC. *Anticancer Res* 2018;38(7):3789-3796.
833 Ai L, Xu A, Xu J. Roles of PD-1/PD-L1 Pathway: Signaling, Cancer, and Beyond. *Adv Exp Med Biol* 2020;1248:33-59.
834 von Pawel J, Bordoni R, Satouchi M*, et al.* Long-term survival in patients with advanced non-small-cell lung cancer treated with atezolizumab versus docetaxel: Results from the randomised phase III OAK study. *Eur J Cancer* 2019;107:124-132.
835 Chan TA, Yarchoan M, Jaffee E*, et al.* Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. *Annals of oncology : official journal of the European Society for Medical Oncology* 2019;30(1):44-56.

836 Stewart EL, Tan SZ, Liu G*, et al.* Known and putative mechanisms of resistance to EGFR targeted therapies in NSCLC patients with EGFR mutations-a review. *Transl Lung Cancer Res* 2015;4(1):67-81.
837 Murtuza A, Bulbul A, Shen JP*, et al.* Novel Third-Generation EGFR Tyrosine Kinase Inhibitors and Strategies to Overcome Therapeutic Resistance in Lung Cancer. *Cancer Res* 2019;79(4):689-698.
838 Ramalingam SS, Vansteenkiste J, Planchard D*, et al.* Overall Survival with Osimertinib in Untreated, EGFR-Mutated Advanced NSCLC. *New England Journal of Medicine* 2019;382(1):41-50.
839 Shi LZ, Fu T, Guan B*, et al.* Interdependent IL-7 and IFN-γ signalling in T-cell controls tumour eradication by combined α-CTLA-4+α-PD-1 therapy. *Nature Communications* 2016;7(1):12335.
840 Yang Y, Jin G, Pang Y*, et al.* Comparative Efficacy and Safety of Nivolumab and Nivolumab Plus Ipilimumab in Advanced Cancer: A Systematic Review and Meta-Analysis. *Frontiers in Pharmacology* 2020;11(40).
841 Puri S, Shafique M. Combination checkpoint inhibitors for treatment of non-small-cell lung cancer: an update on dual anti-CTLA-4 and anti-PD-1/PD-L1 therapies. *Drugs Context* 2020;9:2019-2019-2012.
842 Yarchoan M, Albacker LA, Hopkins AC*, et al.* PD-L1 expression and tumor mutational burden are independent biomarkers in most cancers. *JCI Insight* 2019;4(6).
843 Schalper KA, Carvajal-Hausdorf D, McLaughlin J*, et al.* Differential Expression and Significance of PD-L1, IDO-1, and B7-H4 in Human Lung Cancer. *Clin Cancer Res* 2017;23(2):370-378.
844 Zhang ML, Kem M, Mooradian MJ*, et al.* Differential expression of PD-L1 and IDO1 in association with the immune microenvironment in resected lung adenocarcinomas. *Mod Pathol* 2019;32(4):511-523.
845 Kozuma Y, Takada K, Toyokawa G*, et al.* Indoleamine 2,3-dioxygenase 1 and programmed cell death-ligand 1 co-expression correlates with aggressive features in lung adenocarcinoma. *Eur J Cancer* 2018;101:20-29.
846 Takada K, Kohashi K, Shimokawa M*, et al.* Co-expression of IDO1 and PD-L1 in lung squamous cell carcinoma: Potential targets of novel combination therapy. *Lung Cancer* 2019;128:26-32.
847 Sakuishi K, Apetoh L, Sullivan JM*, et al.* Targeting Tim-3 and PD-1 pathways to reverse T cell exhaustion and restore anti-tumor immunity. *J Exp Med* 2010;207(10):2187-2194.
848 Kwapisz D. The first liquid biopsy test approved. Is it a new era of mutation testing for non-small cell lung cancer? *Ann Transl Med* 2017;5(3):46.
849 Calder PC. Immunological Parameters: What Do They Mean? *The Journal of Nutrition* 2007;137(3):773S-780S.
850 Claus M, Dychus N, Ebel M*, et al.* Measuring the immune system: a comprehensive approach for the analysis of immune functions in humans. *Arch Toxicol* 2016;90(10):2481-2495.
851 Karasaki T, Nagayama K, Kuwano H*, et al.* An Immunogram for the Cancer-Immunity Cycle: Towards Personalized Immunotherapy of Lung Cancer. *J Thorac Oncol* 2017;12(5):791-803.
852 Dimitrov S, Gouttefangeas C, Besedovsky L*, et al.* Activated integrins identify functional antigen-specific CD8+ve T cells within minutes after antigen stimulation. *Proceedings of the National Academy of Sciences* 2018;115(24):E5536-E5545.

Appendix 1- Raw data values provided from Definiens from multiplex IHC image analysis for 92 cases from the LLP cohort



Appendix 2– Factors used in multivariate analysis modelling in Chapter 7.

Primary TPS	TC Area in FOXP3-PD1-CD8 [mm^2]
Secondary TPS	IM Area in FOXP3-PD1-CD8 [mm^2]
Gender	TC-TE Area in FOXP3-PD1-CD8 [mm^2]
Age	TC-NE Area in FOXP3-PD1-CD8 [mm^2]
T Stage	TC Area in GRANZB [mm^2]
monoplex PD-L1	IM Area in GRANZB [mm^2]
TumorCore_monoplex PD-L1	TC-TE Area in CD68-PDL1-CD3 [mm^2]
Density_PD-L1+CD68-CD3-[1/mm2]	TC-NE Area in CD68-PDL1-CD3 [mm^2]
manual_TPS_score_monoplex	
H&E	
Histology	
Status (1= dead)	
FU (Months)	
PDL1+ Cells in TC-TE	
CD8+ Cells in TC	
CD8+ Cells in TC-TE	
CD8+ Cells in TC-NE	
CD8+ Cells in IM	
FoxP3+ Cells in TC	
FoxP3+ Cells in TC-TE	
FoxP3+ Cells in TC-NE	
FoxP3+ Cells in IM	
CD3+ Cells in TC	
CD3+ Cells in TC-TE	
CD3+ Cells in TC-NE	
CD3+ Cells in IM	
PD-1+ Cells in TC	
PD-1+ Cells in TC-TE	
PD-1+ Cells in TC-NE	
PD-1+ Cells in IM	
CD68+ Cells in TC	
CD68+ Cells in TC-TE	
CD68+ Cells in TC-NE	
CD68+ Cells in IM	
Granzyme B in TC [%]	
Ratio: GranB:CD8 TC	
Granzyme B in IM [%]	
Ratio T-Cells CD8+ [TE/NE]	
Ratio CD8+/FOXP3+ Cells	
TC Area in CD68-PDL1-CD3 [mm^2]	
IM Area in CD68-PDL1-CD3 [mm^2]	