Using primary tumour transcriptomes to better understand immune responses to melanoma

Joanna Maria Poźniak

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
School of Medicine
Faculty of Medicine and Health

January 2019

The candidate confirms that the work submitted is his/her own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Whole thesis contains work from the following jointly authored publication:

Genetic and environmental determinants of immune response to cutaneous melanoma. Joanna Poźniak, Jérémie Nsengimana, Jonathan P. Laye, Sally J. O'Shea, Joey Mark S. Diaz, Alastair P. Droop, Anastasia Filia, Mark Harland, John R. Davies, Tracey Mell, Juliette A. Randerson-Moor, Sathya Muralidhar, Sabrina Hogan, Sandra N. Freiberger-Rupp, Mitchell P. Levesque, Graham P. Cook, David T. Bishop, Julia Newton-Bishop, Cancer Research 2018; in press

I performed all bioinformatic and statistical analyses, literature search and wrote the paper; I, Jonathan P. Laye, and Sathya Muralidhar performed immunohistochemical staining and scoring; Sally J. O'Shea performed histopathological analyses; Joey Mark S. Diaz, Alastair P. Droop, Anastasia Filia, Mark Harland, and David Timothy Bishop contributed to CNA data generation; Jonathan P. Laye, Juliette A. Randerson-Moor, and Tracey Mell designed the microarray experiment; Mitchell P. Levesque provided melanoma cell lines data from the University of Zürich Hospital, Switzerland; Sandra N. Freiberger (University of Zürich) generated melanoma cell lines data, Sabrina Hogan (University of Zürich) analysed melanoma cell lines data; Graham P. Cook provided immunological insight; John R. Davies provided statistical input; Jérémie Nsengimana, David Timothy Bishop and Julia Newton-Bishop developed the concept and supervised the study; Jérémie Nsengimana, Jonathan P. Laye, and Graham P. Cook, and Julia Newton-Bishop provided guidance in writing of manuscript.

Acknowledgements

This project was funded by the European Union's Horizon 2020 research and innovation programme Marie Skłodowska Curie Actions-ETN MELGEN, under grant agreement No 641458.

I would like to express my sincere gratitude to my supervisor Prof. Julia Newton-Bishop for her guidance, support, patience, and giving me the opportunity of undertaking the PhD studies while being a mother. She has set an example of excellence as a researcher and a woman in science.

I would like to thank Prof. David Timothy Bishop and Dr Jérémie Nsengimana, my supervisors for the scientific directions and remarkable training in statistics and epidemiology.

I am very grateful to past and present members of the Leeds Melanoma Research Group who had previously recorded clinical data, sectioned the FFPE tumour blocks, sampled the genetic material and generated data on gene expression, tumour mutation status and copy number alteration. I thank the patients and their families for consenting in participation in this study.

I would also like to thank:

Dr Jonathan P. Laye for training and assistance in immunohistochemical staining and its scoring; as well as for his support in scientific writing.

Dr Juliette Randerson-Moor for constant personal support and assistance in formatting my thesis.

Prof. Graham P. Cook for his feedback in immunological knowledge and his overall motivation and support.

Prof. Ulf Klein for providing knowledge and insight into NF-κB signalling.

Dr Alastair Droop for initial assistance in coding and analysis of the protein-protein interaction networks.

My fellow office mates Rohit Thakur and Joey Mark Santiago Diaz for their support and encouragement, and especially Sathya Muralidhar for deep scientific and philosophic conversations and “breakdown” sessions.

The Marie Skłodowska-Curie “MELGEN” European Training Network for valuable training and opportunity for collaborations.

Last but not least I would like to thank my family for their continuous care through my whole PhD journey. Most importantly to Victor and Hugo for their love and keeping my mind and spirit healthy day by day.

Abstract

Strong immune responses to melanoma predict improved survival and better responses to immunotherapies. However the host-tumour interactions are not yet fully elucidated, especially for early stage tumours and increasing understanding of this interaction was the aim of this thesis. The analyses were carried out using 703 primary melanoma transcriptomes generated from the Leeds Melanoma Cohort (LMC), detailed clinico-histopathological, and additional genomic data. Bioinformatics was applied to infer the immune environment within the tumours.

Using a modified "Immunome Compendium" developed by Angelova et al. [1], consensus clustering was applied to identify three immune subgroups: low, intermediate and high, associated with survival. Differentially expressed genes between the immune subgroups in the LMC were identified and analysed in the context of networks and pathways using Reactome FIViz. The oncogene *MYC* was identified as a nodal gene for the Low and *NFKB1* for the High Immune Subgroup. The expression of both genes showed significant association with protein scores from immunohistochemistry and with copy number alterations. The genes from NF-κB and IFN-γ pathways were more frequently deleted and *MYC* was amplified in the Low Immune Subgroup. These observations were considered as immune evasion mechanisms in primary melanoma.

Furthermore, it was observed that *MYC* expression *was* negatively correlated with many antigen processing and presentation genes (*HLA-B , HLA-C*, *B2M*, *TAP1 and ERAP1*) in the LMC and patient-derived melanoma cell lines and with HLA-B at a protein level in the LMC. My hypothesis was that MYC drives immunosuppression at least in part by reducing antigen presentation by the histocompatibility complex.

Finally, it was demonstrated that smoking was detrimental for melanoma specific survival overall, with the strongest effect for patients with tumours classified in the High Immune Subgroup suggesting a specific effect on tumour progression and it is a clear demonstration of an environmental modifier of survival.

Abbreviations

Adoptive Cell Therapy	ACT
American Joint Cancer Committee on Cancer	AJCC
Antigen Presenting Cell	APC
ATP Binding Cassette Subfamily B Member Transporter	TAP
B Cell Receptor	BCR
Beta-2-Microglobulin	B2M
cDNA-mediated Annealing, Selection, extension and Ligation	DASL
Cell type Identification By Estimating Relative Subsets Of RNA Transcripts	CIBERSORT
Chemokine (C-X-C motif) ligand 3	CXCL3
Chimeric Antigen Receptor	CAR
Cluster of Differentiation	CD
Confidence Interval	CI
Consensus Immunome Cluster	CIC
Conserved Helix-Loop-Helix Ubiquitous Kinase	CHUK
Copy Number Alteration	CNA
Cumulative Density Function	CDF
Cytotoxic T-Lymphocyte-Associated Protein 4	CTLA-4
Database for Annotation, Visualization and Integrated Discovery	DAVID
Dendritic Cell	DC
Defective Ribosomal Products	DRiPs
Electronic Nicotine Delivery System	ENDS
Endoplasmic Reticulum	ER
Endoplasmic Reticulum Aminopeptidase	ERAAP
Endoplasmic Reticulum Aminopeptidase 1	ERAP1
European Genome-phenome Archive	EGA
False Discovery Rate	FDR
Fluorescence-Activated Cell Sorting	FACS
Forkhead Box P3	FOXP3
Formalin-Fixed Paraffin Embedded	FFPE

Functional Interaction	FI
Gene Ontology	GO
Gene Set Enrichment Analysis	GSEA
Guanosine Diphosphate	GDP
Guanosine Triphosphate	GTP
Hazard Ratio	HR
Hematoxylin and Eosin	H&E
Human Genome Organisation	HUGO
Human Leukocyte Antigen	HLA
immunohistochemistry	IHC
Inhibitor of nuclear factor kappa-B kinase subunit beta	IKBKB
Interleukin 6	IL6
Interleukin 6 Receptor	IL6R
Interleukin 6 Signal Transducer	IL6ST
Interleukin-1 Receptor-Associated Kinase-like 1	IRAK1
Interleukin-1 Receptor-Associated Kinase-like 2	IRAK2
kappa-B kinase subunit gamma	IKBKG
Kyoto Encyclopedia of Genes and Genomes	KEGG
Leeds Melanoma Cohort	LMC
Major histocompatibility complex class I, B	HLA-B
Major histocompatibility complex class I, C	HLA-C
Melanoma Associated Antigens	MAGE
Melanoma Specific Survival	MSS
Minor Allele Frequency	MAF
Mitogen Activated Protein Kinase	MAPK
Mitogen-Activated Protein Kinase kinase kinase 14	MAP3K14
Mitogen-Activated Protein Kinase kinase kinase 7	MAP3K7
Multiplex Ion Beam Imaging	MIBI
Myeloid Derived Suppressor Cell	MDSC
Myeloid differentiation primary response protein MyD88	MYD88
Natural Killer	NK

Next Generation Sequencing	NGS
NF-kappa-B inhibitor alpha	NFKBIA
Nuclear Factor Kappa B Subunit 2	NFKB2
Nuclear factor NF-kappa-B	NFKB
Number of samples	N
Overall Survival	OS
Overrepresentation Analysis	ORA
Pattern Recognition Receptors	PRRs
Phosphatidylinositide 3 Kinases	PI3K
plasmacytoid Dendritic Cell	pDC
Polo like kinase 1	PLK1
Programmed Death 1 Receptor	PD-1
Programmed Death Ligand 1	PD-L1
Protein Phosphatase 2 Scaffold Subunit Alpha	PPP2R1A
Protein-Protein Interaction	PPI
Proto-oncogene cRel	REL
Quality Control	QC
quantitative Polymerase Chain Reaction	qPCR
Receptor Tyrosine Kinase	RTK
Regulatory T cell	Treg
Senescence-Associated Secretory Phenotype	SASP
Signal Transducer and Activator of Transcription 1-alpha/beta	STAT1
Singular Value Decomposition	SVD
Standard Error	SE
Stimulator of Interferon Genes	STING
Support Vector Regression	SVR
T Cell Receptor	TCR
T helper cell	Th
The Cancer Genome Atlas	TCGA
The Melanoma Institute of Australia	MIA
Transcription factor p65	RELA

Transcription factor RelB	RELB
Transporter 1, ATP Binding Cassette Subfamily B Member*:*	TAP1
Tricarboxylic Acid	TCA
Tumour Infiltrating Lymphocytes	TILs
Tumour Microenvironment	TME
Tumour size, Lymph Node affected, distal Metastases	TNM
Tyrosine-protein kinase 2	JAK2
Tyrosine-protein kinase 1	JAK1

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Aims of this thesis

- To apply appropriate bioinformatic methods of immune microenvironment characterisation to gene expression data from formalin-fixed paraffin embedded (FFPE) primary cutaneous melanomas from individuals in a large population-based cohort.
- To identify prognostic subgroups based on the bioinformatically identified immune infiltration profiles.
- To identify and describe molecular and environmental immunosuppressive factors in primary cutaneous melanoma.

1.2 Aims of this chapter

The main aim of this chapter was to give an overview of the current knowledge that is essential for this thesis:

- To describe relevant knowledge about cutaneous melanoma.
- To document known predictors of melanoma specific survival (MSS).
- To describe the most commonly disrupted molecular pathways in melanoma and the available related targeted therapies.
- To introduce what is published with respect to host immunity-tumour interactions and some therapies that might modulate these interactions.
- To give an overview of the basic immune responses.
- To describe the methods of characterising immune cells within the tumour microenvironment.

1.3 Cutaneous melanoma

Human skin (cutis) consists of three layers: the epidermis, dermis and subcutis. The basal layer of the epidermis (stratum basale) contains melanocytes, which are melanin

producing cells. In the skin, melanocytic benign proliferations, which are usually self-limiting are known as melanocytic naevi or “moles”. Persistent proliferation occurs in some people and then the naevi become unusually large (>5mm in diameter), which are known as dysplastic or atypical naevi. Even though proliferation is more persistent in these naevi, it usually eventually stops. When the proliferations become uncontrolled, a proportion of naevi may develop into cutaneous melanoma, however this malignant neoplasm may equally arise from skin which is considered as normal. Understanding the drivers of proliferation is very important in melanoma and other cancers in order to be able to control them therapeutically. In this thesis, I report the use of tumour transcriptomics to understand this process and how the host tries to control the tumour by deploying immunological responses. For brevity, the term “melanoma” will be used to mean cutaneous melanoma throughout this thesis.

Reported sunburn and sunbathing are established risk factors [2] and this led to the development of the intermittent exposure hypothesis [3]. It is hypothesised that the continued observed increase in melanoma incidence results from the increased exposure to intense sun, especially on holidays. In the UK the current trends towards tanned skin colour and accessible sunny holidays for people whose skin type (pale) is not photo-adapted might explain the rise in melanoma incidence. In the UK, the incidence of melanoma has been increasing since 1990 (it has more than doubled - 128%) [4].

1.4 Melanoma prognosis

1.4.1 Disease staging using the American Joint Cancer Committee on Cancer system (AJCC)

Clinico-histopathological features which predict survival for melanoma are used by clinicians and pathologists to assess the patient prognosis. The formal staging system of melanoma used in the UK is that developed by the American Joint Cancer Committee on Cancer (AJCC). I have used the 7th edition [5][6] as this was the current version when the data for my study were collected. The AJCC system categorises melanoma into four main stages and here I use a broad summary. In truth, the classification is rather more detailed than indicated below:

- Stage I: early primary tumours thinner than 1mm (Breslow thickness, see below)
- Stage II: primary tumours which are thicker but without evidence of spread to the lymph nodes;
- Stage III: evidence of melanoma spreading to the regional lymph nodes and palpable lumps;
- Stage IV: more aggressive disease spreading to distant organs (viscera, lung, brain, etc.).

Within stage I and II melanoma, the staging is determined initially by the pathologists who record the Breslow thickness which is the depth of the melanoma from the superficial (granular) layer of the epidermis to the deepest part of the tumour. Breslow thickness is measured under the light microscope using a millimetre scale and the thicker the depth the worse patient's prognosis [5][6].

Stages I and II are further refined by presence/absence of tumour ulceration (also assessed using microscopic examination). Ulceration is the absence of an intact epidermis covering the melanoma [7][8]. Patients presenting with an ulcerated melanoma have a shorter survival than those with a non-ulcerated tumour [9]. For stage I, an additional measure to categorise melanoma is mitotic rate. It is the mitotic count (number of actively dividing cells) per tumour area of 1 mm^2, which is manually counted by pathologists using the light microscope. A mitotic rate greater than 1/mm^2 in stage I tumours is associated with decreased patient survival and has been used in the 7th Edition of AJCC staging. However it has been recently shown that a high mitotic rate indicates poor prognosis and aggressive disease even in thicker tumours (>1mm, n=1524) [10]. For this reason, in this thesis, the mitotic rate was analysed independently of AJCC staging for the whole study cohort.

The measurement of Breslow thickness, the detection of ulceration, and the presence of metastases is used to categorise the tumours into the so-called TNM classifiers (Tumour size, Lymph Node affected, distal Metastases) [5]. Individuals whose tumours are thicker than 1mm are usually then offered a staging procedure known as sentinel node biopsy. In this procedure dye and a radioactive tracer are used to locate the lymph node into which lymph from the tumour drains. The dye collects in the sentinel node: the first lymph node into which lymph drains. This node is removed and if that and or other nodes are found to contain melanoma cells then the patient is said to

have stage III melanoma. Patients may develop nodal tumour masses, which can be felt and this is also stage III disease but stage IIIb or IIIc.

Within the late stages, the number of metastases and the organs affected (stage IV) as well as the number of lymph nodes involved (stage III and IV) predict melanoma survival [5][6]. Metastases to non-pulmonary visceral sites predict the poorest prognosis, such as the liver and brain. There are other clinico-pathological factors which also independently predict prognosis but are not included in AJCC staging, such as lymphovascular invasion, primary tumour site, patient sex and age, as well as the lymphocytes infiltrating the tumour. They are briefly described below.

1.4.2 Lymphovascular invasion

Lymphovascular invasion is a term that describes the histological detection of invasion within the lumen of vascular and lymphatic vessels by melanoma cells, within or around the tumour. It is known to be associated with an increased likelihood of metastasis to the sentinel lymph nodes [11][12]. A recent study reported that lymphovascular invasion (assessed histologically) independently predicts poor survival [11].

1.4.3 Primary tumour site

Non-ocular melanoma occurs in different anatomic locations: on the head and neck, trunk (most common), limbs, genitals, rectum and mucosal sites. Primary tumours located at the trunk, head and neck are known to be associated with a poorer prognosis than the ones on the extremities [13]. Other studies showed that tumours located on acral sites (fingers, palms, soles, and nail beds) have poor prognosis [14][15], which could be due at least in part to late diagnosis [16]. Similarly melanoma detected in genital and rectal areas were reported to have lower survival rates [17][18]. Overall, unsurprisingly, the tumours occurring on sites where they are easy to detect and surgically remove are associated with the best prognosis, however there might be biological differences among tumours in different body sites. It is generally accepted that tumours exposed to sun have higher rates of C>T mutations caused by UV radiation than sun-protected tumours [19]. These mutations may induce the generation of neo-antigens (described later in this chapter), which might potentially attract immune cells to kill the lesion and result in better survival for patients having

tumours in sun-exposed sites.

1.4.4 Sex

Sex is an independent predictor of melanoma death [20][21]. Female patients have a lower risk of melanoma death than males consistently when analysed within pre- and postmenopausal ages categories across the stages III and IV [20]. The basis of this risk difference is not yet fully understood.

1.4.5 Age

Age of melanoma patients is an important predictor of survival. Increasing age is an independent predictor of melanoma death [22], which is currently hypothesised to relate to impaired host responses to the tumour [22][23] and/or changing collagen matrix [24] impacting on the tumour invasive front. It has also been observed that older patients at diagnosis are more likely to be men than women [22][25], which makes the effects of age and sex difficult to disentangle. In multivariable analyses, both factors appear to have independent effects on survival.

1.4.6 Tumour infiltrating lymphocytes (TILs)

More than a century ago, immune cells occurring within melanoma were thought to be causative for skin neoplasia due to a process of inflammation [26]. The idea was challenged at the International Congress of Skin Cancer conference in Sydney in 1972 suggesting instead the anti-melanoma function of lymphocytes [27]. The tumour infiltrating lymphocytes (TILs) term was introduced in 1969 by Wallace Clark and denoted the histologically detected presence of lymphocytes within and around primary melanomas, postulated to be taking part in a host response to the cancer [28]. Clark et al. [28] recognised that there were different patterns of lymphocytes within the tumours and classified TILs as absent, non-brisk and brisk (Figure 1.1).

Brisk TILs have been shown to predict an improved outcome (overall and disease free survival) in some studies [29][30], but others have failed to prove an independent survival advantage [31][32]. The immune response might vary over time during melanoma progression, and that fact could explain why the survival benefit of TILs is not observed consistently across studies. Another plausible explanation of this inconsistency is the heterogeneous nature of the immune infiltrates within and around

tumours. There is furthermore, a large degree of inter-observer variation: the scoring can be pathologist-dependent, which has so far limited its utility [33][34][35]. Nonetheless, a recent study based on a large sample size (n=1865) did confirm that patients with higher counts of brisk TILs have very good prognosis [36]. Even if inter-observer variation can be controlled there may be biological differences which could not be captured using standard histopathology: TILs scoring does not indicate exactly which type of lymphocytes invade the tumour. Some might be immuno-suppressive, while others might be more anti-tumorigenic.

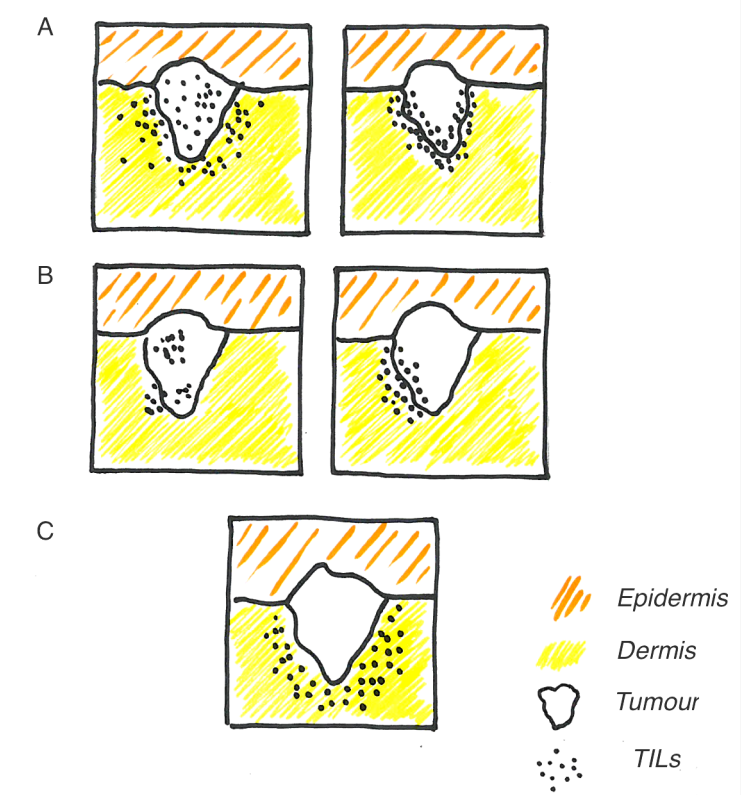


Figure 1.1: Clark's classification of tumour infiltrating lymphocytes

(A) Representation of brisk TILs, which are present within the tumour or infiltrating the entire base of the tumour. (B) Representation of non-brisk TILs, which are observed in one or more foci of tumour. (C) TILs absent within the tumour, where no lymphocytes have infiltrated melanoma (even though they could be seen around it).

The Melanoma Institute of Australia (MIA) has proposed a modified version of the Clark's classification, introducing new grades ranging from 0 to 3 [36]:

- Grade 0: TILs absent
- Grade 1: a mild/moderate focal or a mild multifocal TIL infiltrate
- Grade 2: a marked focal, either a moderate/marked multifocal, or a mild dispersed TIL infiltrate
- Grade 3: a moderate/marked dispersed TIL infiltrate.

Like Clark's classification, this grading positively correlates with melanoma specific survival [36]. In this thesis, only the Clark's classification was used assessed by unselected clinical dermatopathologists and single observer Dr Sally O'Shea as separate variables to test their cross-validation. Throughout later analyses I will point out whose classifications were used.

1.5 Key oncogenic pathways, genetic mutations

Melanoma progression is associated with activation of pathways that regulate tumour growth and proliferation, which I will discuss in this section.

1.5.1 Mitogen activated protein kinase (MAPK) signalling pathway

MAPK signalling is a regulator of cell cycle, differentiation, migration, proliferation and apoptosis [37]. In healthy cells, this pathway is triggered either by receptor tyrosine kinases (RTKs) binding to their accompanying ligands, or integrin adhesion molecules on the cellular matrix or the cell membrane leading to further activation of RAS (small GTPase) by changing its state from GDP (guanosine diphosphate) to GTP (guanosine triphosphate) [37][38]. Activated RAS activates RAF (serine/threonine kinase), activating MEK1 and MEK2, then ERK1 and ERK2, resulting in their translocation to the nucleus and regulation of several transcription factors (Figure 1.2) [37][38]. The first most common driver mutation in melanoma is *BRAF*, which results in uncontrolled activation of the MAPK pathway. The *BRAF* point mutation V600 is a substitution of valine by another amino acid. *BRAF* mutations occur in ~50 % of cutaneous melanoma[39] (with the most common substitution of E (glutamic acid) occurring in ~90% of these mutations and K (lysine) occurring in the remainder 10%).

Targeted inhibitors or BRAF +/- MEK inhibitors are used to inhibit the MAPK pathway. Recent clinical trial results indicated that combinatory treatment delays therapy resistance and improves progression free and overall survival [40]. *NRAS* is the second

most commonly driver mutated gene in melanoma, also activating the MAPK pathway, with a frequency of 20-25% [41]. The point mutations occur in codons G12, G13 and Q61. However there is no treatment available in clinical practice to date that targets NRAS as yet [42].

1.5.2 Phosphatidylinositide 3 Kinases (PI3K) signalling pathway

Activation of the PI3K pathway is also involved in melanoma proliferation, cell growth and survival [43]. PI3K is activated by G-protein-coupled and tyrosine kinase receptors. Activated PI3K produces phosphatidylinositols, which bind to AKT, restraining it at the cell membrane. AKT restrained in the cell membrane can be phosphorylated/activated [44][45], resulting in the mediation of phosphorylation of downstream proteins (e.g. mTOR), which regulate cell cycle, proliferation and survival (Figure 1.2). Oncogenic RAS is known to regulate the PI3K pathway, leading to uncontrolled activation of this pathway and activating cancer associated genes. In normal cells, PTEN inhibits this pathway, by interrupting the reaction of AKT with phosphatidylinositols. In melanoma, 40-60% of tumours have deletions or mutations in *PTEN* [46] and they are more frequent in BRAF mutated tumours [46]. It was reported that approximately ~20% of melanomas have *PTEN* inhibited/deleted together with *BRAF* mutations [47][46] and that both of these aberrations cooperate in metastatic induction [47]. PI3K and AKT inhibitors have been developed, however they are not commonly used in the clinic. Another gene that is functionally important in melanoma progression is *NF1*, which, like *PTEN*, is also involved in the negative regulation of RAS. Mutations of *NF1* occur in approximately 4% of melanomas [46].

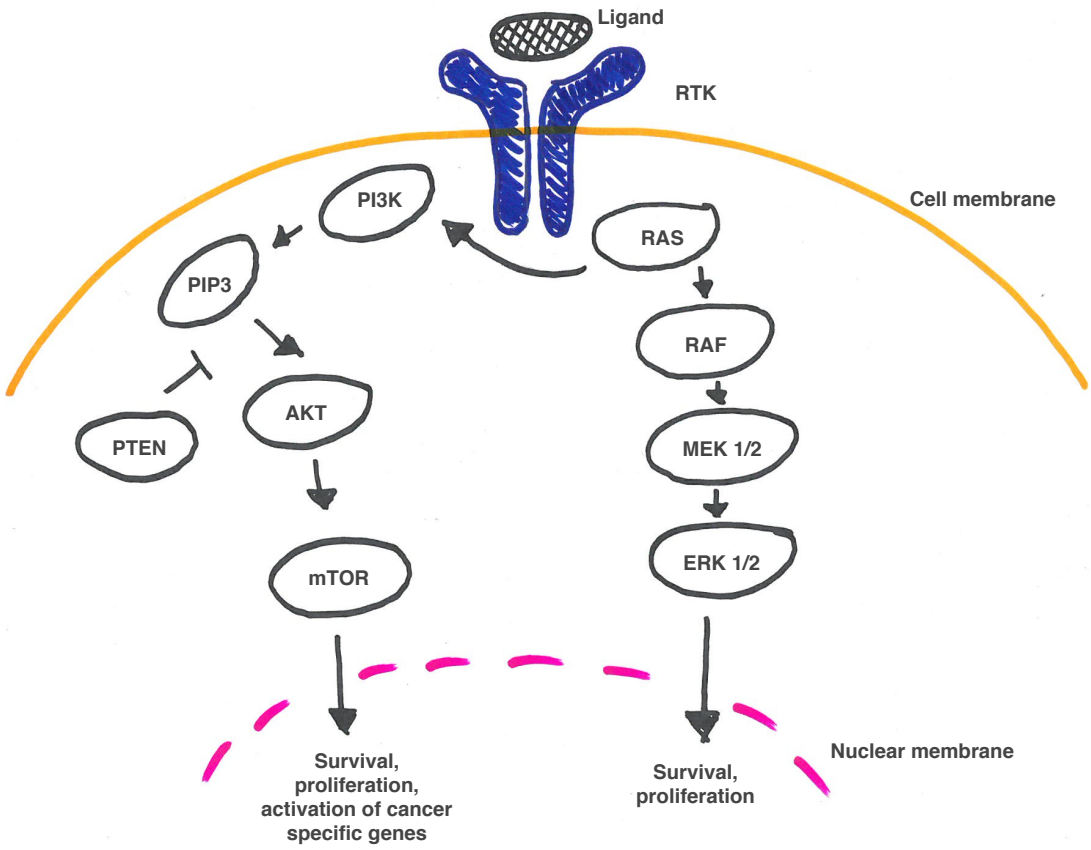


Figure 1.2: Simplified overview of the MAPK and PI3K pathways

1.5.3 Genetic mutations

A study designed for large scale mutation detection identified some more mutated genes in cutaneous melanoma: *TP53, CDKN2A, MAP2K1, PPP6C, RAC1, SNX31, TACC1, and STK19* [48], however with low frequency. This study included 15 primary tumours, 30 metastatic samples, and 76 short-term cultures derived from metastatic tumour tissue from 95 melanomas of cutaneous, 5 of acral, 2 of mucosal, 1 of uveal, and 18 of unknown primary origin.

Additionally a study conducted three years later using 67 primary and 265 metastatic samples confirmed some of these mutations: T*P53, CDKN2A, MAP2K1, PPP6C, and RAC1,* and identified some new mutations relatively frequent in melanoma: *ARID2*, *IDH1*, *DDX3X*, *RB1* [46].

Moreover, melanoma was shown to be a cancer type with the highest prevalence of somatic mutations in a study comparing mutational load across various cancer types [49]. This study identified 21 signatures of mutations, one of them (the Signature 7)

was characterised by C>T mutations known to be caused by ultraviolet light and this signature was particularly high in melanoma [49].

Apart from the mutations some frequent copy number changes in specific genes were identified in melanoma. For example, commonly amplified gene was *MYC* [50] and deleted were *CDKN2A* [46][48]and *PTEN* [50][46][48].

1.6 The immune response

Generally, the immune system is described as having two main components: innate and adaptive. The innate immune responses are mediated by germline encoded receptors and these responses are rapid reactions to pathogens (sometimes described as spontaneous), playing a role in both the immediate broad defence against challenges and in subsequently activating secondary more specific or adaptive immune responses. Pathogens are first recognised, by innate/phagocytic cells using pattern recognition receptors (PRRs), then they are ingested and killed. There are two main types of PRRs: Toll-like (extracellular pathogen detectors) and NOD-like (intracellular pathogen detectors). They recognise foreign DNA and RNA and this recognition triggers the production of IFN-α and -β (Toll-like receptors) and NF-κB (NOD-like receptors). Another similar process of foreign nucleic acid recognition is stimulator of interferon genes (STING) situated on the endoplasmic reticulum and when it is active it leads to transcription of interferon genes [51]. This signalling was shown to have significant impact on cancer immunity [52]. The major immune cells in innate immunity are:

- macrophages
- neutrophils
- dendritic cells
- mast cells
- eosinophils
- basophils
- natural killer cells.

The adaptive immune responses are mediated by somatically rearranging antigen receptors (which is a continuous random process). The adaptive responses are less rapid (as specific B and T cell clones must be triggered by other immune cells and

proliferate to provide effective coverage) but more specific than the innate response and its main function is the development of highly specific effector cells and memory responses, from two main types of lymphocytes: T and B cells. Both these cells have several subtypes, which have distinctive functions. For T cells these are:

- effector
- helper 1, 2 and 17
- regulatory
- cytotoxic
- memory
- natural killer T cell (NKT, different from NK cells)
- gamma-delta;

and for B cells:

- plasma cells
- memory B cells
- regulatory B cells.

T cells mature in the thymus, hence their name, and they can be either CD4+ or/and CD8+ (which means they express CD4 or/and CD8 glycoprotein on their surface). All subsets of T cells express CD3 and T cell receptors (TCR), forming a TCR complex. This complex interacts with HLA class I (CD8 T cells) and II (CD4 T cells). B cells however, mature in the bone marrow and express the B cell receptor (BCR). The BCR is a transmembrane receptor complex, comprised of the antigen binding component (an immunoglobulin molecule) and associated signalling chains (CD79A and B, analogous to the CD3 chains of the TCR). The B cell's key function is antibody (immunoglobulin) secretion. The role of antibodies is in neutralising pathogens by direct binding to them. B cells also capture antigens using immunoglobulin and present it to T cells via HLA class II; a T cell with specificity for the processed antigen captured by the antibodies then provides helper function (co-stimulation and cytokines release) that allows the B cell to differentiate into an antibody secreting plasma cell.

All immune cells are derived from the pluripotent hematopoietic stem cells from the bone marrow, which have undergone differentiation mainly due to gene expression changes in different tissue types (haematopoiesis), shown in Figure 1.3. The first step

of differentiation is into common lymphoid and common myeloid progenitor cells. The first one matures into T cells, B cells and natural killer cells, the latter one however matures into all the other immune cells. Overall in healthy individuals the differentiation is fluid, but at the same time well controlled [53]. The classification of immune cells is mainly based on the expression of surface molecules, which are characterised by the use of standardized cluster of differentiation (CD) nomenclature [54], but functional tests are often used to differentiate particular subtypes. However, the understanding of subgroup phenotypes and functions continues to evolve.

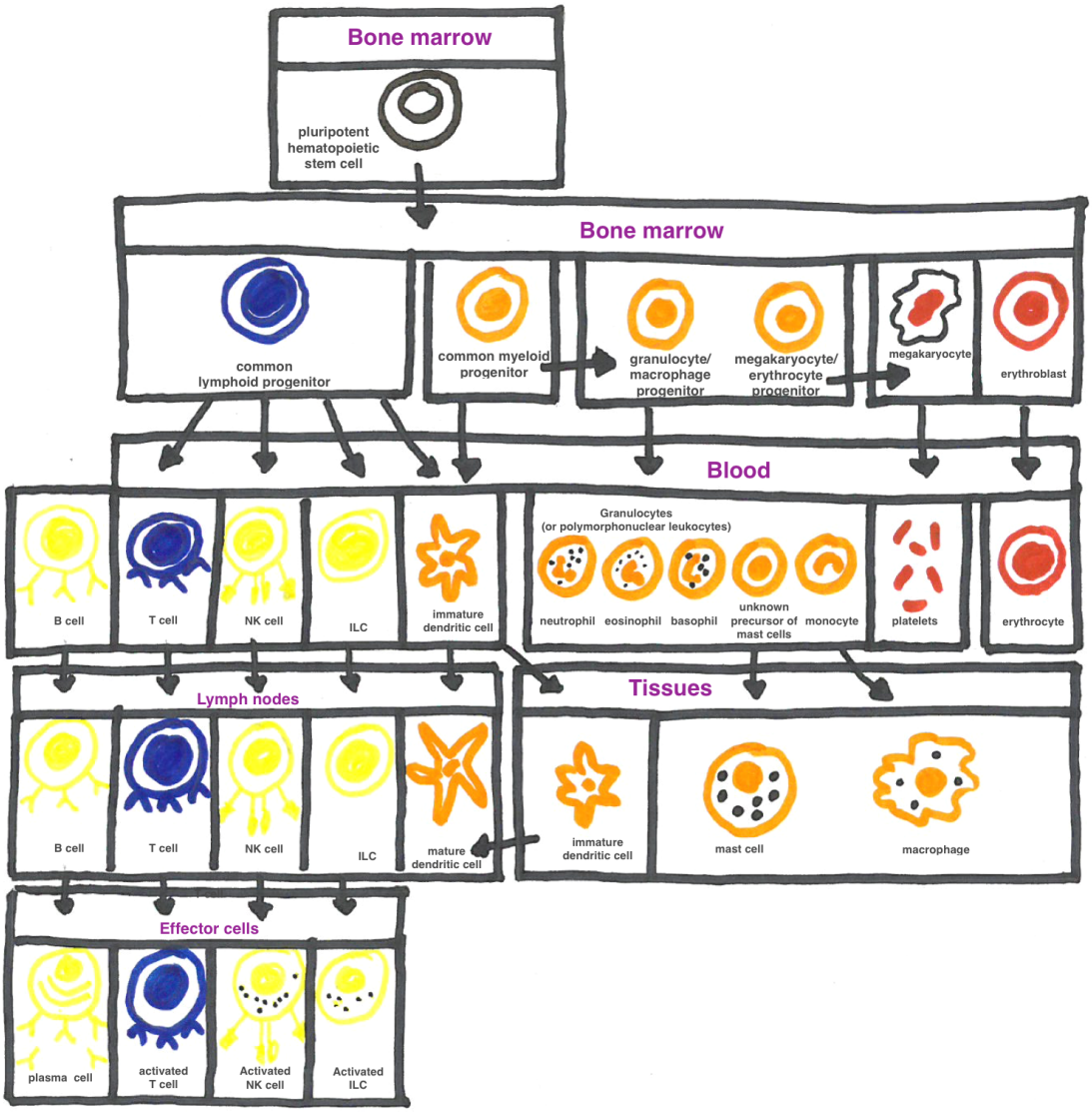


Figure 1.3: Haematopoiesis, process of differentiation of immune and blood cells [53]

Immune response is a complex process in which many different subtypes of immune cells take part. It is possible that in a fight against cancer all of those immune cells may

be involved simultaneously or at different stages, therefore understanding this process is needed but is very challenging [55][56].

1.6.1 Host immune response to melanoma

The immune system is a defence mechanism that primarily protects an individual against pathogens but it has become increasingly apparent that it also plays a crucial role in the defence against cancer. The first scientist who conceived the idea of the immune system potentially controlling cancer proliferation was Paul Ehrlich in 1908 [57]. However, it was only a half century later in 1957 that the formal hypothesis of "cancer immunosurveillance" was proposed by Sir Frank Macfarlane Burnet and Lewis Thomas [58].

For melanoma, the first evidence that the immune response, mainly the TILs might play a role in melanoma was the discovery by Clark et al. in 1969 [28]. TILs were later on proven to be associated with an improved melanoma specific survival (as described before). Melanoma has been said to be the most immunogenic cancer over the past decades, based largely upon the observations made by clinical researchers. One of the earlier and most striking observations was the spontaneous regression (disappearance) of melanoma lesions in certain patients [59]. It is a very rare process speculated to be caused by a rapid cytotoxic (tumour killing) specific immune response [60], or by stimulation of the immune cells by the melanoma cells present in the lymph nodes [61].

Other studies reported that some melanoma patients develop vitiligo (hypopigmentation), an autoimmune process that destroys melanocytes (acquired pigment loss), and that these patients have a better prognosis [62]. It was suggested that this higher rate of survival could be explained by higher CD8+ T cell responses to melanoma cells in those patients possibly due to recognition of same antigens produced by both melanocytes and melanoma [63][64].

The prominent role of the immune response controlling melanoma was also shown while studying patients who received organ transplants who are typically given immunosuppressive drugs to avoid the graft rejection. The studies were designed to test the hypothesis that transplant recipients were at higher risk of developing skin cancer/melanoma. It had become clear that organ recipients indeed were at higher risk of developing this type of cancer due to lack of systemic immune surveillance [65].

There were some transplantation cases where the organ of the donor contained melanoma cells (unknowingly to the surgeons) and because of the immunosuppressive drugs the recipients received during and after transplantation, they developed melanoma metastasis to distant organs [65][66]. These observations imply both that melanoma cells can exist in healthy organs for many years yet retain the ability to proliferate, and that in healthy individuals' immune responses to melanoma may keep tumours in check.

Overall, melanoma tumours in order to be recognised and attacked by the immune system need to provide a supportive niche for the immune cells to infiltrate melanoma and activate their effective functions. These functions are mainly induced by tumour antigens. Tumour antigens can be classified into:

- Cancer- germline antigens – expressed by cancer cells and adult reproductive tissues (e.g. melanoma associated antigens (MAGE family))
- Differentiation – expressed by cancer cells and limited range of normal tissues (e.g. produced by melanocytes and melanoma cells to which T cell tolerance is incomplete, such as MART-1, melan-A or Tyrosinase)
- Overexpressed – expressed in cancer cells and normal tissues, but significantly overexpressed in tumour cells
- Viral – expressed by cancer cells as a result of viral infection
- Neoantigens – expressed in cancer cells and absent from the normal tissues peptides considered to be a result of degraded abnormal proteins, which can be products of genomic mutations [67][68][69].

Tumour antigens are processed and presented to the immune system via MHC class I molecules. Simplistically, the abnormal endogenous proteins are ubiquitinated and fragmented into peptides by the proteasome. Next, these peptides are transported from cytosol to endoplasmic reticulum by ATP Binding Cassette Subfamily B Member Transporter (TAP) and trimmed by Endoplasmic Reticulum Aminopeptidase (ERAAP). Finally these peptides are bound and presented to the immune cells by Major Histocompatibility Complex class I (Human Leukocyte Antigen (HLA) class I: HLA-A, HLA-B, HLA-C), formed with its major component Beta-2-microglobulin (B2M)[53] (Figure 1.4).

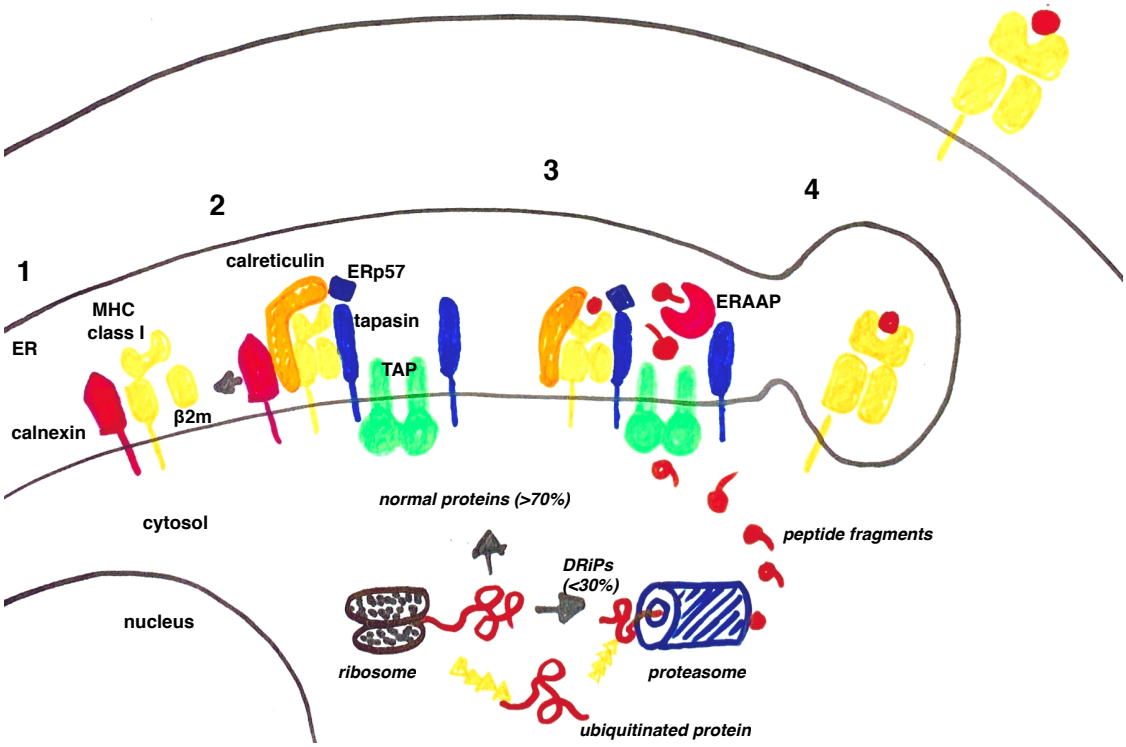


Figure 1.4: Schematic representation of antigen processing and presentation via MHC class I molecules [53]

(1) Partially folded MHC class I α chains bind to calnexin till β_2-micorglobulin binds. (2) When MHC class I α:β_2m complex is released from calnexin, then binds to a complex of chaperone proteins (calreticulin, ERp57) and then to TAP via tapasin. (3) Cytosolic proteins and defective ribosomal products (DRiPs) are degraded to peptides by the proteasome and delivered to the ER by TAP. (4) A peptide binds to MHC class I molecule and finishes its folding. Next, the MHC class I molecule is released from the TAP complex and exported to the cell membrane [53].

As mentioned earlier melanoma is considered to be a cancer type with the highest prevalence of the somatic mutations [49], which is considered to potentially lead to higher numbers of neoantigens, hence being more immunogenic.

Melanoma can also be recognised by innate responses, mainly by Natural Killer (NK) cells [70] and Dendritic cells (DC) [71], for which neoantigen presentation is not essential. NK cells have a cytotoxic function: killing infected cells and tumour cells. NK cells do not use an antigen receptor, but instead use an array of activating and

inhibitory receptors to detect changes in expression levels of host cell surface molecules. Activated NK cells kill target cells using granzymes and perforins, which induce apoptosis of those target cells and produce TNF, GM-CSF and IFN-gamma to modulate innate and adaptive immunity [53]. DCs, however, process and present antigens (via HLA class I and II) to T cells. They are sometimes described as professional antigen presenting cells (APCs). There are two broad classes of dendritic cells: conventional (myeloid) and plasmacytoid (which are much less abundant), which produce high levels of IFN-γ and IFN-α [72][73].

1.6.1.1 *Regulation of host immune response*

Although it is evident that melanomas can provoke anti-tumour responses and can be eliminated by that process, at some point some of the tumour cells clearly acquire the ability to evade the immune reaction, which results in tumours that grow faster with increased ability for invasion and metastasis [74]. For example, the downregulation of components of antigen processing and presentation and IFN-γ signalling [55] contribute to immune evasion, as well as upregulation of β-catenin signalling, which inhibits the expression of chemokine (CCL4) that attract the infiltration of CD103+ DCs into the tumour [75][76].

Another type of immune evasion mechanism reported is the up-regulation of checkpoint molecules like PD-L1, PD1 and IDO or chemokines by tumour or immune cells that activate and attract regulatory immune cells. However the upregulation of checkpoint molecules was recently shown to be associated with higher immune infiltrates which have a good prognostic value [77]; the expression of checkpoint molecules within tumour microenvironment may simply reflect a homeostatic response to activation of T cells functioning as a feedback loop. Furthermore, Myeloid Derived Suppressor Cells (MDSCs) can have the ability to supress the immune response against melanoma. MDSCs originate from the myeloid linage (stem cells from the bone marrow). These are a heterogeneous group of cells, which have an ability to suppress T cells in pathological conditions, including in cancer overall and melanoma in particular [78][79][80]. Similar overall function to MDSCs have Regulatory T cells (Tregs). Tregs are CD4 positive expressing the forkhead box P3 (FOXP3) protein which is a transcription factor. As their name implies, they take part in the regulation of immune responses. They express checkpoint molecule Cytotoxic T-lymphocyte-

associated protein 4 (CTLA-4) (see “Immunotherapies” below). In some cancer types, Tregs have been reported to be associated with a poor survival while in other cancers it is the opposite [81].

Moreover, other components of tumour microenvironment (TME), such as stromal cells, cancer-associated fibroblasts (CAFs), vascular components, and adipose cells, and their complex interactions might also regulate the effectiveness of immune responses against melanoma [82]. All of the them can release different types of cytokines affecting of the anti-tumour immune responses as well as tumour progression [82]. For example in melanoma lymphatic vessels were shown to have an immuno-suppressive role as well as a role in inducing the anti-tumour local inflammation/immune response [83][84]. Another study reported that the changes in collagen matrix (related to age) were associated with impaired host immune responses to the tumour [24]. Last but not least, in melanoma CAFs were reported to contribute to decrease of melanoma susceptibility to be killed by NK cells [85], and in general in cancer, they are recognised for induction of excessive inflammation within the tumour [86].

Finally, the tumour-host interactions might be also affected by genetics, age of the patient, microbiome, viral infections, exposure to sunlight, and immune-modifying drugs [87]. For example, less sunlight, viral infection and younger age are associated with higher immune responses [87]. Moreover, smoking and Vitamin D are known to affect systemic immune responses [88][89], but their role in tumour-host interaction is not yet understood in melanoma. During my thesis I had access to reported patient’s smoking habits and vitamin D blood levels, hence I have explored only these two and described in details in Chapter 6.

1.6.2 Immunotherapies

Over time, the accumulation of data indicating that immune responses are important in melanoma has led to the development of therapies that can boost immunity. Despite a slow start in this field, the current lines of immunotherapies are effective in a significant proportion of melanoma patients.

The most promising therapeutic option for stage IV melanoma (advanced metastatic stage) are immunotherapies, most commonly checkpoint blockade. Immune checkpoint receptors (which are targeted by immunotherapies) negatively regulate T

cell activation, and function to limit potential self-induced damage to the host. CTLA-4 is mainly expressed on activated T-cells and Treg and is closely related to the CD80/86 ligand, as well as costimulatory molecule CD28. Both CD28 and CTLA-4 compete for binding to CD80/CD86 [90][91]. When the receptor is bound to CD80/86, a signal is produced to switch off the activated T-cell alternatively - CTLA4 on Treg blocks CD28 binding to CD80/86 [90][91]. Programmed death ligand 1 (PD-L1), is another checkpoint molecule which may be expressed by antigen presenting cells or tumour cells. It binds to programmed death 1 receptor (PD-1), which is expressed on macrophages, B cells and activated T cells, NK cells, NKT, cells and subsets of DC cells [91], causing T cell inactivation. The CTLA-4 inhibitor Ipilimumab (Yervoy) was approved by the FDA for metastatic melanoma treatment in 2011 [92], followed by Pembrolizumab (Keytruda) and Nivolumab (Opdivo), which block PD-L1 [93][94].

The unprecedented therapeutic benefits reported from combined checkpoint blockade (anti-PD-1 and anti-CTLA4) have shown the potential to exploit the current body of knowledge on the immune response to cancer, but still only around 58% of treated patients survive 3 years [95]. Note that these combinations are still in clinical trials and have so far only been followed-up for a short period of time.

Checkpoint blockade is associated with significant toxicity (immune related adverse events such as inflammation of multiple organs), which causes significant morbidity even treatment related deaths. The toxicity issues are even more problematic where checkpoint blockade is used in the adjuvant setting. Most recently a study showed that neoadjuvant (first step treatment before the main therapy) setting for immunotherapies might be more promising for melanoma patients, however still the toxicity rates are very high and more studies are needed to overcome this problem, for example by dose reduction [96][97].

Another type of hopeful immunotherapy is adoptive cell therapy (ACT). This type of therapy was introduced in 1998 by Steve Rosenberg, which was based on isolating TILs from the tumour, culturing and propagating the cells *ex vivo*, and finally injecting them back to the melanoma patients [98]. Subsequently, similar technique was introduced by isolating peripheral immune cells and engineering them with chimeric antigen receptors (CARs). The idea behind this treatment is to engineer the patient's immune cells and make them more sensitive to cancer cell recognition. T cells are extracted

from the tumour, cultured *in vitro*, engineered by introducing chimeric antigen receptor (CAR) into their genome and transferred back to the patient's body [99][100]. CAR is a treatment combination of antibodies, TCR and co-stimulatory molecules. Specifically, CAR is composed of extracellular portion, which usually are antigen binding domains derived from variable domains of antibodies, and which are designed to recognise specific tumour antigens. The intracellular signalling module of the CAR is derived from T cell signalling proteins, mainly CD3ζ (CD3 molecule and ζ-chain) and additional costimulatory domains from receptors such as CD28, CD134 (OX40), and CD137 [101]. When CAR is activated by tumour-associated antigen, it has the ability to activate different pathways at the simultaneously, for example T cell effector and costimulator functions [99][100], which results in over-activation of T cells. As for combined checkpoint inhibitors [102], around 50% of melanoma patients currently benefit from adoptive T cell transfer therapy [103][104].

It is important to study immune responses in melanoma because it is needed to understand the failure of immune surveillance and resistance to immunotherapies. The increased understanding will help in designing more effective and potentially less toxic therapies. There is a possibility that environmental or lifestyle factors might modify the host-tumour interactions [87], henceforth studying them might lead to important findings that could be added to the patient care.

1.6.3 Established methods to characterise immune cell subtypes within tumour

1.6.3.1 Histopathological quantification of immune cell infiltration

Excised melanoma samples are normally prepared for examination by fixation in formalin and embedding in paraffin wax (FFPE), although for large samples portions of the tumour can be stored fresh-frozen (cryopreservation) for research purposes. FFPE is the preferred method in pathology laboratories because the tissue block can be stored for a long time and be reanalysed if needed for further diagnoses. On the other hand, cryopreservation is preferable for many research processes since the genetic and protein material is not degraded by formaldehyde and higher temperature storage, which is the main disadvantage of FFPE method. From both FFPE or cryopreserved tissue, sections of the tissue are cut, stained (predominantly with Hematoxylin and

Eosin (H&E) for histopathological characterisation) and examined using light microscopy. The extent of the TILs can be assessed by eye; this is a routine part of diagnostic histopathological examination of melanoma within the clinical setting as described earlier in this chapter.

However, to distinguish many of the individual cellular subpopulations within the infiltrate, immunohistochemistry (IHC) is required. This method of staining relies on the use of antibodies that specifically bind to an epitope (part of an antigen recognised by an antibody) expressed by immune cells. The nomenclature of immune cell subtypes has been developed on the basis of proteins expressed on their cell surface in blood. This method of measuring immune subtypes within the tumour is labour intensive and challenging if more than a handful of antigens are to be tested simultaneously. For simultaneous multi-antigen detection, multiplex immunochemistry is required, usually with usage of fluorochrome rather than standard chromogenic reactions [105]. Although this method is commonly used it carries some technical limitations, such as signal overlap between fluorophores [105]. To overcome this problem Garry Nolan and colleagues at Stanford University developed a more advanced method: Multiplex ion beam imaging (MIBI), where up to 100 antigens can be detected on FFPE tissues utilising metal isotopes reporters [106] which can be washed off the surface allowing consecutive probes to be used. This method is not yet widely used however because it is too expensive and requires specialized training for performance and analysis of the results. In summary, all types of IHC require a reasonable amount of tumour tissue and are time consuming especially when assessing numerous immune cell populations in a large number of samples.

Flow cytometry (or FACS, for Fluorescence-activated cell sorting) of disaggregated tissues offers an alternative antibody-based method to measure the numbers and proportions of immune cells in tissues. However, in this case the tissue has to be fresh – cells need to be alive in order to obtain specific antigen-antibody binding [107]. This is rarely available for melanoma primaries at least, due to the very small size of the tumours. The concerns are around freezing tissues and effectively destroying their architecture precluding accurate staging of the tumour. Furthermore, flow cytometry is based on the detection of surface markers (Cluster of Differentiation, CDs) of the immune cell types, few of which are exclusive to individual cell populations, such that

multiple antibodies are required to differentiate between immune cells. Another key difficulty associated with FACS is tissue preparation, mainly its step of disaggregation, which might lead to destruction of surface antigens especially when the cells stick to each other which is often the case in carcinomas. Moreover, isolation of the cells from their original environment might induce different gene expression and protein patterns during this process.

1.6.3.2 *Bioinformatic methods of immune cells scoring within tumours*

Genomic and transcriptomic tumour profiling has become commonplace, as microarrays and sequencing techniques improved and costs fell leading to increased accessibility. Fresh or cryopreserved tumours are sampled, nucleic acids extracted and isolated and then assayed to generate detailed profiles. DNA or cDNA from RNA (depending on the research aim) can be used to generate different types of genetic profiles: genome sequences, copy number changes, promoter methylation and whole genome transcriptomes, or any combinations of the above. However, in order to have all genetic profiles per sample a reasonable amount of genetic material is required. Genetic profiling assays were needed with a high enough sensitivity to allow accurate signal detection when a very small amount of genetic material is available, and preferably effective in formalin fixed tissue.

The transcriptomic data utilised in this thesis were generated with arrays developed specifically to allow analysis of degraded RNA which occurs in formalin fixed blocks. This technological advance has allowed the research group to generate a uniquely large collection of tumour derived transcriptomes from FFPE primaries (as described in the Methods section).

In my thesis one of the tested hypothesis was that the bioinformatic analysis of tumour-derived gene expression profiles would allow inference of the abundance of a large number of immune cells subsets infiltrating the tumours using these cells specific markers.

I explored a number of bioinformatic approaches that have recently been published: CIBERSORT, the Immunome compendium reported by Bindea et al. [108], subsequently updated by Angelova et al. [1]. There are other methods in literature which infer immune cell subtypes [109][110][111][112], however they were not considered in this

thesis because they either cover a smaller spectrum of cell subtypes [109][111] or only RNA sequencing data may be applied to them [110], or they appeared relatively late during my PhD (November 2017) [112].

1.6.3.2.1 CIBERSORT – "LM22"

Newman et al. designed and validated a leukocyte gene signature matrix (immune sorting template) termed LM22 [113]. In their study, 547 genes were used to distinguish 22 human hematopoietic cell phenotypes:

The design of the approach was based on isolating immune cell types from the blood and generating whole genome transcriptomes for each purified cell type on microarrays (HGU133A) (Figure 1.5). After learning the particulars of each cell subtype in a machine learning approach (linear support vector regression), CIBERSORT (Cell type Identification By Estimating Relative Subsets Of RNA Transcripts) can generate an estimate of the relative proportions of these cells in any tissue using their transcriptomic data. Newman et al. developed an algorithm to address the problem of mixture deconvolution (where a signal, in this instance a measure of gene expression was a result of a signal derived from the tissue and something else which might be thought of as noise). The deconvolution algorithm was intended to produce a purer signal and where the deconvolution fails then the implication is that the data quality cannot be relied upon. The "P value of deconvolution" gives the CIBERSORT user a measure of whether the sample was well deconvolved or not. It is an empirical p-value estimated from simulation or resampling methods.

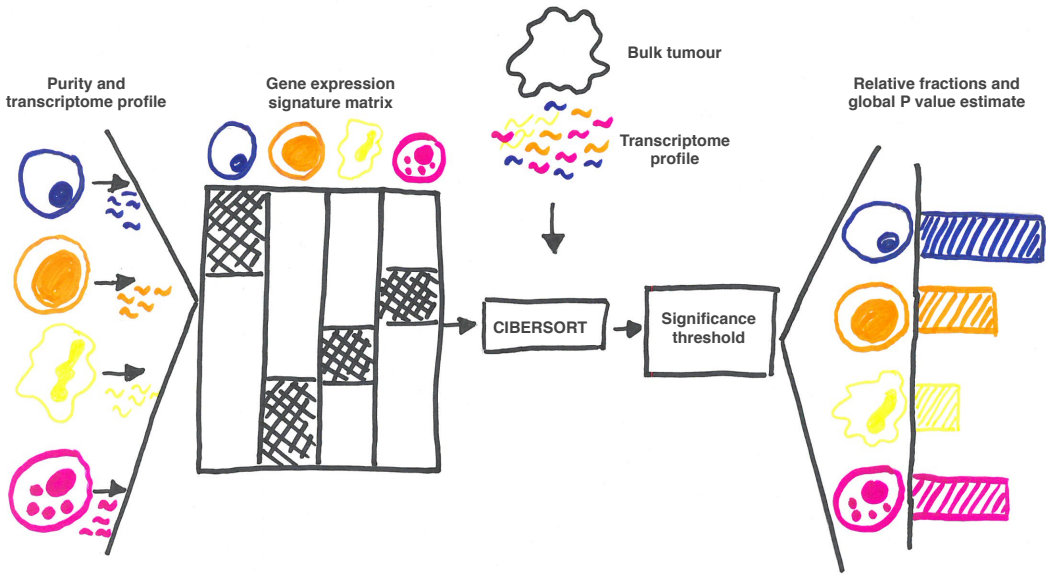


Figure 1.5: Overall representation of CIBERSORT design and application to leukocyte deconvolution (adopted from Gentles et al. [113]**)**

The gene expression data used as input into CIBERSORT can be derived from any tissue type. However, the algorithm was developed using RNA-seq data from fresh tissue and from a personal communication with the authors, it is unclear whether it could be applied to array data generated from FFPE tissue. As a way of evaluating this method, I applied CIBERSORT to the array data in the LMC study since it is one of the most well-developed tool with a dedicated online-accessible software and it is the only one (or one of a few) to provide a metric (p-value) indicating success or failure to infer immune cell presence.

1.6.3.2.2 Bindea et al. – the "Immunome"

Bindea et al. (2013) [108] used six publicly available gene expression datasets (from Gene Expression Omnibus and Array Express databases) derived from purified immune cell subsets to establish a reference - the Immunome compendium. In their study, the gene expression profiles from immune cells, the normal mucosa and colon cancer cell lines, blood and lymph vessels were tested using pairwise correlation to identify a unique signature of each of them. For example, if the genes expressed by blood vessels/normal mucosa correlated with some of the immune genes, then these immune genes were removed from the final gene list. The compendium contained 577

genes characterising 24 cell types and the control tissues (colon cancer cell lines, normal mucosae, blood and lymph vessels).

Each cell subtype has its genes assigned, but some of the genes can be expressed by multiple cell subtypes. Bindea and colleagues validated the method using multiple techniques and experimental approaches (including DNA microarrays (Affymetrix), quantitative polymerase chain reaction (qPCR), FACS, immunohistochemistry, and mouse models) and strong correlations were found with all these approaches. The inferred presence of cytotoxic, T cell, B cell and other cell subsets in the colon cancer tissue was strongly associated with improved patient survival [108]. We have applied this method to the FFPE melanomas of the LMC and found similar results [76]. However, since its first publication, the Immunome compendium was extended to more cell subtypes and was based upon more genes from extensive searches of a much larger number of sources [1]. In this thesis, I used this updated version (see below) and made some further developments (see Methods section) inspired from the earlier work conducted in our group based on the Bindea et al.'s signature [76].

1.6.3.2.3 Updated Immunome by Angelova et al.

The new Immunome by Angelova et al. [1] identifying uniquely expressed genes by immune cell subtypes, was an extension of Bindea et al.'s strategy [108], with however some different starting assumptions. Principally, the newer approach comprises 36 publicly available datasets from Gene Expression Omnibus and Array Express from purified immune cell from blood in various diseases. It was a significant increase in data sets used compared to Bindea et al. study, which used only 6 sources. All datasets utilised by Bindea et al.'s Immunome were also included in Angelova et al.'s. Among the genes described to be unique per immune cell score by both methods, only 112 genes are common to the 2 methods (Figure 1.6). Presumably, the absence of a large number of genes of the initial Immunome compendium (Bindea et al.'s) in the updated version (Angelova et al.'s) means that by those genes were no longer considered cell-specific using more extensive data sources.

The immune cell types proposed by each described method are shown in the Table 1.1: Immune cell types proposed by all the three methods and their existence in each of them.

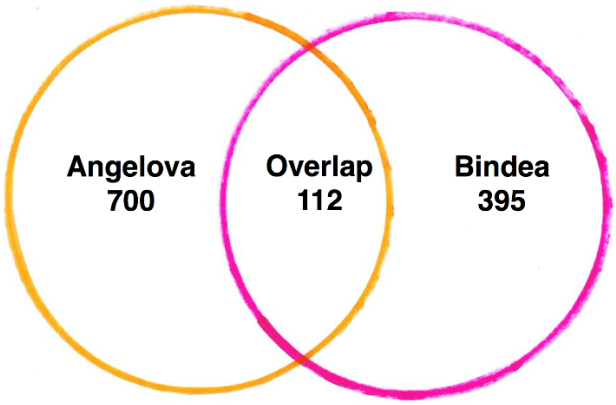


Figure 1.6: Intersection of genes from Angelova et al. and Bindea et al.'s Immunome compendiums

Table 1.1: Immune cell types proposed by all the three methods and their existence in each of them

Immune cell type	*CIBERSORT*	*Bindea*	*Angelova*
activated B cells			x
activated CD4			x
activated CD8			x
activated dendritic cells	x	x	
activated mast cells	x		
B cell		x	
B cells memory	x		x
B cells naïve	x		
central memory CD4			x
central memory CD8			x
cytotoxic cells		x	
dendritic cells		x	x
effector memory CD4			x
effector memory CD8			x
Eosinophils	x	x	x
immature B cells			x
immature dendritic cells		x	x
M0 macrophages	x		
M1 macrophages	x		
M2 macrophages	x		

Immune cell type	*CIBERSORT*	*Bindea*	*Angelova*
Macrophages		x	x
mast cells		x	x
memory dendritic cells			x
Monocytes	x		x
myeloid derived suppressor cells			x
natural killer cells CD56bright		x	x
natural killer cells CD56dim		x	x
natural killer cells		x	x
natural killer T			x
Neutrophils	x	x	x
NK cells activated	x		
NK cells resting	x		
plasma cells	x		
plasmacytoid dendritic cells		x	x
resting dendritic cells	x		
resting mast cells	x		
T cells	x		x
T cells CD4 memory activated	x		
T cells CD4 memory resting	x		
T cells CD4 naïve	x		
T cells CD8	x	x	x
T cells follicular helper	x	x	x
T cells gamma delta	x	x	x
T central memory cells		x	
T effector memory cells		x	
T helper 1		x	x
T helper 17 cells		x	x
T helper 2		x	x
T helper cells		x	
T regulatory cells	x	x	x

CIBERSORT and Angelova et al.'s Immunome provide valuable and promising methodology to characterise immune cells within the tumours. All the methods deliver some distinct and common immune cell types to be characterised. As shown in the Table 1.1 all of the methods proposed very specific immune cell types, which might be actually difficult to define their functional differences. Nevertheless, I explored utility

of two of them applying them to primary melanoma FFPE-derived gene expressions. More technical details of these methods and how I applied them are described in the next chapters.

Chapter 2

Materials and Methods

This chapter summaries the Leeds Melanoma Cohort from which the data used in my work was generated. Moreover, additional datasets, and standard statistical tests used in majority of the chapters of the thesis are described. The more advanced statistical and bioinformatic methods applied to only one chapter will be described in full details in the relevant chapter.

2.1 The Leeds Melanoma Cohort (LMC)

A cohort of 2184 primary cutaneous melanoma patients was recruited in the period 2000 to 2012 (67% recruitment rate) [114] by Julia Newton-Bishop and research staff. The participants were predominantly recruited from a geographically defined area (Yorkshire up to the river Tyne) with some additional recruitment from specialist centres in order to recruit a larger number of people with rare tumours e.g. acral tumours, and people having sentinel node biopsies.

Lifestyle data, co-morbidities and drug exposures including supplements were collected from participants by questionnaire. Clinical data were extracted from medical records and in order to derive melanoma specific survival (MSS), follow up was conducted using the national cancer registry, medical records in primary, and secondary care and by annual questionnaires completed by consenting participants. The median follow-up for the cohort at the time of the analysis of the data presented in this thesis was 7.5 years. The LMC was reviewed by the North East – York Research ethics committee (Jarrow, Tyne and Wear, UK) and received ethical approval - MREC 1/3/57 and PIAG 3-09(d)/2003. The histological features of the tumours were reported by clinical dermatopathologists. Dr Sally O'Shea from our research group (former clinical PhD student) also carried out a single observer review of as many of the slides from tumours used to generate the transcriptomes as were available. Dr O'Shea derived a number of measures of the tumour and its environment which were found to be reproducible in inter and intra-observer agreement studies.

2.1.1 Sampling of the FFPE primary tumour blocks

Sampling of the FFPE primary tumour blocks was performed predominantly by Dr Jonathan Laye. 5µm primary tumour sections were cut using a microtome and mounted onto Superfrost™ glass slides (Solmedia, Romford, UK). The sections were dried overnight and then hot-plated for a minimum of 20min at 70°C and subsequently following a standard protocol Haematoxylin and Eosin (H&E) staining was performed.

All the resulting H&E-stained sections of primary tumours were reviewed using light microscopy by Prof. Julia Newton-Bishop and Dr Jonathan Laye. Using a fine-tipped marker pen, the area of the slide corresponding to the invasive tumour was marked for sampling, consistently ensuring that the region selected had the least stromal content and the least infiltrating lymphocytes. The intent was to ensure comparability between tumours. Necrotic areas were avoided. The marked slide was then used to orientate the tumour block so that it could be sampled using a 0.6mm diameter microarray needle (Beecher Instruments Inc, USA) taking a core horizontally. Depending upon the size of the primary tumour, sampling was performed to yield up to 2 tissue cores. Although the 2184 participants were ascertained from the population rather than from hospital (and therefore potentially biased) series, the protocol demanded that blocks should not be destroyed by sampling less the patient need their sample for clinical testing later. Hence we did not sample approximately half the blocks. Unfortunately, two batches of arrays were lost attributed to technical difficulties related to unexpected loss of reagent function over time. The requirement to avoid block destruction for living participants did mean that there was some bias to sampling of thicker tumours. The mean thickness for those sampled was 3.0mm compared with 2.2mm for the whole cohort.

The DNA and mRNA were isolated from the cores by Dr Jonathan Laye and Dr Ross Jewell, preceded with tissue deparaffinisation and digestion. DNA was extracted using Qiagen AllPrep® DNA/RNA FFPE kit and Qiagen QiAamp® FFPE tissue kit. RNA was isolated using High Pure paraffin RNA kit (Roche Diagnostics Ltd, Burges Hill, UK). All the nucleic acid extraction procedures were performed as recommended in manufacturers' protocols. The extracted DNA was used for generation of Copy Number Variation data and mutational (*NRAS*, *BRAF*) data, while mRNA was used for the generation of gene expression data.

2.1.2 Measurement of the gene expression

The mRNA was extracted from 820 tumour cores (703 unique patients and 117 duplicates). Gene expression was assayed by an external service provider ServiceXS (Leiden, Netherlands) using the Whole-Genome DASL (cDNA-mediated Annealing, Selection, extension and Ligation) HT12v4 assay (Illumina®) in 3 batches. This array was generated for use on formalin fixed tissue in that the probes were short: about 50 bases in length in order to enable profiling of partially degraded RNA, which resulted due to FFPE archival methods. The generated raw expression data were normalised by Dr Jérémie Nsengimana. Firstly, the data were background-corrected and quantile-normalised in R using the package Lumi [115]. Singular value decomposition (SVD) was applied in package Swamp [116] in R in order to evaluate the association between the top principal components and technical variables: batch, chip, age of FFPE block, and RNA concentration. The variables that were found to be associated with these top components were adjusted out, and SVD was applied again with and without data permutation to evaluate the remaining "biological" variability in the data. In order to detect outliers normalized full intensity plots were examined. Among sample duplicates, in the final data set the sample with the highest number of detected genes was retained. Generally, the median and interquartile range of genes that were detected per sample were 14,784 ($P<0.05$, range 14,153– 15,304), which was consistent with other studies, which used DASL arrays in melanoma [117][118][119][120]. Some of the DASL probes were designed to hit:

- all splice isoforms of a gene
- or specific splice isoform of a gene, for which multiple isoforms are known to exist
- or one known single splice isoform
- or multiple isoforms - more than one and fewer than all of the splice isoforms of a gene,

however, the majority of probes were the ones targeting all splice isoforms. For some of the genes more than one probe was designed and for most of my analyses I examined all the probes, then chose the probes with the highest proportion detected (throughout the samples at the $P<0.05$), and eventually annotated them to the gene names. However, in the application of one of the immune cell scoring methods

(Angelova et al.) to the LMC transcriptome I filtered out the probes that were designed to hit only one specific isoform of a gene. I did not apply this filtration to another immune cell scoring method - CIBERSORT because this idea emerged long after I decided to drop the latter method.

The LMC gene expression data are accessible from the European Genome-phenome Archive (EGA) - accession number: EGAS00001002922. These data were the principal data used in this thesis.

2.1.3 Mutation data

2.1.3.1 *NRAS and BRAF mutations from pyrosequencing*

The DNA was used to carry out the *NRAS* and *BRAF* mutation screening using pyrosequencing [121] by Dr Philip Chambers, Genomics facility, Leeds. The pyrosequencing primers were designed to detect *BRAF* codon 600, *NRAS* codon 61, and *NRAS* codons 12 &13. These analyses were performed prior to my arrival in Leeds. During my PhD we collaborated with the David Adams' laboratory at the Wellcome Sanger Institute to mutation screen a large number of genes see below.

2.1.3.2 *Gene panel mutation screen from next generation sequencing (NGS)*

The mutation data were generated at the Wellcome Sanger Institute in Cambridge using DNA extracted from the tumours in Leeds. The data were pre-processed by Sofia Chen (Marie Skłodowska-Curie PhD student), supervised by David Adams. Briefly, from the LMC 521 samples were sequenced on Illumina HiSeq4000. Mutation data were only available from 319 of the 703 tumours from which transcriptomic data were generated (Figure 2.1) as the 0.6mm cores derived DNA had been exhausted in the process of making libraries for copy number data generation in Leeds.

To detect mutations Sofia used targeted capture custom design, with Agilent SureSelectXT baits diluted 1/24 (https://www.agilent.com/cs/library/catalogs/public/5991-7099EN.PDF).

The human reference genome assembly GRCh37d5 was used for mapping the reads. This step performed using BWA mem and the duplicates were marked using Picard MarkDuplicates. For SNPs the calling of mutations was completed using CaVEman

v.1.11.2 and the variant annotation was accomplished by using VAGrENT. In order to annotate functional consequences of the indels Pindel v.2.2.2 in together with VEP was used to. Variants with the low quality were excluded.

Moreover, variants which had both minor allele frequency (MAF) <0.10 and a coverage <30x within the tumour sample were filtered out. The mutational load was provided as three variables: mutation count per megabase, exonic mutation per megabase, and nonsynonymous mutation per megabase. For my analyses, which were mainly related to immune responses I used the mutation count per megabase, which could be considered as the most representative of neoantigen load (explained in Chapter 1). The mutational load per megabase per sample was obtained by dividing the number of mutations per patient by 5.2 (the sequenced regions were 5.2 megabases long). This variable, was categorized into tertiles, because these were predictive for melanoma specific survival, as initially showed by Sofia Chen's analyses (private communication).

2.1.4 Somatic copy number alteration (CNA) data

Dr Anastasia Filia (former PhD student) generated the whole-genome DNA libraries for the Next Generation Sequencing (NGS) using the NEBNext® UltraTM DNA Library Prep kit for Illumina® (indexed primers) (New England BioLabs, UK) in Leeds. NGS was run on Illumina GAII or HiSeq sequencer. The CNA data from NGS reads were generated by Dr Anastasia Filia and Dr Alastair Droop as described in Filia et al. (under revision, Scientific Reports, October 2018), however in my thesis the method of the CNA data analysis was slightly altered as compared to the previous one and it was carried out by my colleague Marie Skłodowska-Curie PhD Student, Joey Mark Santiago Diaz CNA data were generated from 303 FFPE tumours, including 276 for which the transcriptomic data were also available (Figure 2.1). From the sequenced tumour DNA samples, the copy number profiles were estimated from the next generation sequencing (NGS) output.

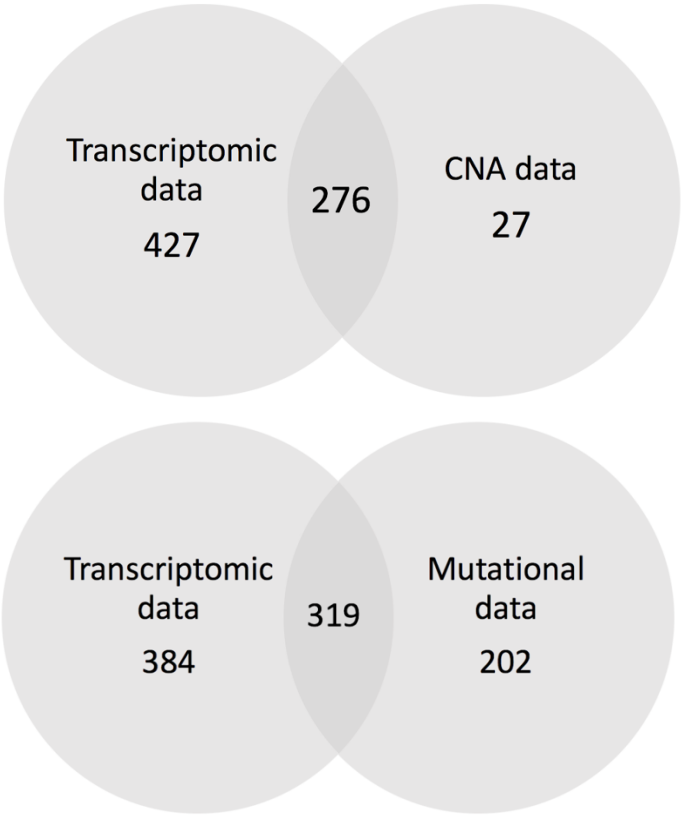


Figure 2.1: Venn diagrams representing the numbers of the samples available for transcriptomic and CNA or mutational data

2.1.5 Immunohistochemistry

Immunohistochemical staining was performed by Dr Jonathan Laye and myself using IntelliPath FLX detection reagents (MenaPath, A. Menarini Diagnostics, UK) according to manufacturer's protocol.

The tissue blocks for immunohistochemistry were chosen based on the availability of gene expression data from the tumour and the identification of blocks that could be resampled or were no longer needed for clinical purposes due to the patients' death due to any cause. This was therefore a small sample biased in terms of thickness and associated mortality given the proportion of samples used from patients who had died of melanoma. The samples were the only ones available however and as a reasonable proportion of people had died of causes other than melanoma were judged acceptable.

5µm tissue sections were cut and mounted on Superfrost Plus™ slides (Solmedia, Romford, UK), dried overnight and hot plated for 1hr at 70°C. The slides were immersed into an antigen retrieval solution (MenaPath Access Revelation, pH 6.4) and placed into

a pressure cooker to perform antigen retrieval by heat: the tissue sections were heated to 125°C sustained for 2min followed by cooling to 90°C which was then maintained for a further 10 seconds. The slides were then washed in warm MenaPath Wash Buffer to remove antigen retrieval solution and any remaining wax, and then in distilled water. The area of the tumours was circumscribed using a wax pen and the slides returned to MenaPath Wash Buffer. Each tissue section was treated with 100μl MenaPath Peroxidase Blocking agent for 20min to quench endogenous peroxidase activity within the tissue. After briefly rinsing in MenaPath Wash Buffer, the tissue sections were treated with 100μl MenaPath Background Blocking Agent with Casein for 20min to limit non-specific antibody binding. After further rinsing in MenaPath Wash Buffer, each tissue section was incubated with 100μl primary antibody, diluted in Zymed antibody diluent (Thermo Fisher Scientific, UK) in accordance with previously optimised conditions, and incubated at room temperature for 1h. The slides were then rinsed 3 times for 3min each in MenaPath Wash Buffer before the addition of 100μl MenaPath HRP-polymer to each tissue section and incubated for 30min at room temperature. The slides were again washed 3 times for 3min in MenaPath Wash Buffer and then treated with MenaPath Purple Chromogen for 10min. Slides were rinsed for 1min in tap water and then underwent counterstaining for 30sec in Mayer's Haematoxylin, followed by 1min rinse in running tap water. The tissue sections were 'blued in Scott's tap water for 2 min, rinsed again in running tap water for 1 min and underwent dehydration in in a series of 100% ethanol washes (4 washes of 3 min each) and 'cleared' in xylene (3 washes of 3 min each). Coverslips (Solmedia, Romford, UK) were applied onto the tissue sections using DPX mounting medium (Solmedia, Romford, UK).

A purple chromogen was chosen instead of brown in order to avoid confusion with melanin. However, we did feel ultimately that the disadvantage of this colour of chromogen is that occasionally if the staining is not strong enough it might be difficult to assess the positive nuclear staining which would be purple from unstained blue ones.

2.2 Data not originating from LMC

In addition to various histological and molecular data included in the LMC, in this thesis I have also analysed data that are independent to that cohort. Three such datasets

were used: The Cancer Genome Atlas (TCGA) metastatic melanoma [46], established melanoma cell line cultured in house and patient derived cell lines cultured at the University of Zürich.

2.2.1 The Cancer Genome Atlas - TCGA

I downloaded the gene expression (measured by RNAseq) and survival data from The Cancer Genome Atlas (TCGA) metastatic melanoma, produced using fresh frozen tissues. (http://www.cbioportal.org/data_sets.jsp, downloaded in 2016). I have chosen to analyse only the metastatic tumours, due to problematic survival data for primary tumour. Basically Dr Jérémie Nsengimana observed that these tumours had much worse prognosis than the metastatic ones, which was counterintuitive. In general, this study used tumours that were highly selected/biased, at very advanced stages and with limited clinical metadata.

Nevertheless, I used in total 339 metastatic samples, representing 80% of the total TCGA dataset at that time. The participants in this study were patients from different locations, hence it was not data from a population-based cohort like LMC. The overall survival (OS) was used in survival analyses performed in this dataset as there was no record of the cause of death. The follow up was as follows: Median=353 days, Min=12 days, Max=3798 days.

2.2.2 Established melanoma cell lines

Melanoma cell lines SkMEL28 (ATCC) and MeWo (courtesy of Professor Alan Melcher, formerly at Leeds, storing samples in the Leeds CRUK cell line bank) were cultured in our lab and their gene expression produced and pre-processed by Dr Anastasia Filia (former PhD student) and Dr Alastair Droop. I used these data to identify genes in the Angelova et al.'s Immunome which are expressed by melanoma cells and could therefore not be considered specific to immune cells. The total RNA was extracted following the Qiagen RNA mini kit (QIAGEN). The measurement of the gene expression was performed using the Affymetrix GeneChip® Human Genome U133 plus 2.0 Array by the service provider at the Patterson Institute, University of Manchester. The data pre-processing and quality control was performed by Alastair Droop using Robust

Multi-chip analysis (RMA) [122] and affyPLM [123] packages in R available in Bioconductor (https://www.bioconductor.org).

2.2.3 Patient derived melanoma cell lines from Zurich University

These data were produced and analysed by a colleague Marie-Curie PhD student Sabrina Hogan based at the University of Zürich, under the supervision of Prof. Mitchell P. Levesque.

Primary melanoma cells were isolated from clinical samples at the University of Zurich as described previously [124]. Melanoma cell cultures were obtained from remaining tumour material of patient biopsies after sampling for histological review, using the selective adherence method [124]. The protocol used to design these patient-derived cultures was adapted to discourage fibroblast cell growth and was reported to be successful in 70% of samples submitted to the live cell biobank.

Total RNA from melanoma cell cultures (N=103) was extracted using the Qiagen RNA mini kit (QIAGEN) according to the manufacturer's protocol. RNA capture was performed with the TruSeq RNA Library Prep Kit v2 (Illumina) and the library was sequenced on a HiSeq4000. RNA counts were quantified from single-end reads using STAR aligner [125].

2.3 Statistical methods

All the statistical methods were performed in STATA (SE) v14.1. Some of the manipulations of the data (e.g. transposition) were implemented in R v3.3.2 and RStudio v1.0.136.

2.3.1 Variables used in statistical analyses of the LMC samples

2.3.1.1 *Categorical variables*

- **AJCC stage** was used as described in the Introduction and was recorded in 3 levels: stage I, II, and III, N=695
- **Ulceration** was recorded as a binary variable: "yes" vs "no", N=703
- **TILs** as described in the Introduction: absent, brisk and non-brisk, measured by clinical histopathologists (N=553). The single observer Dr Sally O'Shea, (N=601) reported in the same way, absent, brisk and non-brisk, as described in the

introduction (Figure 1). Both of these measures were used separately as a cross-validation between the measures from the clinic and single observer. Dr O'Shea also generated the measures of lymphocytes infiltrate around the cored region, however for smaller number of samples, hence I excluded them in my analyses

- **Site of melanoma** was used as either binary ("Limbs" vs "Rest") or with 4 levels: "Limbs" (N=299), "Head and neck" (N=80), "Trunk" (N=233) and "Other" (N=90), in total N=702. "Other" represent rare, non-sun exposed tumours such as:
    - ear, nose, thorax (N=5)
    - acral (N=19)
    - anal (N=5)
    - cervix (N=1)
    - foot (N=18)
    - hand (N=5)
    - nodal with no known primary (N=3)
    - penis (N=3)
    - perineal (N=1)
    - subungual (N=14)
    - vaginal (N=3)
    - vulval (N=13)
- **Sex** was used as a binary variable (females, males), N=703
- **Smoking** was analysed as categorical variable as "ever" (N=324) vs "never" smokers (N=334); Total N=658 and "smokers" (N=543) vs "still smokers" (N=85)
- **Vascular invasion** assessed by the presence of melanoma cells within blood vessels recorded as "yes" vs "no", by clinical histopathologists reporting clinical samples N=627
- ***NRAS*/*BRAF* Mutation data:** *NRAS* mutant vs *BRAF* mutant vs "double wild type", N=575
- **Mutational load:** mutation count per Mega Base (MB) was categorised into tertiles, N=319

2.3.1.2 *Continuous variables*

- **Breslow thickness** was measured as described in the Introduction: median=2.3mm, min=0.33mm, max=20mm, N=692
- **Mitotic rate** was quantified as mitotic count within tumour per mm^2: median=3, min=0, max=83, N=596
- **Age at diagnosis** – age of the patient was recorded at diagnosis of primary melanoma: median=58.3, min=18.29, max=81.25, N=703
- **Duration of smoking** was recorded as a number of years smoked by participants: median=0, min=0, max=60.9, N=654. More than half of the patients had never smoked (see previous section)
- **Vitamin D** levels (nmol/L) in the blood were measured in the NHS biochemistry laboratory in Leeds, in a sample taken at recruitment (add average time from diagnosis to recruitment). Concentrations of 25-hydroxyvitamin D_2 and D_3 (nmol/L) were measured in 100 µL of cryopreserved serum by liquid chromatography tandem mass spectrometry. The levels were adjusted for season, by rescaling every patient's record to winter: median=39.5nmol/L, min=0nmol/L, max=191.2nmol/L, N=549. This adjustment was based upon the observation that levels are on average 20nmol/L higher in summer months than at the end of winter
- **Melanoma specific survival (MSS)** was used in survival analyses in LMC: median=6.3years, min=0.45years, max=14.3years, N=703

2.3.2 Statistical tests

Parametric and non-parametric tests were used depending on the distribution of a tested variable.

I used the Student t-test (parametric test) or Mann Whitney U (non-parametric test) for comparison of two groups. For more than two groups comparisons, I used the Analysis of variance ANOVA (parametric test) or the Kruskal-Wallis (non-parametric) test. For multiple pairwise comparisons, after the above tests, I used a post hoc Dunn's test.

I ran correlation analyses using the Pearson (parametric test) or Spearman's rank correlation (non-parametric) coefficients. For multiple testing, I utilised Bonferroni correction and Benjamini-Hochberg tests, depending on the required stringency and the research area common practice.

2.3.2.1 ***Survival analyses***

Survival analyses were carried out using Cox's proportional hazards regression (univariable and multivariable). Survival data were visualised using the Kaplan Meier approach.

The survival data consist of one observation per case with the data being either of the form: (i) case dies at time t since diagnosis (termed "failure") or (ii) case still alive at time (t) after diagnosis (censored observation).

In the analyses presented here, the failure event will be defined in each analysis but will usually reflect "death from melanoma". Cases who have died from causes at time t after diagnosis other than melanoma are censored at that time.

The Cox proportional hazards model assumes that the probability density (probability per unit time) for "dying" at time t (given that an individual survived until t) is equal to a hazard function, $h(t)$, times a function of predictor variables. For convenience, the effect of the predictor variables is modelled as $\exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_i x_i)$ where $(x_1, x_2, \ldots x_i)$ are predictor variables and $(\beta_1, \beta_2, \ldots \beta_i)$ are the weights associated with the corresponding predictor variable which are typically estimated during analysis by maximum likelihood; the estimates of β_1, β_2 etc. are log "hazard ratios" (HRs). The insight provided by Sir David Cox was that using the likelihood formulation and this precise model, the likelihood is independent of the unknown baseline hazard $h_0(t)$. HR estimated to be greater than 1 implies decreased survival (increased hazard) while HR < 1 implies improved survival. The significance of the Cox proportional hazards was estimated by the appropriate likelihood ratio tests with the statistical significance being assessed with the chi-square test.

Chapter 3

In silico characterisation of tumour immune microenvironment

3.1 Aims

- To evaluate, develop, and apply bioinformatic approaches to characterisation of immune cell subtypes infiltrating primary melanomas.
- To cluster primary melanoma transcriptomes based upon inferred immune cell scores.
- To identify immunologically distinct and prognostic primary melanoma subgroups.

In this chapter, I applied two of the three bioinformatic methods of immune cell quantification *in silico* described in the Introduction to the LMC transcriptomic dataset: CIBERSORT [113] and Angelova et al.'s Immunome [1]. As explained earlier, the Angelova et al.'s Immunome is an extension of the system described by Bindea et al. [108] and used by Dr Jérémie Nsengimana in our research group previously [76]. The scores generated based on Bindea et al.'s Immunome were used for some selected comparisons in this thesis.

3.2 Methods

3.2.1 Application of the CIBERSORT method to the LMC transcriptome

CIBERSORT - a computer algorithm of gene expression deconvolution to infer immune cell infiltrates [113] was developed using Affymetrix data and has not yet been evaluated in other settings. The first aim of this chapter was to apply it to FFPE tumour array-based transcriptomes (DASL HT12.4). The algorithm can be used interactively online or as a downloadable R package. As described in Chapter 1, it is based on linear support vector regression (SVR) machine learning to deconvolve the cell mixture in gene expression with data resampling to derive an empirical p-value showing the success or failure of the algorithm when applied to each sample. As a default, CIBERSORT uses as reference the LM22, a transcriptomic signature of 22 purified immune cells (547 genes) and it estimates the relative proportion of each of those cells,

with the sum of those proportions adding up to 1. The method failure may occur for example if there are no immune cell infiltrates at all in the tumour.

The principal equation that CIBERSORT algorithm utilises is:

$$\boldsymbol{m} = \boldsymbol{f}\ \boldsymbol{x}\ \boldsymbol{B}, \text{ where:}$$

$\boldsymbol{m}$ = mRNA mixture (measured gene expression in the tumour)

$\boldsymbol{f}$ = vector of the unknown (relative) fractions of each cell type in the tumour

$\boldsymbol{B}$ = GEP (gene expression profile) matrix of the 22 pure cells on the 547 reference genes, the "LM22" signature.

Note that CIBERSORT allows the user to define their own signature of cells to investigate (which can be fewer or more than 22 and may include non-immune cells). It is assumed that the size of matrix $\boldsymbol{B}$ is much smaller than the size of matrix $\boldsymbol{m}$ (i.e. many more genes measured overall than the number of genes characterising the cells of the signature).

I applied the online version of this method (http://cibersort.stanford.edu/) to the LMC dataset, with the default LM22 signature. The deconvolution p-value was produced using Monte Carlo sampling, with 1000 permutations as recommended. For each tumour analysed, the tested null hypothesis was that there were no immune infiltrates of the LM22 signature (see LM22 signature components in the Introduction, Table 1). The algorithm calculates the Pearson correlation between the input gene expression matrix ($\boldsymbol{m}$) and its expectation under the null hypothesis using the inferred cell proportions ($\boldsymbol{f}\ \boldsymbol{x}\ \boldsymbol{B}$). A significant p-value means that the inferred cell proportions are unlikely to be random observations, while a non-significant p-value means that the inferred proportions may well represent noise.

After preliminary checks, I found that the LM22 reference signature had data on a different scale to our dataset, due to different approaches to pre-processing and transformation. For LM22 the maximum value was 42,851.29 and the minimum 461.191. The LMC data were on log2 scale with a maximum of 17. Although a simple power transformation of the LMC dataset would have made the two scales comparable, they would not have been entirely similar (the new maximum value would have been 2^{17}=131072, 3 times larger than in LM22).

To have completely similar scales prior to running CIBERSORT, I collaborated with Dr Victor Boudara (Mathematics department, University of Leeds), who devised the formula described below and I applied it to the LMC dataset. In essence, I applied two data transformations:

a) The first was to transform one interval (LMC data scale) *[y_1 ; y_2]* into a second interval (LM22 scale) *[x_1 ; x_2]* solving the following system of equations for the constants (*a, k*):

1)

$$\begin{cases} k(y_1)^a = x_1 \\ k(y_2)^a = x_2 \end{cases}$$

Dividing x_1 by x_2 gives:

2)

$$\frac{x_1}{x_2} = \frac{(y_1)^a}{(y_2)^a}$$

$$a = \frac{\log\left(\frac{x_1}{x_2}\right)}{\log\left(\frac{y_1}{y_2}\right)}$$

After obtaining the value of *a*, to obtain k I used:

3)

$$k = \frac{x_1}{(y_1)^a}$$

After obtaining both constants *a and k* any value, *y,* contained in the LMC dataset interval *[y_1 ; y_2]* can be transformed into, *y'*, contained in the LM22 signature interval *[x_1 ; x_2],* by using:

4)

$$y' = ky^a.$$

b) The second data transformation was scaling both the LMC gene expressions and the LM22 gene expressions, using standard normal transformation, setting the mean to 0 and the standard deviation 1. These analyses were conducted in R.

3.2.2 Application of Angelova et al.'s Immunome

The Bindea et al.'s Immunome [108] and its later version by Angelova et al. [1] were developed by using purified immune cells and different tissue types as controls, including colorectal cancer cells, because the authors are colon cancer researchers. Before applying this approach to melanoma FFPE samples, I performed gene and cell filtration to adapt the compendium to the melanoma context. The reason for doing this was to ensure as much as was possible that the measured immune cell signals originated effectively from the immune cells, not from melanoma cells or melanocytes[76]. I applied three filters to the Angelova et al.' Immunome.

Firstly, I removed the genes that were expressed among the top 25% across the whole genome in a melanocyte cell line (GSE4570) and in our cultured melanoma cell lines, MEWO and SK-MEL28 (cells cultured and RNA extracted by Dr Anastasia Filia, former PhD student in the lab). SK-MEL28 gene expression was measured in 8 replicates while those from MEWO were generated in 7 replicates. The replicates were averaged within each dataset and genes were ranked based on this average expression. We considered genes that ranked among the top 25% most expressed in each dataset as evidence of their substantial expression by melanoma cells with therefore little support that they are immune cell specific. Our hypothesis here is that we do not expect immune cells in melanoma cell lines. These genes were then removed from the Angelova et al.'s Immunome.

Secondly, I calculated the proportion of genes removed by this filtering step per immune cell type and eliminated the whole cell type altogether if less than 10% of its genes remained. The exception to this was only if I could find concordant evidence from published literature that the reduced number of genes was indeed characteristic of the immune cells in question.

Thirdly, for the remaining immune cells I tested reciprocal correlations between genes assigned to each cell type, as the Immunome compendium approach was based on genes co-ordinately expressed within a particular cell subtype, i.e. a positive correlation between those genes was expected. I reasoned that, for each immune cell subtype, gene expression correlations that were observed in the blood might differ from what we detect in the tumours and if some genes were negatively correlated with

the majority of others, then the evidence of their specificity to that cell type is weakened and they should be removed. In the majority of cell scores, most of genes were co-expressed as expected but in a few cases there were negative correlations and I removed those genes negatively correlating with the majority. The negative or very weak correlations could also be a reflection of technical issues with probes, which failed the detection of the particular genes, particularly in mRNA degraded by the tissue fixation in formalin. This problem could be partially addressed by testing these correlations using a dataset generated from fresh frozen tissue with less degraded mRNA, such as TCGA. The correlation plots between all genes within each immune cell type were constructed in R using "corrplot" package [126][127]. This analysis was conducted in both the LMC and the TCGA dataset (for the replication).

After applying all above described filters, I devised a score for each immune cell type, calculated as the mean of expression values of all genes attributed to that cell, after z-score normalization of the log2 transformed gene expression data as described before [76]. The scripts to create the immune cell scores were written in Bash (Unix shell) with the help from Dr Victor Boudara and then applied in STATA. The same set of genes per immune cell type was applied to LMC and TCGA.

3.2.3 Consensus clustering of LMC tumours based on the immune cell scores

Clustering is a grouping/subdivision of the data (objects) based on their similarities and it can be performed on genes- or samples basis. It assigns the most similar objects to the same group and the most dissimilar ones to distinct groups. Cluster analysis helps to find functional structures within large datasets on a comprehensive level, therefore it became almost a standard method for simultaneously analysing expressions of multiple genes in large experiments [128]. The genomics research field has expanded regarding the development of various classes/algorithms of clustering. For example, one very commonly clustering type is hierarchical clustering, which produces a dendrogram (graphical representation) of similarities between samples and/or genes for example with respect to gene expression values [129]. The dendrogram can be cut an any level to set the final number of clusters after a visual inspection.

The other widely used algorithm is K-means, which is based on partitioning the data and dividing it into a predetermined number of clusters [130]. It has been shown that

K-means usually outperforms hierarchical clustering in finding the optimal number of clusters [131][132][133]. The most frequently used similarity measures are Euclidean or Pearson correlation distances and both perform reasonably well for clustering methods [131][132].

Although clustering provides good solutions to subdivide the data, it faces some difficulties in the identification of non-overlapping groups/clusters and their final number. In fact, visual inspection of dendrograms in HC and predetermining the number of clusters in K-Means can be very subjective. It also encounters the problem of overfitting, when the sample size is relatively small in comparison to the number of analysed genes. To overcome this problem, consensus clustering was introduced, which is a resampling based method of identifying sample classes in microarray data [134]. Subsets of data are resampled multiple times (for example N=1000) from the original dataset and a user-selected clustering algorithm is applied to each subset, assuming a growing number of clusters from 2 to a user-defined maximum K. At each analysis iteration, each sample is placed in one of the clusters and a consensus score is produced, reflecting the sample belonging to each cluster. In an ideal situation a sample belongs to one and only one cluster. All analysis iterations should effectively classify that sample in its cluster, i.e. no iteration should classify it in other clusters. In other words, the sample is put in its cluster with 100% consensus and in other clusters with 0% consensus. In reality however, samples will have varying levels of consensus scores in each cluster, but after consensus clustering each sample is finally assigned to its most likely cluster (one with the highest consensus score). A consensus score is the proportion of iterations classifying a sample in a particular cluster and it is stored in a NxN matrix, where N is the total number of clusters.

A number of metrics are generated in consensus clustering to aid the decision on the optimal number of clusters: the consensus heatmap, the consensus cumulative density function (CDF) and the delta graph of the area under the CDF [134]. The consensus heatmap is a results of clustering of the consensus matrix values using hierarchical clustering and the dendrogram is cut at the chosen k threshold. The heatmaps are symmetrical and their “cleanliness” at each k indicates the stability; the blue colour designates high consensus whereas white no consensus (samples are always classified in different clusters). CDF(x) for a consensus matrix indicates the proportion of pairs

for which consensus score is less than or equal to x $(0 \leq x \leq 1)$ for each k. The optimal number of clusters is the CDF curve for k with the highest proportion of "0" and "1". The delta graph of the area under the curve is the change in the area under the CDF when the number of clusters increases. Using this graph the optimal number of clusters is when the change of the area under the CDF is the smallest when the number of k increases.

I have used the same parameters for consensus clustering (listed below), as proposed previously in consensus cluster analysis [134] within ConsensusClusterPlus [135] package in R to generate tumour subgroups. I decided to proceed with the same parameters, because, it was shown that they result in obtaining the most stable clusters using the gene expression data. I chose K-means as an algorithm for clustering with max K=12, Euclidean distance metric, 5000 repetitions, 80% genes and tumour resampling. Consensus heatmap, CDF and delta CDF plots were used to define the optimal number of clusters.

After sample classification, I tested the association between the obtained clusters and melanoma specific survival using Cox regression and likelihood ratio tests. The immune cell scores were plotted on a heatmap alongside the identified clusters to visualise their association ("pheatmap" package in R).

3.2.3.1 *Comparison of newly obtained immune signature with existing melanoma signatures*

We previously used consensus clustering in our group to identify 6 consensus immunome clusters (CICs) in the 703 LMC tumours based on selected immune genes [76]. For the current analysis, I chose to apply a dimensionality reduction to the genes by using them to calculate scores of 27 immune cells (see previous section) and then use these scores in consensus clustering to identify immunologically different tumour subtypes. I hypothesised that dimensionality reduction prior to clustering would delineate larger and more prognostic subgroups, allowing for enough power for downstream analyses aimed at biological characterisation. This hypothesis was driven by the observation that among the 6 CICs we reported previously [76], 5 had similar profiles of survival (good prognosis, but distinct histological features) and only one was different from them (poor prognosis). However, when the CICs prognosis was adjusted

for AJCC stage the two low immune CICs had a poor outcome. I tested the agreement between the 6 CICs and the new subgroups using Cramer's V statistic.

Back in 2015, consensus hierarchical clustering of 1,500 gene expressions with the highest variance was used by The Cancer Genome Atlas (TCGA) research network to classify 329 melanoma samples. Data were generated using RNA-seq from fresh frozen tissue, mostly from metastatic lesions. Three molecular classes were identified: "MITF low", "keratin", and "immune", with the most favourable survival for the "immune" and least for "keratin" [46]. In this thesis, I will call these classes "TCGA classes". Dr Nsengimana applied the nearest centroid method [136][137] to classify our 703 primary melanomas into these TCGA classes. I assessed the intersection between these TCGA classes and my newly identified clusters using the Cramer's V statistic.

3.2.4 Replication of obtained clusters on independent dataset (TCGA)

In order to assure that the consensus clustering results generated from the LMC were not over-fitted, they needed replication on an independent dataset. For this purpose, I utilised the TCGA metastatic melanomas, restricting the analysis to the metastatic samples (N=339 samples, 80% of the total dataset). These data were downloaded from the cBioPortal website (http://www.cbioportal.org). The reason to exclude TCGA primaries is because we have discovered in our group (Dr Nsengimana) that their survival data records were inaccurate (primary disease was associated with better outcome than metastasis). They could therefore bias the survival analysis of clusters.

To replicate clustering analysis in TCGA, I applied the nearest centroid method [137] in R. Firstly, I obtained the centroid vectors by averaging scores of each immune cell subtype within each cluster from LMC. Secondly, I scored each immune cell subtype for each TCGA sample in the same manner as I had done in the LMC. Thirdly, I calculated the Spearman correlation coefficient between each TCGA sample with each immune cluster centroid. Finally, the sample was assigned to the immune cluster with which it had the highest correlation. Subsequently, I tested the difference in overall survival (OS) between the clusters by using the Cox proportional hazard regression.

3.2.5 Clinico-histopathological characterisation of the obtained clusters

I analysed the biological and clinical validity of the clusters by comparing various clinico-histopathological variables recorded in our cohort (LMC). I tested the differences among the clusters using the chi-square test for categorical variables and the Kruskal-Wallis test for continuous variables. The following variables were tested:

- Age at diagnosis (continuous, years)
- Site of melanoma (%) (Limbs, Head, Trunk vs Rare, explained in Chapter 2)
- Sex (% males)
- *BRAF*-mutated (%)
- *NRAS*-mutated (%)
- Ulceration status (%)
- Breslow thickness (continuous, mm)
- Mitotic rate (continuous, count/mm2)
- AJCC stage (%) (I vs II vs III)
- TILs (%) (Brisk, Non-brisk, unclassified, no TILs), defined by dermato-pathologists and single observer from the research group, Dr S O'Shea
- Smoking (% ever smoked)
- Season-adjusted serum vitamin D at recruitment (continuous, nmol/L)
- Mutational load (3 groups %).

3.3 Results

3.3.1 Application of CIBERSORT to LMC transcriptomes

Without scale transformation (i.e. using the log2 transformed gene expression in the LMC and raw data in the LM22) CIBERSORT was able to successfully deconvolve only 32% of the samples (226 out of 699) at $p \leq 0.005$; 37% (259 out of 699) at $p \leq 0.05$; and 41% (290 out of 699) at $p \leq 0.1$. At the time of CIBERSORT exploration only 699 tumour data were available. The p-value threshold of 0.005 is the software recommended default [138] but I noted that relaxing this threshold did not increase substantially the number of deconvolved samples, therefore I remained at the recommend threshold.

To understand why CIBERSORT produced those results, I assessed the immunological characteristics of the samples deconvolved at $p \leq 0.005$ compared to those not well-

deconvolved. All the LMC samples have been previously classified in 6 consensus immunome clusters (CICs) based on the Bindea et al.'s Immunome [108][76]. The 6 CICs were in strong agreement with TILs reported by the pathologist: the high and intermediate immune clusters (CIC 2, CIC 3 and CIC 5) had highest number of tumours with evidence of brisk TILs while the low immune cluster (CIC 4, CIC 1 and CIC 6) contained the highest number of tumours with "absent" TILs [76]. Upon applying CIBERSORT, the vast majority of successfully deconvolved tumours were in the high immune cluster (CIC 2) and to some extent in the intermediate immune clusters (CIC 3 and CIC 5) (Table 3.1). The majority of poorly deconvolved samples were in the low immune clusters (CIC 4, CIC 1 and CIC 6). The contrast was striking between high immune CIC2 (136/145 well deconvolved) and all low immune CICs (only 5/70 deconvolved in CIC1, 3/171 in CIC 4, 5/100 in CIC6) (Table 3.1).

Table 3.1: Matching patients with CIBERSORT deconvolution status at p≤0.005 to the six "Consensus Immunome Clusters" without scale transformation

	Consensus Immunome clusters [76]					
	CIC 1	***CIC 2***	***CIC 3***	***CIC 4***	***CIC 5***	***CIC 6***
Number of deconvolved samples at P<0.005	5	**136**	42	**3**	28	5
Number of non-deconvolved samples	65	**9**	41	**168**	77	95

These results suggested that CIBERSORT can accurately estimate the proportions of immune cell subpopulations when there is a relatively high level of infiltration in the tumour and is unreliable where there are only a few or no immune cells present in the tumour.

The result from the first scale transformation (equation 4) produced a slight improvement in the total number of deconvolved samples increasing the number only modestly, by 38 samples at deconvolution threshold p≤0.005 and 48 samples at deconvolution threshold p≤0.05. The second data transformation (standard normal) resulted in 30 more samples deconvolved at p≤0.005, 149 additional samples were deconvolved at 0.05 level (totalling 408/699).

3.3.2 Application of Angelova et al.'s Immunome

Before applying this method, I applied QC filters to select the genes most representative of cell subtypes (Figure 3.1). The first gene filter removed 354 genes (genes highly expressed in melanoma cell lines), leaving 458 genes in the list of genes present on the Illumina HT12.4 DASL array used in this study, representing 30 distinct immune cell types (Subset 1, Figure 3.1). The second filter (removal of cell types where the first filtration eliminated >90% of the genes) resulted in the elimination of effector memory CD4+ T cell, activated CD8+ T cell and activated CD4+ T cell scores (Subset 2, Figure 3.1, Appendix A.1.1). The plasmacytoid dendritic cell score (pDCs) however was retained despite having only 1 specific gene (*IL3RA*), because this gene was also unique in representing the pDCs in the previous version of the Immunome compendium [108] and is well-known to be highly expressed in pDCs [139][140]. The gene reciprocal correlations per immune cell were not always positive in the LMC (Figure 3.2, Appendix A.1.2) and eventually the gene filtration based on negatively correlating genes in LMC resulted in a final number of 376 genes representing 27 immune cell subsets (Subset 3, Figure 3.1, Appendix A.1.3).

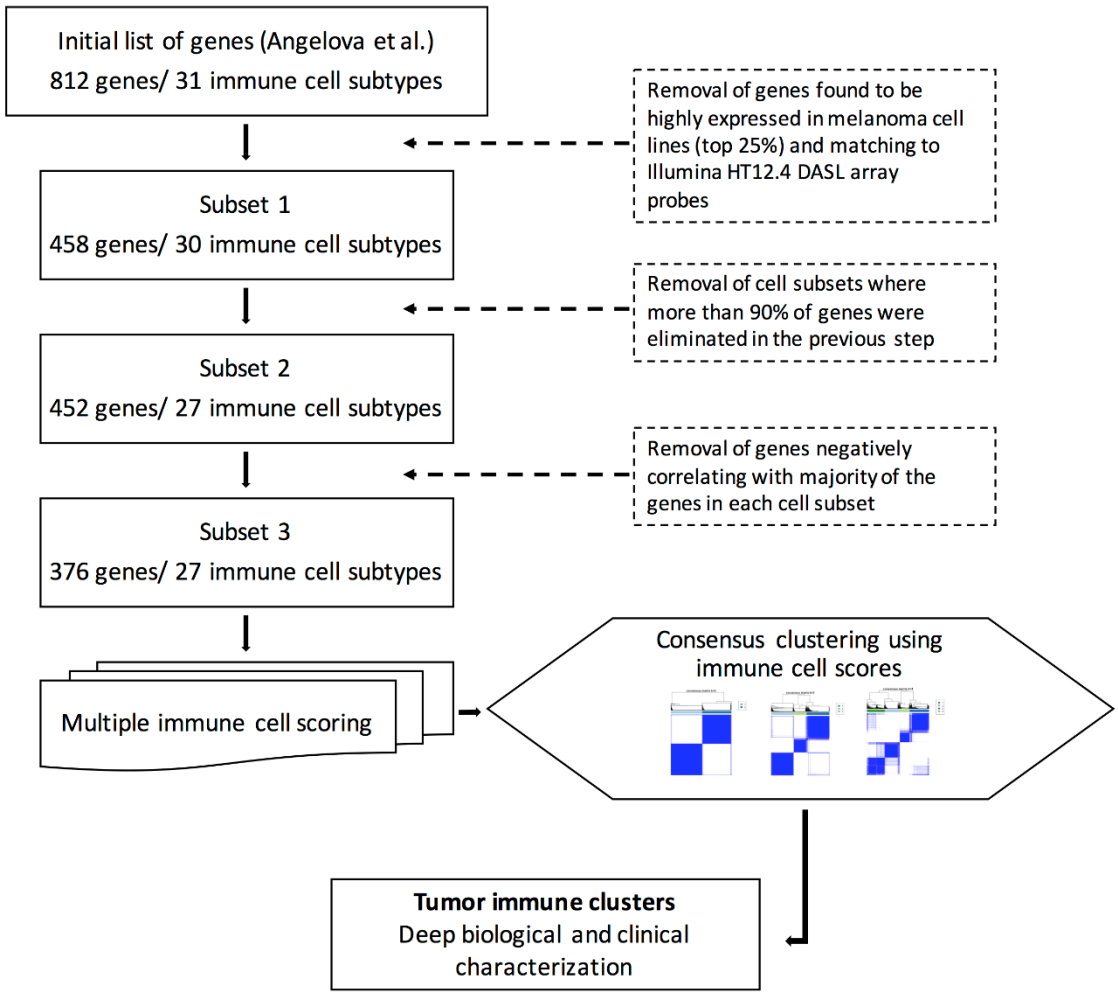


Figure 3.1: QC pipeline for devising the list of genes indicative of specific immune cells infiltrating melanoma

In the TCGA dataset, the reciprocal correlation matrices demonstrated fewer negative correlations for the majority of the immune cells (Figure 3.2, Appendix A.1.2). To further compare correlation matrices in LMC and TCGA datasets using Angelova et al. [1] and Bindea et al.'s [108] Immunome lists, I chose a few cells that are often described in literature by a single strong marker, such as *FOXP3* for Tregs [141] and *GATA3* for Th2 [142] but assigned several "specific" genes in the Angelova et al.'s Immunome. Figure 3.2 illustrates the correlation between tumour expressions of Treg genes by the Angelova et al.'s Immunome in LMC and TCGA. Bindea et al.'s Immunome used only *FOXP3* to represent Tregs. Figure 3 represents the correlation of the genes describing Th2 proposed by Bindea et al. and Angelova et al.'s Immunomes in LMC and TCGA.

As it can be seen from Figure 3.2, *FOXP3,* a widely accepted marker of Tregs, negatively correlated with majority of other Treg genes in the LMC, but not in TCGA. This might be an example of the failure of the probe detecting this gene on DASL array. Therefore, prior to the calculation of the Treg score I removed *FOXP3*, and other genes showing a negative correlation the majority (*MADCAM1,* and *RYR1)*. In general, there was a reasonable number of positively correlating genes in the LMC, and in most cases these correlations were much stronger in TCGA.

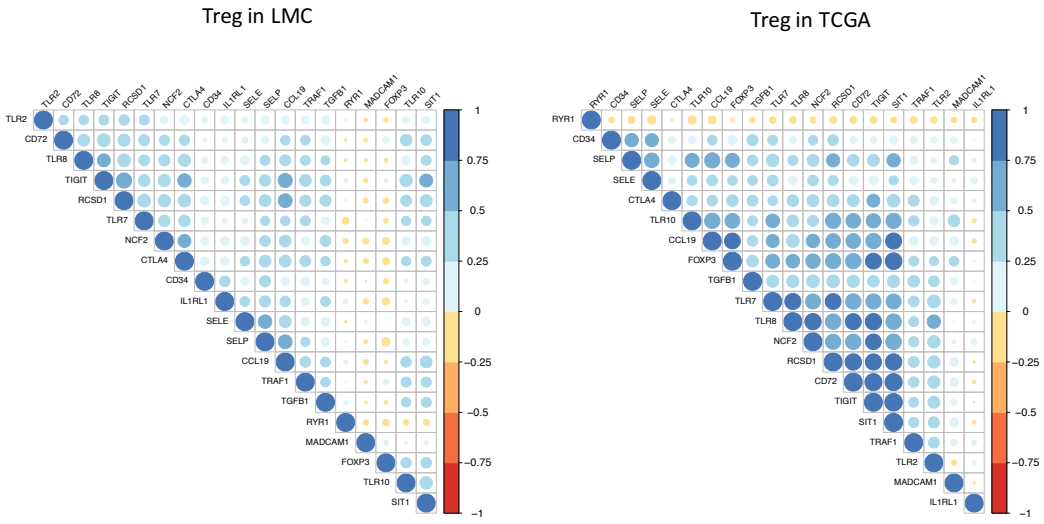


Figure 3.2: An example of correlation matrices of genes for regulatory T cells in the LMC and TCGA datasets

For the Th2 comparison, similar observations were made regarding the quality of the data (Figure 3.3). Overall, the correlations were stronger in the TCGA dataset, for both Angelova et al. and Bindea et al.'s Immunomes. When comparing the genes for scores generated based on Angelova et al. versus Bindea et al.'s Immunome in LMC and TCGA, it could be observed that *GATA3* the marker for the Th2 cell type was negatively correlated with majority of the genes in the Bindea et al.'s Immunome both in LMC and TCGA, however it was not the case in the Angelova et al.'s Immunome. I removed the following genes by the correlation filtration from the Th2 gene specific list: *RMBS3*, *HTR2B*, *ASB2*, *DNAJC12*, and *TMPRSS3*.

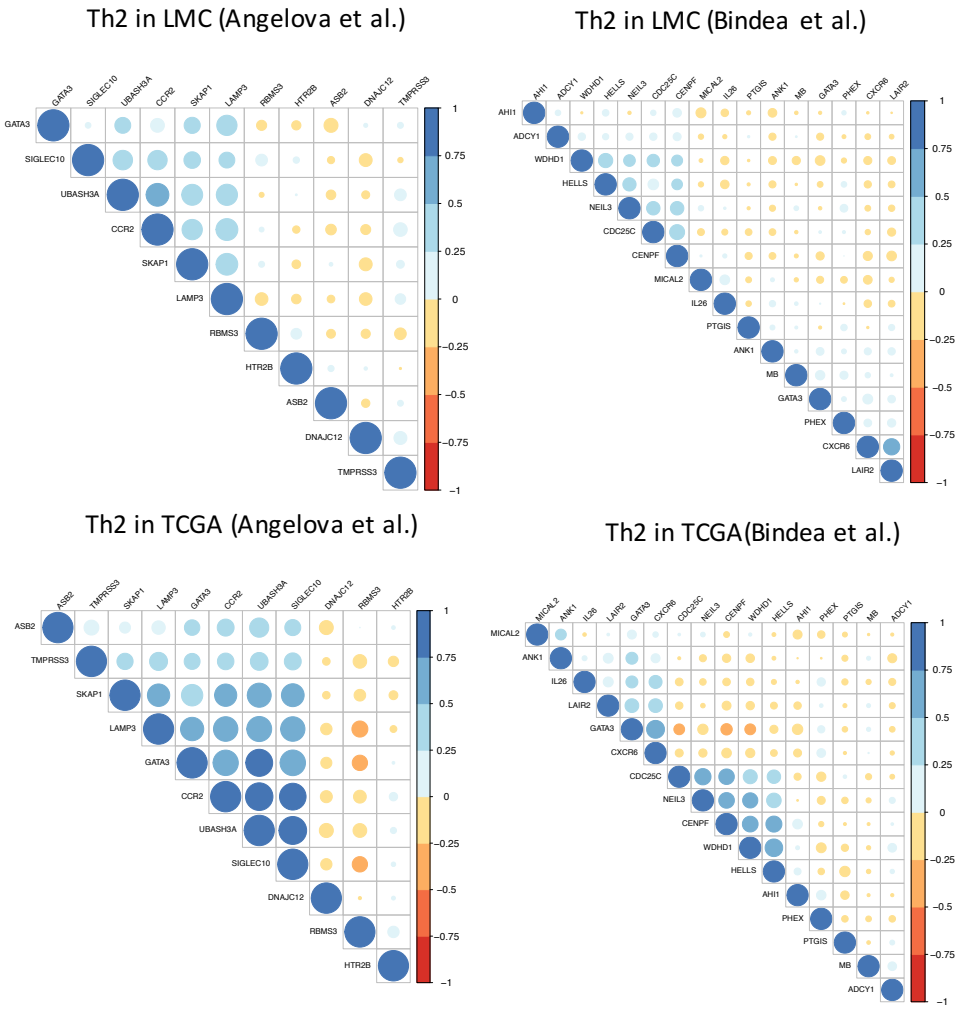


Figure 3.3: An example of correlation matrices of genes for Th2 cells in the LMC and TCGA imputed by Angelova et al. and Bindea et al.'s Immunomes

After these QC steps, I generated a single score for each immune cell subtype by combining tumour expression values for the remaining genes (see Methods). Then I tested the association between each immune cell score with melanoma specific survival (MSS) (univariable analysis) and the results revealed that in the great majority the immune cell scores (17 out of 27 in the LMC) were associated with improved survival after Bonferroni correction for multiple testing (P<0.002) and these observations were also replicated in TCGA (Table 3.2).

Table 3.2: Melanoma Specific Survival (MSS) for each cell type in the LMC and overall survival (OS) in TCGA, using univariable Cox proportional hazard model

Significant associations in the LMC after multiple-testing correction by Bonferroni method are shown in bold.

	LMC (primaries)				*TCGA (metastases)*			
Cell type	***HR***	***P***	***95% CI low***	***95% CI high***	***HR***	***P***	***95% CI low***	***95% CI high***
Activated B cells	**0.60**	**$1.18x10^{-5}$**	0.48	0.75	0.63	$2.79x10^{-6}$	0.52	0.76
Central memory CD4	**0.54**	**$7.31x10^{-6}$**	0.41	0.71	0.52	$5.46x10^{-5}$	0.84	1.27
Central memory CD8	0.72	$9.58x10^{-3}$	0.56	0.92	0.55	$4.33x10^{-7}$	0.89	1.44
Cytotoxic cells	**0.72**	**$3.35x10^{-4}$**	0.60	0.86	0.65	$1.91x10^{-6}$	0.38	0.72
DC	**0.51**	**$4.27x10^{-6}$**	0.39	0.68	0.53	$1.81x10^{-6}$	0.44	0.70
Effector memory CD8	**0.63**	**$2.97x10^{-4}$**	0.49	0.81	0.54	$1.49x10^{-8}$	0.54	0.78
Eosinophil	1.39	$2.13x10^{-2}$	1.05	1.84	0.72	$5.97x10^{-2}$	0.41	0.69
iDC	0.88	$2.87x10^{-1}$	0.68	1.12	0.56	$1.85x10^{-6}$	0.26	0.76
Immature B cells	**0.66**	**$7.16x10^{-4}$**	0.51	0.84	0.61	$1.12x10^{-5}$	0.79	1.09
Macrophages	0.71	$1.20x10^{-2}$	0.54	0.93	0.68	$2.24x10^{-3}$	0.51	1.01
Mast cells	**0.51**	**$3.90x10^{-5}$**	0.37	0.70	0.56	$3.50x10^{-3}$	0.49	0.76
MDSC	**0.64**	**$9.27x10^{-5}$**	0.51	0.80	0.62	$1.58x10^{-5}$	0.44	0.67
Memory B cells	1.03	$8.11x10^{-1}$	0.81	1.31	0.65	$2.06x10^{-3}$	0.50	0.77
Monocytes	0.94	$5.74x10^{-1}$	0.75	1.17	0.65	$7.30x10^{-5}$	0.54	0.87
Neutrophils	0.68	$4.18x10^{-3}$	0.52	0.88	0.49	$4.83x10^{-7}$	0.44	0.69
NK	**0.68**	**$2.39x10^{-4}$**	0.55	0.83	0.64	$4.16x10^{-5}$	0.38	0.83
NK56 bright	**0.71**	**$8.60x10^{-4}$**	0.58	0.87	0.55	$4.08x10^{-7}$	0.49	0.85
NK56 dim	0.78	$3.39x10^{-2}$	0.62	0.98	1.66	$1.14x10^{-3}$	0.53	0.81
NKT	**0.67**	**$2.43x10^{-4}$**	0.54	0.83	0.48	$3.04x10^{-6}$	0.52	0.80
pDC	0.91	$2.09x10^{-1}$	0.79	1.05	0.77	$1.13x10^{-3}$	0.45	0.69
T cells	**0.55**	**$4.58x10^{-6}$**	0.42	0.71	0.55	$2.36x10^{-7}$	1.22	2.26
TFH	**0.69**	**$4.00x10^{-4}$**	0.56	0.85	0.65	$3.78x10^{-6}$	0.35	0.65
TGD	**0.68**	**$1.19x10^{-3}$**	0.54	0.86	0.64	$3.66x10^{-6}$	0.38	0.65
Th1	**0.57**	**$4.55x10^{-6}$**	0.45	0.73	0.58	$1.35x10^{-7}$	0.44	0.69
Th17	0.82	$2.45x10^{-1}$	0.58	1.15	0.44	$2.87x10^{-3}$	0.54	0.78
Th2	**0.51**	**$6.17x10^{-8}$**	0.40	0.65	0.56	$6.08x10^{-8}$	0.53	0.78
Treg	**0.59**	**$8.19x10^{-5}$**	0.45	0.77	0.54	$5.92x10^{-7}$	0.47	0.71

3.3.3 Consensus clustering of LMC tumours based on the immune cell scores

3.3.3.1 *Final number of clusters*

Consensus cluster analysis of the 703 LMC tumours was conducted using the 27 immune cell scores. Examination of consensus heatmaps, CDF and delta CDF area plot (Figure 3.4) allowed me to decide the optimal number of clusters. The change in delta CDF area (Figure 3.4E) suggested 4 or 5 clusters because after that the curve flatlined. Further, using the patterns of the consensus heatmaps, the consensus CDF plot, and the average cluster consensus plot, 3 clusters appeared as the most stable (Figure 3.4).

Having each sample assigned to one of the three clusters, for visualisation purposes, one-off hierarchical cluster analysis of cell scores (but not the samples) was conducted and a heatmap was plotted. The three clusters showed distinct immune phenotypes across the vast majority of cells: one cluster of tumours (N=156) showed evidence of coordinated infiltration of 22 of the 27 immune cells while, at other end, another cluster (N=272) showed the weakest signals for these 22 cells. Between these 2 extremes there was a cluster of tumours (N=275) displaying the 22 immune cells at intermediate level (Figure 3.5). The 3 clusters were called respectively High, Low and Intermediate Immune Subgroups (Figure 3.5).

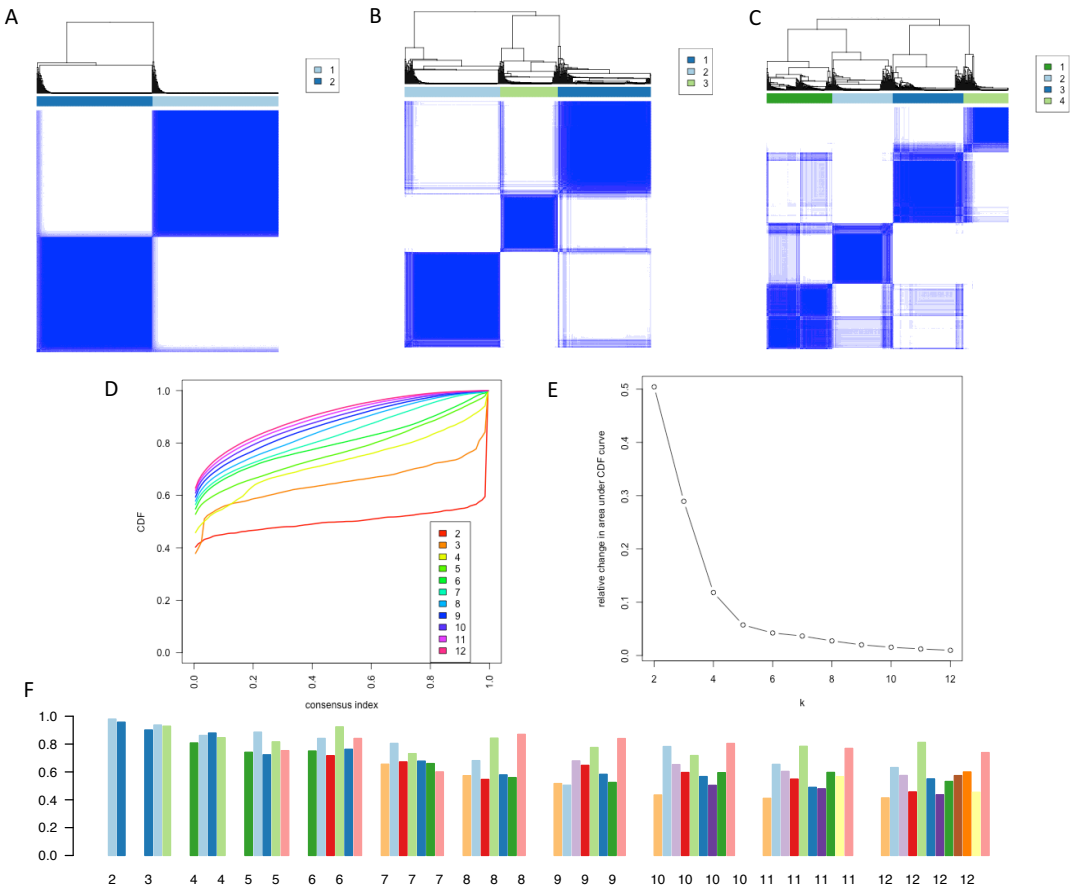


Figure 3.4: Representation of Consensus Clustering Results

(A) Sample dendrogram and heatmap with the number of clusters; The heatmaps are symmetrical and indicate the stability ("Cleanliness" of the heatmaps), the blue colour indicates high consensus (i.e. samples occurring in the same cluster with high incidence in the 5000 repetitions) whereas white indicates no consensus (samples are always classified in different clusters). (A) k=2; (B) k=3; (C) k=4. (D) Cumulative density functions (CDF) of independent runs with k=2 to 12 in 5000 data resampling. The most stable number of clusters is the one with the highest number of "0" and "1" on the x axis. The red and orange curves showed these features. (E) Relative change in area under the CDF with increasing k. (F) Average cluster consensus plot for k=2 to 12 clusters.

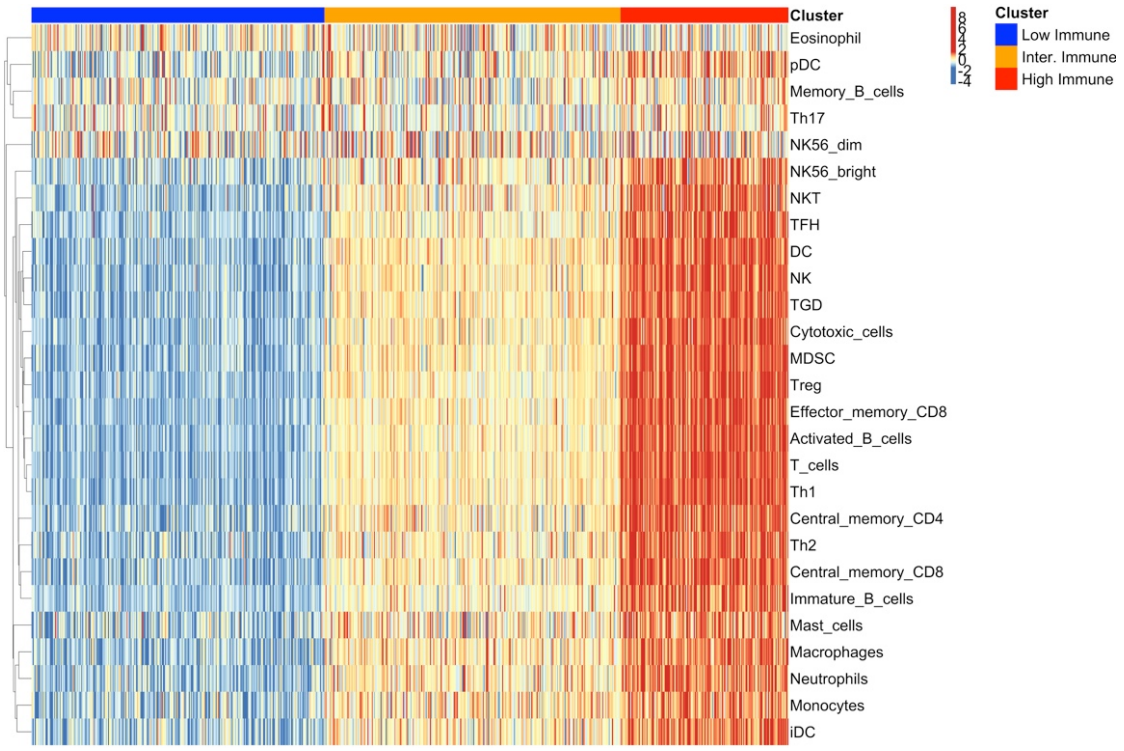


Figure 3.5: Heatmap representing immune cell scores across the three identified immune subgroups

The tumour samples (columns) were fixed according to the consensus clustering output while the immune scores were hierarchically clustered (rows).

3.3.3.2 *MSS for the obtained clusters*

The three immune subgroups were associated with MSS: a significantly lower hazard of melanoma death was observed for patients assigned to the High compared to Low (Hazard Ratio (HR)=0.5, P=0.001, 95% CI 0.3-0.7); and Intermediate Immune Subgroups (HR=0.6, P=0.05, 95% CI 0.4-1.0) (Figure 3.6).

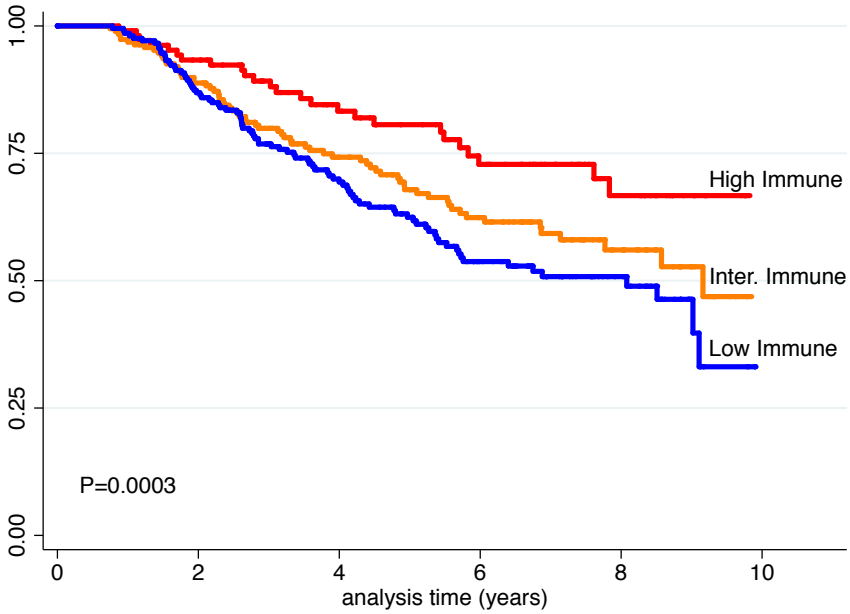


Figure 3.6: Kaplan Meier plot (MSS) for the three immune subgroups

The p-value was obtained from the overall Cox proportional hazard regression.

3.3.3.3 *Comparison of the three immune subgroups with existing molecular melanoma signatures*

As already mentioned, our group published previously 6 Consensus Immunome Clusters (CICs) in this melanoma cohort [76]. The 3 three Immune Subgroups derived here showed a very significant overlap with those 6 CICs (Figure 3.7), with a Cramer's V of 0.72, $P=6.61x10^{-151}$. The High Immune Subgroup was comprised predominantly of three Consensus Immunome Clusters: CIC 2 (High Immune) – 73.1%, CIC 3 (Intermediate Immune/Keratin Poor) – 16.7% and CIC 5 (Intermediate Immune/Keratin High) – 9.6%. The Low Immune Subgroup was predominantly comprised of: CIC 4 (Low Immune/Beta-catenin High) – 58.1%, CIC 6 (Low Immune/Keratin Rich) – 31.3% and CIC 1 (Low Immune/Beta-catenin Low) – 9.2%. These results were therefore consistent. However, the Intermediate Immune Subgroup was composed of tumours from all 6 CICs with a less clear pattern although the most represented groups were CIC 5 and CIC 3 both of which were already described as having intermediate levels of immune infiltrates [76] (Figure 3.7).

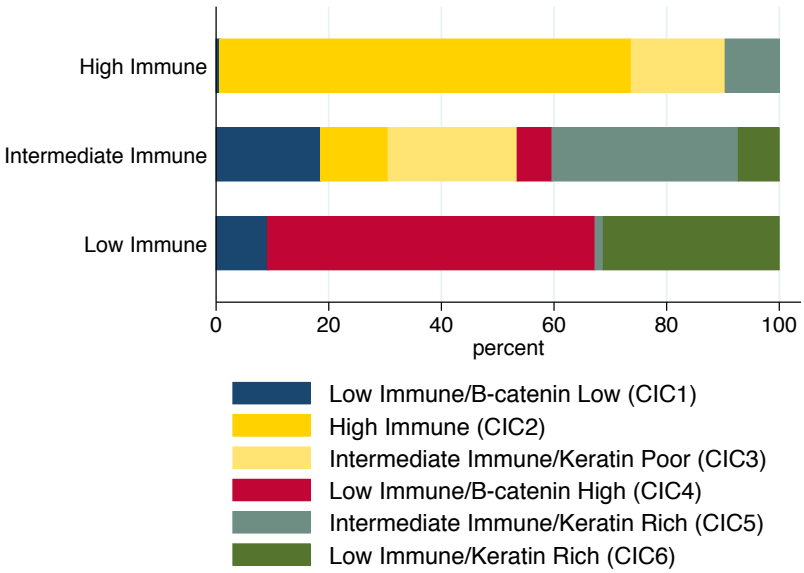


Figure 3.7: Intersection of the three Immune Subgroups and previously published Consensus Immunome Clusters by Nsengimana et al. in LMC [76]

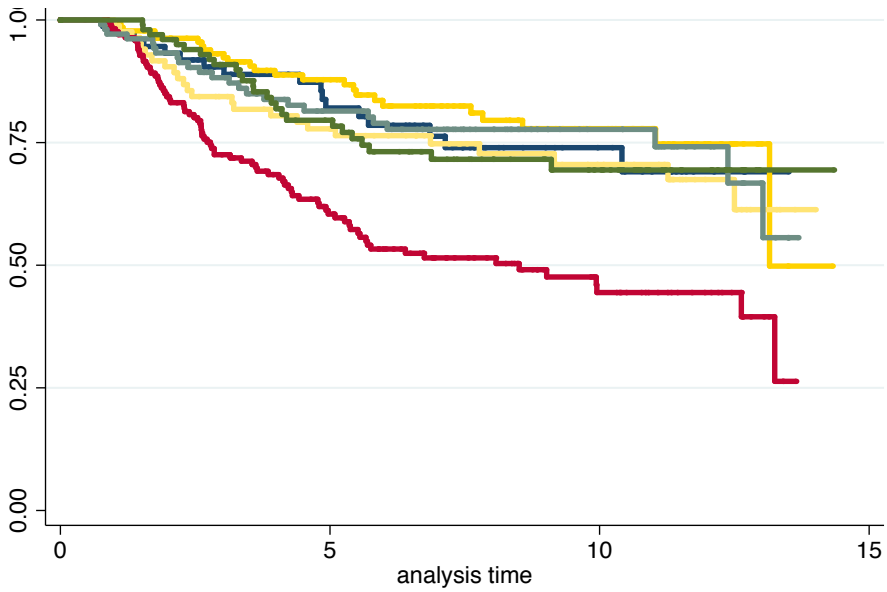


Figure 3.8: Kaplan Meier plot for the 6CIC generated by Nsengimana et al. in LMC [76]

When the 6 CICs were adjusted AJCC stage, both CIC 4 and 6 had equally poor survival. For legend, see Figure 3.7.

The 3 immune subgroups were also compared to another set of 3 melanoma molecular subtypes published by the TCGA called “immune”, “keratin-rich” and “MITF low” classes.

Here too, a good overall concordance was found with Cramer’s V=0.47, $P=2.5x10^{-54}$, albeit weaker than the concordance with the 6 CICs. Specifically, the High Immune Subgroup was comprised predominantly of TCGA “immune” class (85.7%), the Intermediate Immune Subgroup contained 33.7% of “immune”, 46% of “keratin” and 20.4% of “MITF low” tumours while Low Immune Subgroup was mainly a mixture of “keratin” (53.8%) and “MITF low” (41.1%). In terms of association with melanoma-specific survival in the LMC dataset, there is a significant difference between “MITF low” and “immune” subgroups (HR=1.97, P=0.001, 95% CI 1.3-2.9), while the “keratin” and “immune” groups have similar survival profiles (HR=0.98, P=1.97, 95% CI 0.7-1.44). (Table 3.3, Figure 3.10, Figure 3.10). Additionally, I tested whether the “keratin” group in the LMC could manifest higher expression of keratinocyte specific genes, due to sampling of keratinocytes (cells from the outermost layer of the skin) along with tumour cells from the thin tumours. I tested the expression of a gene called filaggrin (*FLG*), that is known to be expressed in keratinocytes [143] across the TCGA classes in the LMC samples. The results showed that the *FLG* expression was significantly higher in the “keratin” group ($P=3.3x10^{-65}$), (Figure 3.11). I have also tested the Breslow thickness among these group, and it showed to be the lowest in the “keratin” group ($P=4.4x10^{-11}$), (Figure 3.12). For both of the analyses Kruskal Wallis test was used.

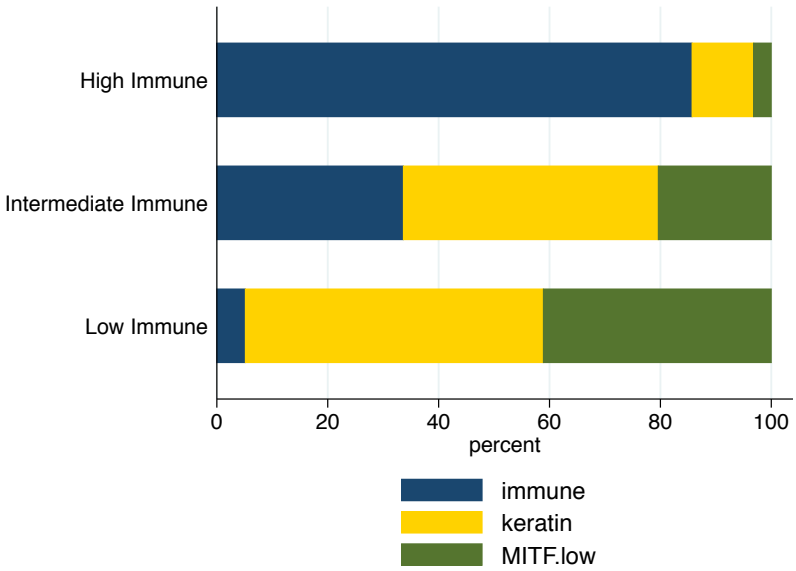


Figure 3.9: Intersection of the three Immune Subgroups and the TCGA subtypes applied to the LMC (N=560)

Table 3.3: Survival analysis of the TCGA classes applied to LMC (N=560)

	HR	***P***	***95% Conf. Interval***	
Immune	-	-	-	-
Keratin	0.98	0.9	0.66	1.44
MTF low	1.97	0.001	1.34	2.90

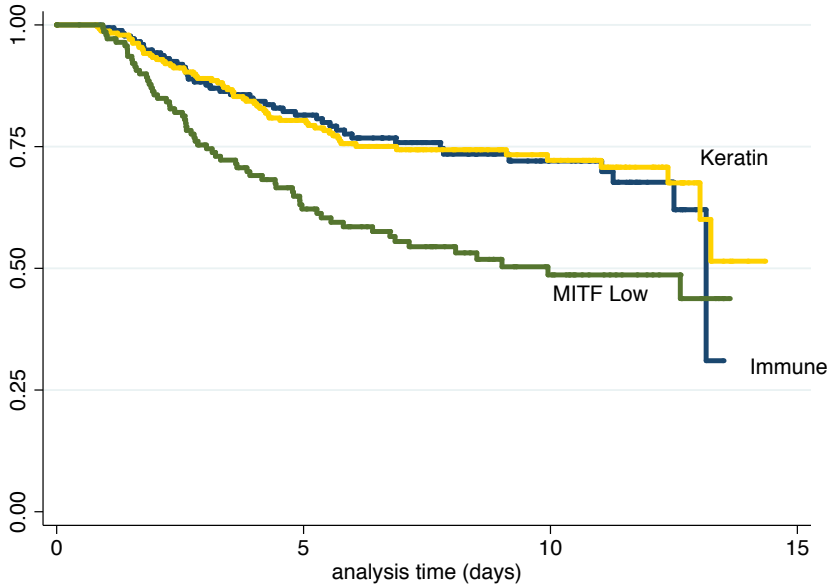


Figure 3.10: Kaplan Meier plot for the TCGA classes applied to LMC (N=560)

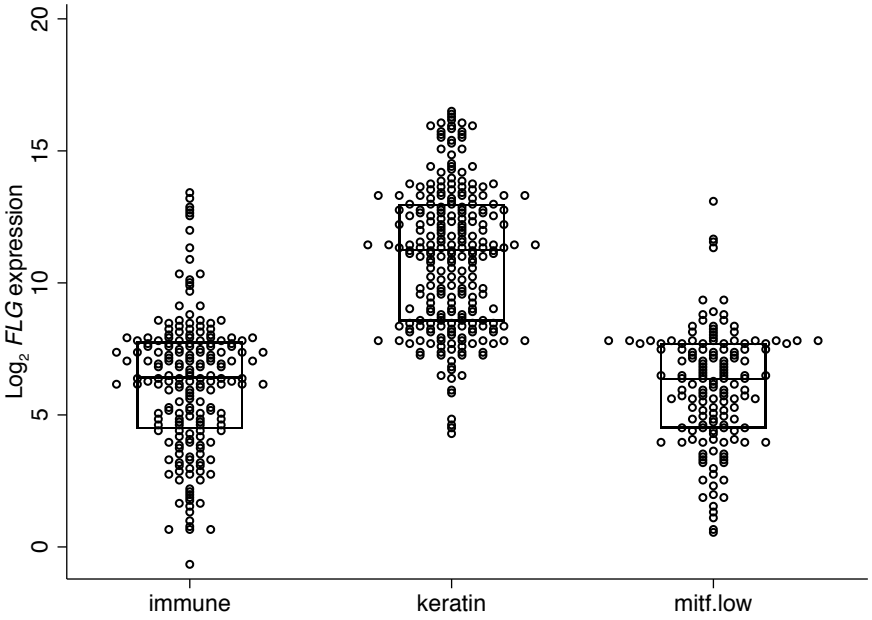


Figure 3.11: Expression of *FLG* across the TCGA classes in the LMC

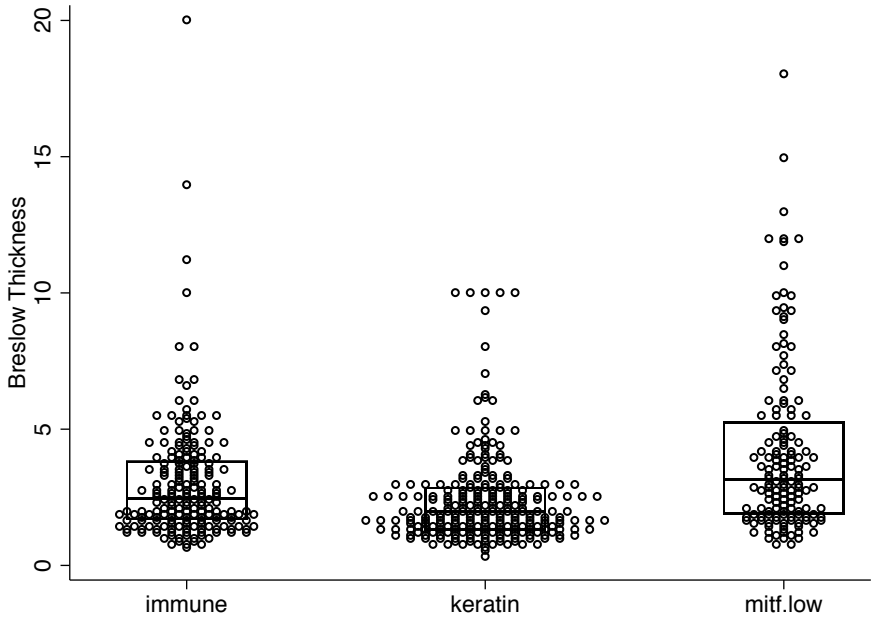


Figure 3.12: Breslow thickness across the TCGA classes in LMC

3.3.4 Validation of the 3 Immune Subgroups in TCGA melanoma metastases

After classifying TCGA metastatic tumours in the 3 Immune Subgroups using the nearest centroid method, a heatmap was drawn in a similar way as in the LMC (Figure 3.13). Similar overall patterns of immune cell scores were found (Figure 3.13) and the association with survival analysis showed a strong agreement with earlier results from the LMC: High Immune exhibited a reduced death hazard with compared to Low Immune Subgroup (HR=0.3, P=1.0x10^{-7}, 95% CI 0.2-0.5) or to the Intermediate Immune Subgroup (HR=0.5, P=4.8x10^{-5}, 95% CI 0.4-0.7) (Figure 3.14).

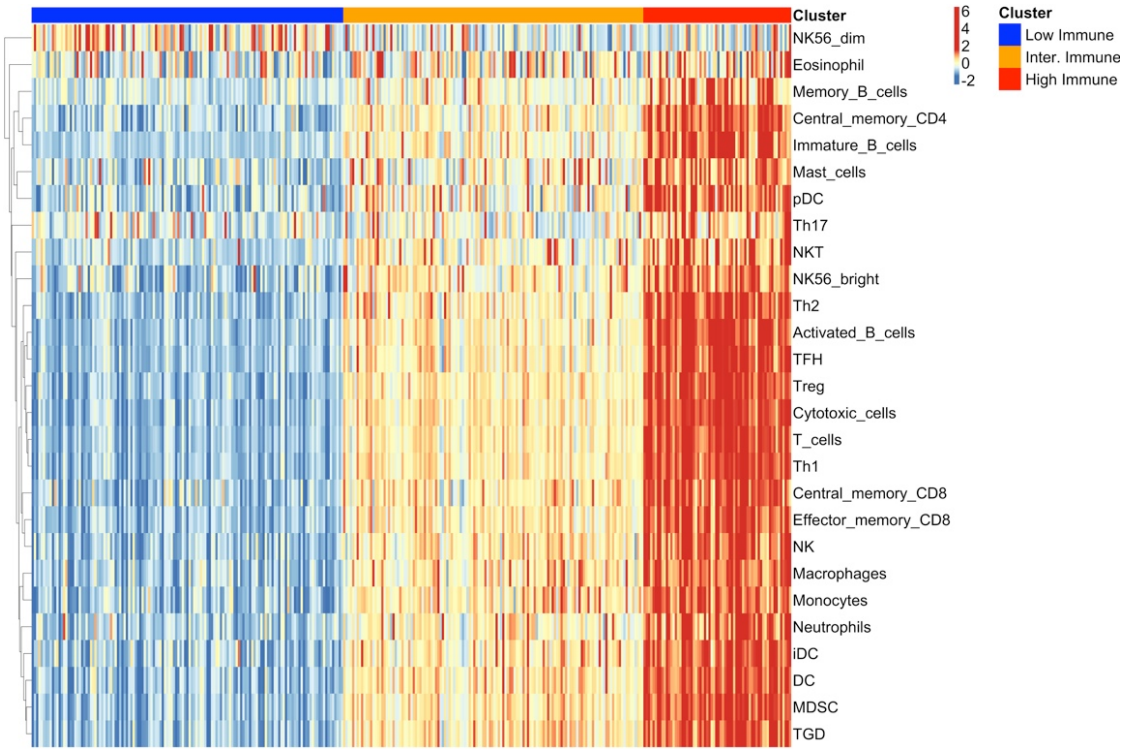


Figure 3.13: Heatmap of immune cell scores across the replicated three Immune Subgroups in TCGA

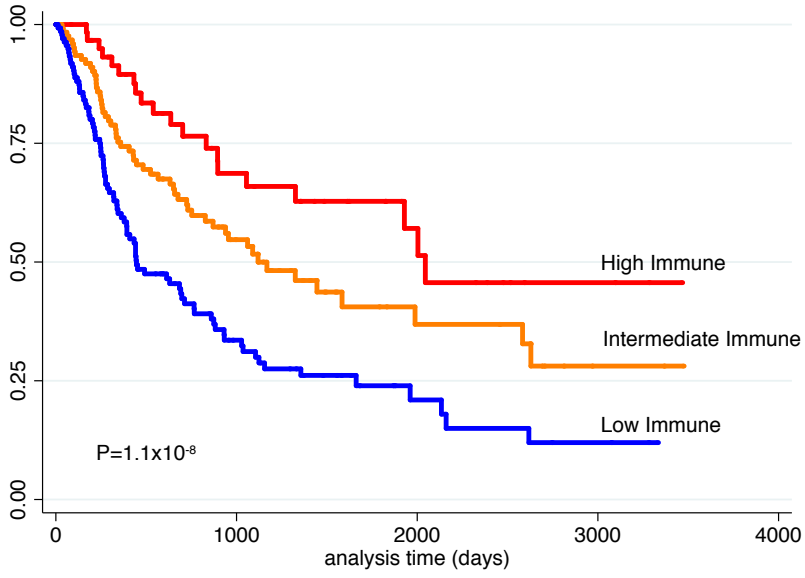


Figure 3.14: Kaplan Meier plot for OS differences between the three replicated Immune Subgroups in the TCGA dataset (p-value from Cox proportional hazard regression)

3.3.5 Clinico-histopathological analysis of the Immune Subgroups

In the LMC data, Breslow thickness, mitotic rate and the proportion of tumours without TILs (from both clinical dermatopathologists and a single observer in our research group) were significantly lower in the High Immune Subgroup compared to other groups (Table 3.4). As expected, the most significant difference was in TILs, with tumours classified in High Immune group more frequently displaying brisk TILs ($P=4x10^{-8}$). The Low Immune Subgroup had the lowest proportion of tumours with a *BRAF* mutation and the highest proportion with *NRAS* mutation, although these results were borderline significant. Site of melanoma was significantly different among Immune Subgroups, tumours arising on the rare sites (non-sun exposed) being more frequent in the Low Immune Subgroup. AJCC stage, sex, age at diagnosis, site, smoking and serum vitamin D at diagnosis (season-adjusted) were not significantly different between the three Immune Subgroups (Table 3.4).

Table 3.4: Associations of clinico-histopathological characteristics with the three immune subgroups in the LMC

Characteristic	***Low Immune***	***Intermediate Immune***	***High Immune***	***P-value (N)***
Number of participants (703)	272	275	156	
Melanoma death (%)	36.0	28.7	18.8	0.001 (666)
Age at diagnosis (median, years)	58.3	55.7	59.9	0.6 (703)
Site of melanoma				0.02 (702)
Limbs (%)	38.6	45.6	44.2	
Head (%)	11.0	10.9	12.8	
Trunk (%)	31.6	33.9	34.6	
Rare (i.e. sun protected) (%)	18.8	9.5	8.3	
Sex (% males)	43.0	44.7	50.0	0.4 (703)
BRAF-*mutated (%)*	40.5	50.4	51.2	0.06 (582)
NRA*S-mutated (%)*	29.8	24.4	16.8	0.03 (574)
Ulcerated (%)	36.0	33.5	28.9	0.32 (703)
Breslow thickness (median, mm)	2.4	2.3	2.0	0.004 (692)
Mitotic rate (median, count/mm^2)	4.0	3.0	2.5	0.0002 (596)

Characteristic	***Low Immune***	***Intermediate Immune***	***High Immune***	***P-value (N)***
AJCC stage (%)				
I	29.7	34.4	38.5	0.17 (695)
II	51.3	52.6	48.1	
III	18.96	12.96	13.46	
TILs (%) (clinic dermatopathologists)				
Brisk	8.5	13.2	27.3	$4.0x10^{-7}$ (553)
Non-brisk	62.2	65.0	55.3	
Unclassified	7.5	10.0	12.9	
No TILs	21.9	11.8	4.6	
TILs (%) – (single observer, S O'S)				
Brisk	4.0	9.8	22.7	$3.62x10^{-8}$ (601)
Non-brisk	84.1	84.6	73.8	
No TILs	12.0	5.6	3.6	
Smoking (% ever smoked)	47.2	51.6	48.7	0.6 (658)
Season-adjusted serum vitamin D at recruitment (winter median, nmol/L)	40.1	41.2	36.1	0.2 (549)
Mutational load				
Low	33.6	26.1	44.6	0.1 (319)
Intermediate	31.2	37.8	28.7	
High	35.2	36.0	26.5	

3.4 Summary

- Immune cell scores were characterised based on Angelova et al.'s Immunome.
- The tumour samples were clustered based on immune cell scores and three immune subgroups were obtained having distinct survival.
- The three immune subgroups were replicated in the independent metastatic melanoma dataset (TCGA).
- The three subgroups defined by inferring immune cell infiltration had correspondingly different histologically detected TILs, the Low Immune tumours having fewer TILs.
- The reported environmental factors of interest e.g. smoking and vitamin D levels, and the mutational load did not vary across the three immune subgroups.

3.5 Discussion

3.5.1 Immune cell scores

In general, meaningful characterisation of immune cells infiltrating tumours from potentially noisy and heterogeneous FFPE tumour gene expression data is challenging. The aim is important however, as understanding host-tumour interaction might inform lifestyle advice for newly diagnosed melanoma patients and understanding of primary and acquired resistance to checkpoint therapies. The concept is that bioinformatic analysis of transcriptomic data will allow deconvolution of complicated relationships between tumour and stromal cells *in situ* i.e. without the need to disaggregate tissues. Moreover, that this would be possible for large numbers of tumours ascertained from populations and therefore more typical of the generality of tumours rather than selected large tumours big enough to sample and disaggregate.

As mentioned in the introduction, there are several methods for inferring immune cell score, available. In my thesis I have explored CIBERSORT and developed a method based on Angelova et al.'s Immunome. I did not test all of the approaches either because they were reported to cover a relatively small spectrum of cell subtypes [109] or because they required RNA sequencing data (as indicated in the R package code the provided by the authors) [110], or they were published relatively late during my PhD (November 2017) [112].

Generally, these methods do not allow distinction to be made between strong signals related to a large number of cells or fewer cells but with higher gene expression per cell. Moreover, the gene-cell specificity may be context-dependent, i.e. genes could be expressed by more than one cell type (immune or non-immune) depending on the tissue or local microenvironment.

Nonetheless, I have explored some of them in my thesis. I first evaluated the utility of an existing cell scoring method (CIBERSORT), which uses as reference gene signatures of flow cytometry sorted circulating immune cells (LM22). This method has some valuable features that other methods do not have. One of them is that it is available as a software easily accessible, it is user friendly, and no bioinformatic knowledge is needed in order to run the algorithm. Another one is that it provides the empirical p-value indicating the level of confidence the user should put in the gene expression

deconvolution for each sample. Moreover, the LM22 signature is comprehensive and contains some immune cell subsets with subtle differences which might have a relevance to melanoma. For example M0, M1, and M2 macrophages are held to be distinct in this signature and macrophage polarity has been shown to play a role in cancer prognosis [144].

However, when applied to the LMC dataset, CIBERSORT showed to have some limitations. In fact, the algorithm inferred successfully immune cell proportions in a minority of tumours.

Varying the p-value threshold did not make a significant difference. Similar results were also found after applying different data transformation strategies. However, and interestingly, the tumour samples where CIBERSORT failed to estimate immune cell proportions were more likely to be recorded as having no TILs and were classified and Consensus Immunome Clusters 1, 6 and 4 which all have been associated with lack of immune infiltration [76]. It can be concluded therefore that failure of CIBERSORT to find immune cell proportions with a significant p-value is not an artefact but a true reflection of very low levels of immune infiltration. I nevertheless consider this as a limitation of the method because it is unclear how to use the cell proportions estimated with low level of confidence (P>threshold). In its set up, the total of all 22 immune cell proportions is 100%, meaning that if there are no immune cell at all in the tumour, proportions of all 22 are still estimated but there is no indication on what they actually represent. The use of a relative rather than absolute quantitation may be more appropriate to compare the extent of different immune cells within a tumour rather than comparing tumours.

Since CIBERSORT was unconvincing, I adopted and modified another method of immune cell scoring. The Immunome proposed by Angelova et al. [1] was explored in-depth in this chapter and the scoring approach I developed using this list was the basis for further analyses in my thesis (next chapters). Advantages of the Angelova et al.'s Immunome are that it covers a large number of immune cells (31), and can be adapted to a specific cancer under study by performing additional (quality control) QC. It allows user-defined immune cell scoring on an absolute rather than a relative scale. However, this approach has an additional caveat apart from the ones mentioned above. Using melanoma and melanocyte cell lines, I found that almost half of the genes proposed by

Angelova et al. were expressed at significant levels by melanoma cells or melanocytes. It is this observation that prompted the decision to conduct some QC steps (gene filtration) before developing the immune scoring method. After this filtering, further explorations showed that some of the genes deemed to be specific to certain purified immune cell subtypes according to Angelova et al. [1] actually displayed a negative correlation with each other in the tumours, implying either the presence of technical issues or that these genes might be less useful markers of those cell subsets in melanoma than they would be in the peripheral the blood. In favour of the potential bias from technical variation we observed that gene expression correlations within immune cell type were less likely to be negative in TCGA melanoma samples, which were produced from fresh frozen tissue, therefore with less degraded mRNA compared to the LMC. One example was the Treg score. When analysing the gene correlations for Treg score, the well-known marker for this cell type – *FOXP3* was negatively correlated with majority of genes assigned to this cell type in LMC, but not in TCGA. Consequently, I concluded that the there is a possibility that in the LMC transcriptome the probe designed for this gene was not efficient enough, as it also could be the case for other probes.

It is important to indicate, that when it comes to Bindea et al.'s Immunome the Treg score (only the *FOXP3* gene, for which the probe in LMC possibly could not be efficient) could not be confidently used.

When I compared the Th2 score using the list of genes proposed by Bindea et al. and Angelova et al.'s Immunomes I showed that *GATA3* (well recognised marker for Th2 cells) was negatively correlated with the rest of the proposed genes, but it was not observed using Angelova et al.'s Immunome both in LMC and TCGA.

The examination of correlation of genes assigned to Treg and Th2 scores in TCGA and LMC reassured me that the Angelova et al.'s Immunome is more reliable than Bindea et al.'s Immunome at least for characterising these two particular cell types. Because of this reason and the fact that the Angelova et al.'s Immunome was devised from more microarray data than Bindea et al.'s. I used this list, after using the devised filters for scoring immune cells in the LMC. Subsequently I proceeded with using these immune cell scores for samples clustering.

3.5.2 Clustering

Consensus clustering is widely used method to find stable sample subgroups in high-dimensional data. I have chosen this approach over one-off clustering because it generates helpful metrics to indicate the optimal number of clusters which is unavailable in standard clustering [134]. This approach allowed me to identify 3 stable immune subgroups, which were well-replicated in TCGA metastases, both in terms of overall cell score patterns and in their associations with survival. The concordant results were reassuring that the cell scoring method was reproducible in independent datasets.

Replication of the 3 immune subgroups in TCGA metastatic samples suggested that the immune infiltration and exclusion mechanism spans the whole range of disease progression, i.e. both primary and metastatic tumours may have different levels of immune response.

Additionally, the identified immune subgroups were consistent with the earlier reported 6 CICs [76]. Besides using different Immunomes (Bindea et al. vs. Angelova et al.), the other differences between the earlier report [76] and the current analysis are a more in-depth QC filtering and tumour clustering using cell scores rather than the expression of individuals genes. I hypothesized that dimensionality reduction by using the immune cell scores in tumour clustering could be more informative and less affected by noise. The 3 identified immune subgroups had distinctive survival profiles. In comparison with 6 CIC and TCGA (“immune”, “keratin” and “MITF low”) signatures, which did not show as distinct survival patterns for each group, the newly identified immune subgroups showed to have advantage over these signatures. Interestingly in the TCGA (not in the LMC) the “keratin” signature which predicted the poorest prognosis was comprised of approximately 74% primary melanomas [46]. In the LMC I showed that “keratin” group was comprised of samples that were thinner and contaminated by keratinocytes (high *FLG* expression) comparing to other groups. Patients with thin tumours are known to have favourable survival, and this could explain why “keratin” group (the thinnest tumours) had comparable survival to the “immune” group in the LMC. Taken together, this observation supports our concern that in TCGA primary melanomas (mostly categorised as “keratin”) have the poorest survival, and should be analysed with caution.

The immune cell scores were highly correlated and had similar patterns of distribution across the immune subgroups. I have moreover observed that immune cell scores, for which the reciprocal gene correlations per score were weak (e.g. Eosinophil, pDC, Memory B cells, NK dim), were less distinct among clusters (or contributed less to clustering) than those having high correlations, and all of them did not significantly predict prognosis. This observations, questioned the reliability of these particular immune cell scores for further analyses.

The high correlation of the majority of immune scores with each other suggests that multiple immune surveillance mechanisms (together with immune checkpoint pathways) are co-ordinately activated to combat cancerous melanocytic lesion, as it has been reported previously [77][76]. However, there is also a concern that applying computational methods to gene expressions for cell sorting may not have enough power to detect subtle differences in the abundance or the functions of these cells. There might be epigenetic or post-translational changes that regulate the immune cell function that are not detectable in transcriptomes. Nevertheless, the High Immune Subgroup had the highest expression of almost all the immune cells and the best melanoma specific survival among the three immune subgroups consistently with current literature.

3.5.3 Clinico-histopathological associations with obtained clusters

Upon testing the clinico-histopathological associations among the three immune subgroups, AJCC stage, age at diagnosis, sex, ulceration, vitamin D or reported smoking did not vary across the immune clusters, meaning that these variables were independent of the strength of the immune response in this analysis. The three immune subgroups showed good overlap with TILs measurements, which gave support to the view that using inferred immune cell scores do represent the presence of more immune cells. Moreover, overall I observed clear association of the high immune responses negatively correlated with tumour thickness and mitotic rates. Tumour site was associated with immune subgroup status and this was driven by the finding that the majority of samples from “rare sites” (those arising on ‘sun-protected’ sites) were classified as low immune. It is well-documented that exposure to ultraviolet light induced C>T mutations with may play neo-antigenic role and it has been reported that

melanomas in sun-protected body sites have fewer genomic mutations, and therefore fewer neoantigens, potentially reducing the attraction of TILs [19]. This is a credible explanation for the observed worse survival and higher stage at diagnosis of patients with these rare melanomas, as well as the probability of late diagnosis [16]. However, the mutational load, which we assumed that could reflect the neoantigen load, was not shown to be significantly different across the three immune subgroups. This observation was surprising, because it has already been shown that total mutational/neoantigen load correlated with better responses to immunotherapies for metastatic melanoma patients [145][146][147] and higher immune cell scores for the TCGA melanoma data [148]. However, Spranger et al. in the study analysing TCGA tumour data showed no association of mutational density with level of T cell infiltration in melanoma and other cancers [149]. Moreover, it is important to stress that the mutational load data used in these analyses was selected, meaning that the mutation count was based only for a panel of 555 genes (see Chapter 2), which might not necessarily represent the whole genome/total mutational load. Moreover, the mutation data were generated from thicker tumours and only for 319 samples, potentially introducing bias and statistical power issues. Lastly, some studies suggested that it is a neoantigen quality not quantity that predicts response to immune therapies in melanoma and other cancers [150][151] and it was shown that only ~10 of somatic mutations, which results in amino acid changes were predicted to bind to MHC class I [152].

NRAS-mutated tumours were more often classified into the Low Immune Subgroup and *BRAF*-mutated into Intermediate or High Immune Subgroup, although at a borderline significance level. This evidence could suggest that rather *BRAF* mutated tumour might be more immunogenic than *NRAS* mutated tumours and the literature supports this observation. *BRAF* mutated melanomas have been reported to be immunogenic [153][154][155] and it was reported by Nsengimana et al. (the researcher in our group), that survival benefit from immune cell infiltration was observed in *BRAF* or wild-type tumours while it was absent in *NRAS* mutated tumours in LMC [76]. Moreover, another study analysing 912 higher-risk primary melanomas, demonstrated that *NRAS* mutated tumours had significantly lower TILs than *BRAF* and WT tumours [156]. This results corroborate with the reported evidences that oncogenic Ras reduces antigen

presentation complex (MHC class I) on the surface of lymphoma tumour cells [157]. On a contrary another study showed that patients with *NRAS* mutated melanoma tended to have more favourable survival after treatment from anti PD-L1/PD1 therapy [158], but this study was conducted on relatively small samples size (229) and only metastatic melanomas.

In summary, the above results reassured me that the tumour subgroups obtained from consensus clustering correspond to biologically distinct entities with potentially specific immune surveillance, edition and evasion mechanisms. In the next chapters these three immune subgroups were explored as the basis of identifying modulators of the immune responses to melanoma.

Chapter 4

Identification of genes and pathways associated with immune response to melanoma

4.1 Aims of the chapter

- To identify pathways that are associated with immune response to melanoma using the genome wide differential expression across the three immune subgroups uncovered in Chapter 3.
- To identify the most nodal genes in those pathways using network analysis.
- To validate the role of nodal genes as mediating immunosuppression in melanoma cell lines.

4.2 Methods

4.2.1 Overrepresentation analysis (ORA)

Overrepresentation analysis is a widely utilised method of studying gene expression data in order to identify groups of genes expressed more highly in the analysed groups of subjects (i.e. with different phenotypes) [159]. This process is generally divided into two stages: first, testing differentially expressed genes between the two phenotypes using standard statistical methods; and second, biological interpretation of the identified genes by for example pathway enrichment analyses. The ORA can be only implemented if the sample size is large enough to find significantly differentially expressed genes between the analysed groups [159]. Since the LMC is a large data set, I reasoned that this method would be appropriate.

4.2.1.1 *Testing differentially expressed genes among the three immune subgroups*

In order to identify biological pathways associated with the immune responses to melanoma I carried out an agnostic analysis, which tested the whole transcriptome differences between the analysed tumour immune subgroups.

I tested differential expression of 29,424 probes among the three immune subgroups in LMC using the Kruskal Wallis test in STATA. The significance threshold was set at Bonferroni-adjusted α=0.05 for multiple testing (i.e. $2x10^{-6}$). Subsequently I matched the probes to the gene names and I removed the 376 genes which had been used to create the immune cell scores.

To visualise the expression of significantly differentially expressed genes among the three immune subgroups I generated a heatmap using the "pheatmap" package in R, with hierarchical clustering of the genes (using the Pearson correlation coefficient as distance metric). For differential expression analysis, I only tested the differences between the two extreme immune subgroups – Low vs High using the Mann-Whitney U test.

4.2.1.2 *Biological interpretation of the differentially expressed genes*

I analysed separately the genes that were significantly upregulated in the Low and in the High Immune Subgroup (p-value threshold $2x10^{-6}$) in Reactome FiViz 5.2.0.beta [160] and Centiscape v.2.1 [161] plugins of Cytoscape v.3.5 [162]. Reactome FIViz enables analysis of pathways and protein-protein interaction (PPI) networks using just a list of gene names as an input or gene expression data (mainly used for correlation calculation among the genes). In my analyses I used the gene names that were upregulated either in the Low or in the High Immune Subgroup.

Reactome FiViz has an inbuilt protein functional interaction (FI) network, which was predicted by machine learning approach (naïve Bayes classifier) using protein-protein interaction (PPI) databases such as:

- human physical PPIs catalogued in IntAct [163], HPRD[164], and BioGrid [165]
- human PPIs projected from fly, worm and yeast in IntAct [163] based on Ensembl Compara [166]
- human gene co-expression derived from DNA microarray studies (two data sets [167][168])
- shared GO (Gene Ontology) biological process annotations [169]
- protein domain-domain interactions from Pfam [170]

- PPIs extracted from the biomedical literature by the text-mining engine GeneWays [171].

The naïve Bayes classifier was trained based on the pairwise protein relationships by utilising curated pathways from Reactome [172] (training dataset). Subsequently, in order to predict functional interactions of the pairs of proteins, the trained naïve Bayes classifier was applied. Next, the predicted FIs were merged with the ones from curated pathways databases:

- R – Reactome
- C – CellMap
- K – KEGG
- N – NCI PID
- P – Panther
- B – BioCarta,

where the curation in this instance means that the pathway databases were created by researchers who carefully verified the evidence from each source of interaction.

The pathway enrichment analysis was based on the hypergeometric test, which tests if the submitted list of genes or proteins contains more genes/proteins for a given pathway than would be expected by chance. The significance of that test is indicated by a p-value and its version adjusted for multiple testing by Benjamini-Hochberg method [173], i.e. the False Discovery Rate (FDR) for each pathway from the curated databases mentioned above.

The software generates a table with enriched pathways within a given gene set, where: “GeneSet” is the pathway name, “Ratio of protein in Geneset” indicates ratios of numbers of genes contained in pathways to the total number of genes in the Reactome FI network; “Number of protein in GeneSet” is the number of proteins within a given pathway; and “Protein from network” indicates number of genes from the input gene list per pathway.

To run this analysis, I firstly uploaded the gene names upregulated in the High and Low Immune Subgroups into RectomeFIViz and the PPI network was constructed. Next, I tested the pathway enrichment separately for each list, setting the FDR threshold for calling pathway significant significantly enriched at a stringent level below 0.001.

After having the PPI networks constructed based on the genes of interest from each of the 2 immune subgroups I looked for the most important/influential genes within the networks. In order to identify these genes in the networks, the “betweenness” metric (indicating a key role in communication between proteins) was used as a centrality measure in Centiscape [161]. In technical terms betweenness is the number of shortest paths going through a node (gene) when connecting each pair of nodes in the network. In biological terms the nodes with the highest betweenness centrality are likely to be crucial in holding together/controlling the proteins within the network [161]. I performed the visual adjustments (size, colour of the nodes as well as edges shapes) of the networks in another software package called Gephi v.0.9.1 [174].

4.2.2 Patient derived melanoma cell lines analysis

After identifying key immunosuppressive gene in the Low Immune Subgroup network, it was necessary to verify its functional role in an independent dataset. In our data, a gene could be expressed by tumour cells, immune cells, both or even by other components of tumour microenvironment. In order to explore the gene expression associations with the nodal gene we therefore collaborated with Mitch Levesque’s group in Zürich to examine the associations with the nodal gene in patient derived tumour cell lines, which lack immune cells. The methodology described (Chapter 2) for the generation of the cell lines is reported to significantly reduce fibroblast cell numbers.

The statistical analysis of the RNAseq gene expression data originating from the 103 melanoma cell lines (described in Chapter 2) was carried out by Sabrina Hogan (Marie Skłodowska-Curie PhD student at the University of Zurich, Switzerland). The Spearman rank correlation test was used to test correlations of the nodal gene (from the network created based on genes upregulated in the Low Immune Subgroup) with the whole transcriptome in the primary melanoma cell lines. Subsequently, I analysed the topmost 50 positively and negatively correlating genes with the nodal gene in Reactome FIViz to test the biological pathway enrichment of these genes.

4.3 Results

Overrepresentation analysis (ORA)

There were 5324 genes differentially expressed across the genome among the three immune subgroups, they are shown Figure 4.1. The heatmap shows a clear clustering of High and Low Immune Subgroups while, the Intermediate Immune Subgroup was essentially a mixture of these extreme subgroups (Figure 4.1). Excluding this middle group, the number of genes differential expressed between the two extreme groups is 5607, with 3324 upregulated in the High and 2283 upregulated in the Low Immune Subgroup.

4.3.1.1 Low Immune Subgroup

The pathways significantly enriched in the genes upregulated in the Low Immune Subgroup predominantly represented proliferation and metabolic functions. The most enriched pathways were the tricarboxylic acid (TCA) cycle and respiratory electron transport (FDR=$5.3x10^{-14}$), mitochondrial translation (FDR=$5.3x10^{-14}$), ribosomal and rRNA processing, and mitosis pathways (Table 4.1, and Appendix A.2.1). The analysis of the PPI network of the genes enriched in the Low Immune Subgroup revealed that proto-oncogene *MYC* had the highest betweenness centrality, suggesting that it was the most pivotal protein in this network (betweenness=31141) (Figure 4.2). The next two genes with relatively high betweenness were: Polo like kinase 1 (PLK1) – betweenness=18634, and Protein Phosphatase 2 Scaffold Subunit Alpha (PPP2R1A) with betweenness of 11667. For further analyses I have chosen *MYC* as an immunosuppressive candidate due to its highest betweenness measure.

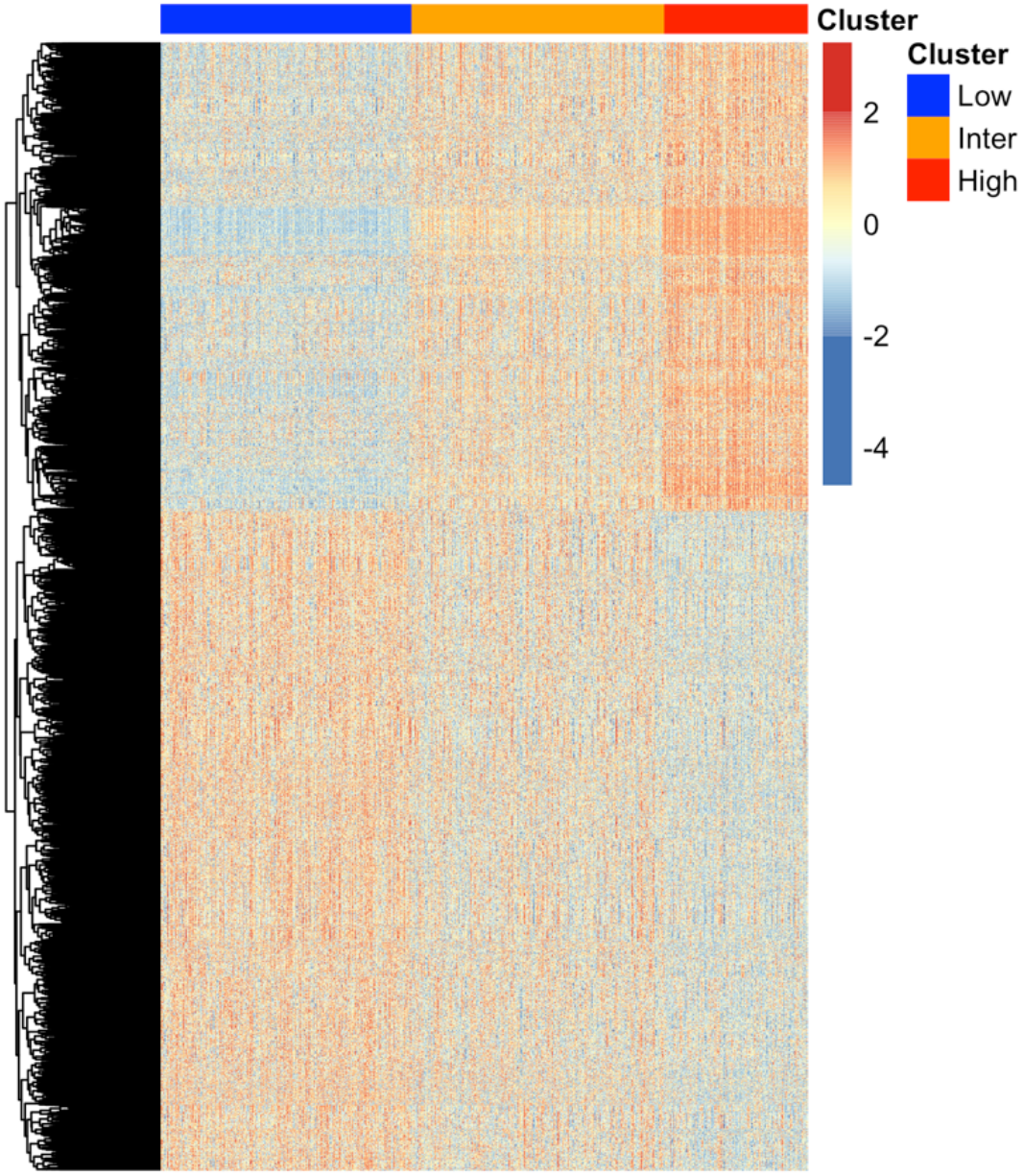


Figure 4.1: Heatmap representing significantly diferentially expressed genes (rows) across the three immune subgroups (columns)

Table 4.1: The top ten enriched pathways in the Low Immune Subgroup

"GeneSet" is the pathway name, "Ratio of protein in Geneset" indicates ratios of numbers of genes contained in pathways to total genes in the Reactome FI network; "Number of protein in GeneSet" is the number of proteins within a given pathway; and "Protein from network" indicates number of genes from the input gene list per pathway. "Nodes" are proteins from the network.

GeneSet	***Ratio Of Protein In GeneSet***	***Number Of Proteins In GeneSet***	***Protein From Network***	***P-value***	***FDR***
The citric acid (TCA) cycle and respiratory electron transport (R)	0.0148	161	66	1.1×10^{-16}	5.3×10^{-14}
Mitochondrial translation (R)	0.0085	92	58	1.1×10^{-16}	5.3×10^{-14}
Ribosome (K)	0.0142	154	59	1.3×10^{-14}	4.2×10^{-12}
Generic Transcription Pathway (R)	0.0455	494	121	9.0×10^{-14}	2.1×10^{-11}
rRNA processing (R)	0.0165	179	62	2.1×10^{-13}	4.1×10^{-11}
Parkinson's disease (K)	0.0131	142	53	7.8×10^{-13}	1.2×10^{-10}
Mitotic Prometaphase (R)	0.0091	99	43	1.0×10^{-12}	1.4×10^{-10}
Mitotic Metaphase and Anaphase (R)	0.0151	164	57	1.8×10^{-12}	2.1×10^{-10}
Oxidative phosphorylation (K)	0.0122	133	50	2.7×10^{-12}	2.8×10^{-10}

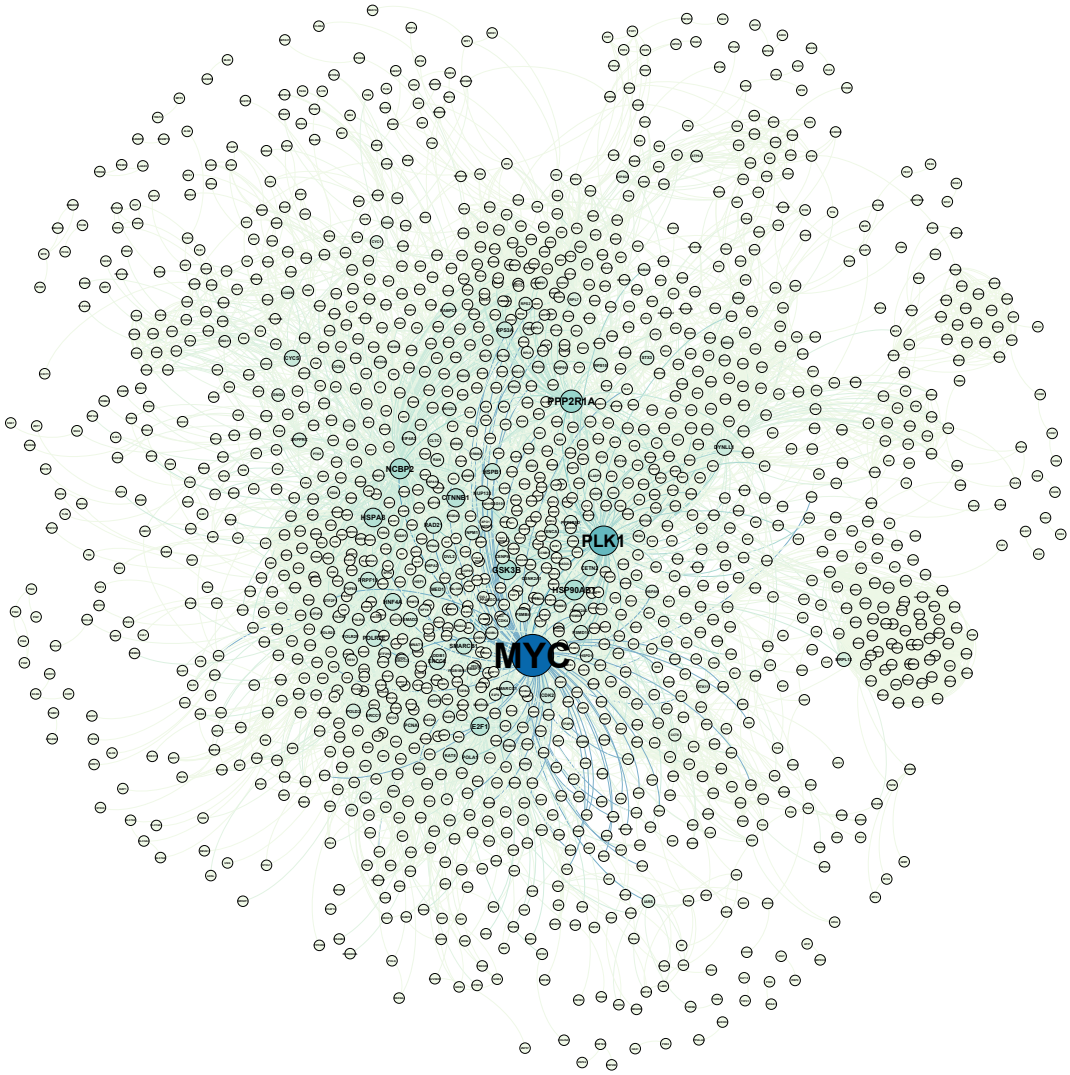


Figure 4.2: PPI network of the genes upregulated in the Low Immune Subgroup

The size of the node indicates its importance (centrality).

Previously our group reported that β-catenin signalling is upregulated in 42% of LMC tumours overall and in 73% of one particular tumour subset with the poor immune responses confirming its immunosuppressive function [76]. In Chapter 3, I showed that the three immune subgroups and 6 CICs (associated with β-catenin signalling pathway) were in a strong agreement. The agnostic analysis of the genes upregulated in the Low Immune Subgroup showed that *MYC* might be the crucial gene involved in the immune evasion, but the gene coding for beta catenin (*CTNNB1)* was also involved in this network. This observation could indicate co-operation of *MYC* and *CTNNB1,* which was not surprising, because it has been suggested that *MYC* is a target of β-catenin [75].

Therefore, I asked what was the relation of *MYC* and *CTNNB1* in terms of association with the Low Immune Subgroup. I dichotomised both *MYC* and *CTNNB1* based on median expression and generated combinations of these new binary variables:

- *CTNNB1*/*MYC* low (N=230)
- *MYC* dominant, i.e. CTNNB1 low/ MYC high (N=122)
- *CTNNB1* dominant, i.e. MYC low/ CTNNB1 high (N=135)
- *CTNNB1*/*MYC* high (N=216).

Figure 4.3 shows that tumours with both *MYC* and *CTNNB1* high were most frequent in the Low Immune Subgroup (47%). This subgroup also contained 18% of tumours with *MYC* dominant and 17% of tumours with *CTNNB1* dominant. The *CTNNB1* dominant subgroup was equally distributed across the three immune subgroups (17%, 23%, 17%), while *MYC* dominant was more frequent in the Low than High Immune Subgroup (18% vs 9%) (Table 4.2).

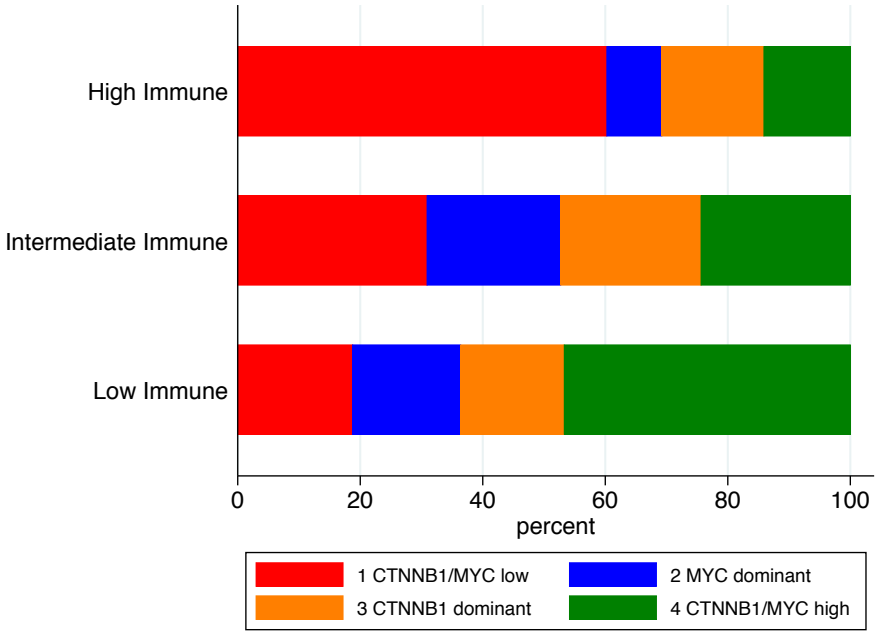


Figure 4.3: Percentages of 4 groups based on *MYC* and *CTNNB1* expression across the three immune subgroups

Table 4.2: Distribution of frequency of the *MYC*/*CTNNB1* groups across the three immune subgroups, Chi2 P=1.7x10^{-20}

	CNNTB1/ MYC low	***MYC dominant***	***CNNTB1 dominant***	***CNNTB1/ MYC high***	***Total***
High Immune (N)	94	14	26	22	156
(%)	60.26	8.97	16.67	14.1	100
Intermediate Immune (N)	85	60	63	67	275
(%)	30.91	21.82	22.91	24.36	100
Low Immune (N)	51	48	46	127	272
(%)	18.75	17.65	16.91	46.69	100
Total (N)	230	122	135	216	703
(%)	32.72	17.35	19.2	30.73	100

4.3.1.2 High Immune Subgroup

The analysis of the genes upregulated in the High Immune Subgroup (n=2283) indicated that they were mostly enriched in genes active in immune pathways, as expected. The top enriched pathways were: Interferon alpha/beta and gamma signalling (FDR=2x10^{-14}), antigen processing and presentation (FDR=2x10^{-14}), chemokine signalling (FDR=2.7x10^{-11}), and NF-κB signalling (FDR=9.8x10^{-11}) (Figure 4.4, Table 4.3 and Appendix A.2.2). All these pathways are known to be activated in various immune cells and in general in immune microenvironment. The up-regulation of these pathways confirmed the up-regulation of the immune cell scores in this immune subgroup.

The nodal gene in this network was a transcription factor *NFKB1* (Nuclear factor NF-kappa-B) encoding the p105/p50 subunit of NF-κB (Figure 4.4). The betweenness measure for NFKB1 was 19838. The second node with highest centrality measure was tyrosine-protein kinase - FYN (betweenness =11863) and Signal transducer and activator of transcription 1-alpha/beta - STAT1 (betweenness =6834).

Table 4.3: The top 12 enriched pathways in the High Immune Subgroup.

"GeneSet" is the pathway name, "Ratio of protein in Geneset" indicates ratios of numbers of genes contained in pathways to total genes in the Reactome FI network; "Number of protein in GeneSet" is the number of proteins within a given pathway; and "Protein from network" indicates number of genes from the input gene list per pathway. "Nodes" are proteins from the network.

GeneSet	***Ratio Of Protein In GeneSet***	***Number Of Protein In GeneSet***	***Protein From Network***	***P-value***	***FDR***
Interferon alpha/beta signaling(R)	0.0063	68	42	1.1×10^{-16}	2.0×10^{-14}
Antigen processing and presentation(K)	0.0071	77	39	1.1×10^{-16}	2.0×10^{-14}
Influenza A(K)	0.0161	175	56	1.1×10^{-16}	2.0×10^{-14}
Interferon gamma signaling(R)	0.0067	73	47	1.1×10^{-16}	2.0×10^{-14}
Signaling by Interleukins(R)	0.0423	460	99	1.1×10^{-16}	2.0×10^{-14}
Herpes simplex infection(K)	0.017	185	57	2.2×10^{-16}	3.3×10^{-14}
Cytokine-cytokine receptor interaction(K)	0.0244	265	67	7.9×10^{-15}	1.0×10^{-12}
Th17 cell differentiation(K)	0.0098	107	39	9.6×10^{-14}	1.1×10^{-11}
Chemokine signaling pathway(K)	0.0172	187	52	2.7×10^{-13}	2.7×10^{-11}
NOD-like receptor signalling pathway(K)	0.0156	170	49	3.7×10^{-13}	3.3×10^{-11}
TNF signalling pathway(K)	0.0101	110	38	9.7×10^{-13}	7.8×10^{-11}
NF-kappa B signalling pathway(K)	0.0087	95	35	1.3×10^{-12}	9.8×10^{-11}

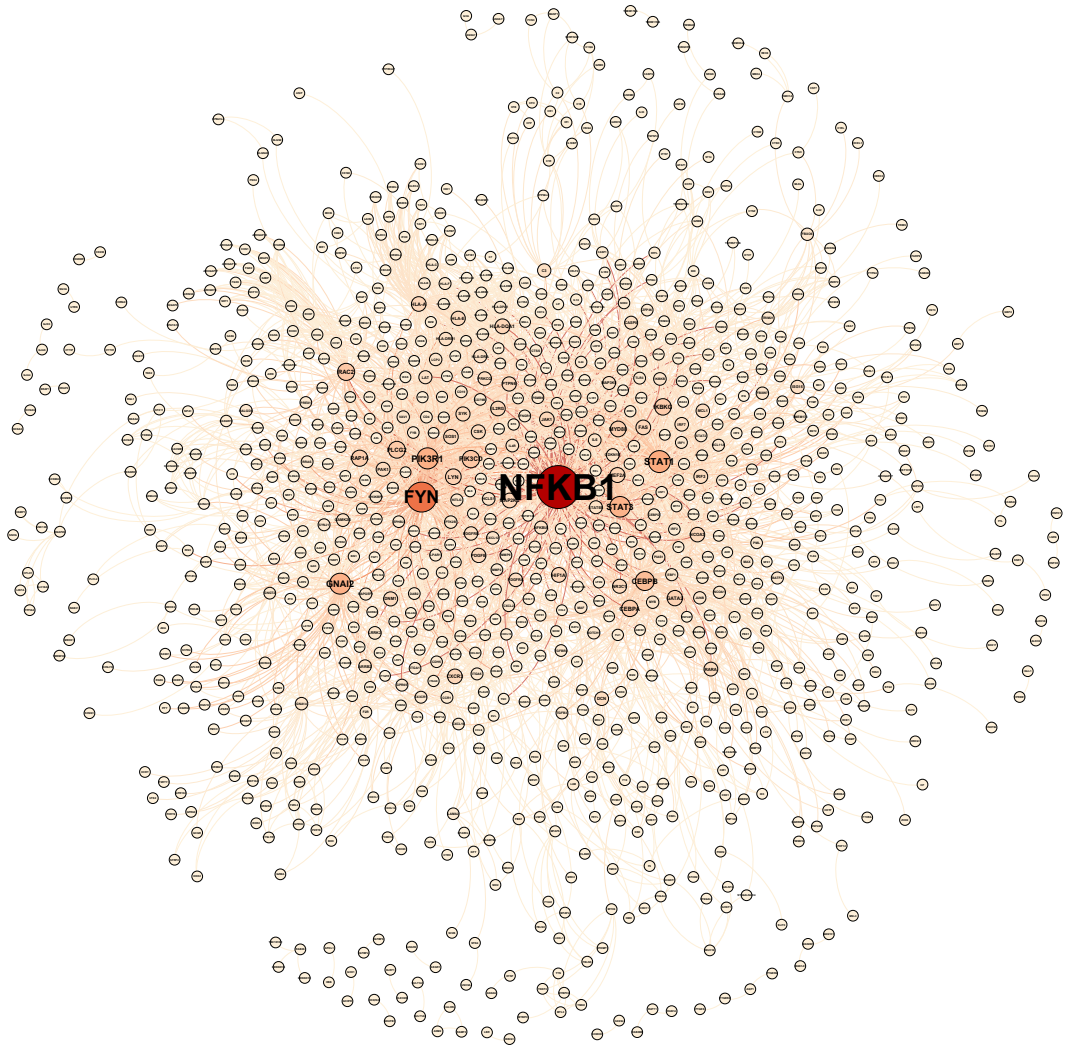


Figure 4.4: PPI network of the genes upregulated in the High Immune Subgroup

The size of the node indicates the gene importance (centrality).

4.3.2 Patient derived melanoma cell lines analysis

The identification of *MYC* as the gene with the highest centrality (betweenness) among the genes of the Low Immune Subgroup network and at the same time having a negative correlation with the High Immune genes suggested that it might play a key role in immune evasion/escape. However, *MYC* is widely expressed in different tissues and different cell types [143] and we have no indication of which cells is expressing it in our dataset. It was therefore important to test whether gene expression correlated with protein expression and which cells within the tumours express that protein. This is discussed in Chapter 5.

Given that the transcriptomic data set was derived from tumour cells, however other cell types of tumour microenvironment (e.g. stroma) could influence the mRNA levels, I hypothesised that looking at the correlates of *MYC* expression in tumour cell lines could give insights about *MYC*'s role in immune evasion restricting the analysis to the pure melanoma cells.

Therefore, I collaborated with Sabrina Hogan (a colleague Marie Skłodowska-Curie PhD student) based at the University of Zürich. The Zürich group generated RNAseq data from 103 patient derived early passage melanoma cell lines. I asked Sabrina to take an agnostic approach to testing the correlations between *MYC* expression and the whole transcriptome in those cell cultures. These cell lines were lacking immune cells, which usually die in tumour cultures, and were fibroblast deficient (as described in Chapter 2). After the analysis, genes were ranked by their correlation coefficient with *MYC*. I was interested in the most negatively correlating genes as they could imply down-regulatory effect of MYC. Interestingly of the 50 genes most significantly negatively correlated with *MYC* five are involved in antigen processing and presentation (Figure 4.5):

- Major histocompatibility complex class I, B (*HLA-B*)*:*
  (R=-0.57, P=1.6x10^{-10})
- Major histocompatibility complex class I, C (*HLA-C*):
  (R=-0.45, P=1.8x10^{-6})
- Beta-2-microglobulin (*B2M*)*:*
  (R=-0.45, P=2.69x10^{-6})
- Transporter 1, ATP Binding Cassette Subfamily B Member (*TAP1*)*:*
  (R=-0.44, P=4.45x10^{-6})
- Endoplasmic Reticulum Aminopeptidase 1 (*ERAP1*)*:*
  (R=-0.44, P=5.0x10^{-6}).

In LMC the correlation of *MYC* with the genes above were:

- *HLA-B* (R=-0.17, P=3.7x10^{-6}*)*
- *HLA-C* (R=-0.15, P=7.4x10^{-5})
- *B2M* (R=-0.17, P=8.5x10^{-6})
- *TAP1* (R=-0.25, P=1.1x10^{-11})

- *ERAP1* (R=-0.17, P=5.5x10^{-6})

And in the TCGA metastatic samples:

- *HLA-B* (R=-0.24, P=5.1x10^{-6})
- *HLA-C* (R=-0.22, P=3.7x10^{-5})
- *B2M* (R=-0.24, P=5.5x10^{-6})
- *TAP1* (R=-0.17, P=0.001)
- *ERAP1* (R=-0.14, P=0.006).

When these 50 genes most negatively correlating with MYC in cell lines were tested for pathway enrichment in Reactome FIViz, the most significantly enriched pathways were class I MHC mediated antigen processing & presentation (FDR=1.2x10^{-6}), and interferon gamma singling (FDR=1.2x10^{-4}) (Table 4.4). On the other hand, the 50 genes most positively correlating with MYC were enriched in senescence-associated secretory phenotype (SASP) (FDR= 6.16x10^{-3}) and the cell cycle (FDR= 6.16x10^{-3}) (Table 4.5). Altogether these results showed that *MYC* was negatively correlated with immunity and positively with senescence and proliferation, which were similar to the ones obtained using the LMC transcriptome.

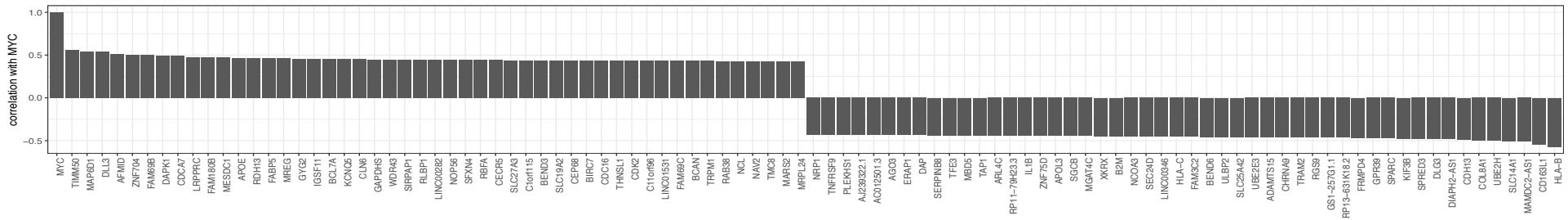


Figure 4.5: The 25 genes most positively and negatively correlated with *MYC* in melanoma cell lines data (Analysis by Sabrina Hogan)

Table 4.4: Enriched pathways in the gene list negatively correlated with *MYC* in patient derived melanoma cell lines

"GeneSet" is the pathway name, "Ratio of protein in Geneset" indicates ratios of numbers of genes contained in pathways to total genes in the Reactome FI network; "Number of protein in GeneSet" is the number of proteins within a given pathway; and "Protein from network" indicates number of genes from the input gene list per pathway. "Nodes" are proteins from the network.

GeneSet	*Ratio Of Protein In GeneSet*	*Number Of Protein In GeneSet*	*Protein From Network*	*P-value*	*FDR*	*Nodes*
Class I MHC mediated antigen processing & presentation(R)	0.0174	189	5	3.25E-08	1.20E-06	HLA-B, TAP1, HLA-C, SEC24D, B2M
Antigen processing and presentation(K)	0.0071	77	4	8.67E-08	1.56E-06	HLA-B, TAP1, HLA-C, B2M
Interferon gamma signaling(R)	0.0067	73	3	1.04E-05	1.25E-04	HLA-B, HLA-C, B2M
Phagosome(K)	0.0142	154	3	9.54E-05	8.59E-04	HLA-B, TAP1, HLA-C

Table 4.5: Enriched pathways in the gene list positively correlated with *MYC* in patient derived melanoma cell lines

"GeneSet" is the pathway name, "Ratio of protein in Geneset" indicates ratios of numbers of genes contained in pathways to total genes in the Reactome FI network; "Number of protein in GeneSet" is the number of proteins within a given pathway; and "Protein from network" indicates number of genes from the input gene list per pathway. "Nodes" are proteins from the network.

GeneSet	*Ratio Of Protein In GeneSet*	*Number Of Protein In GeneSet*	*Protein From Network*	*P-value*	*FDR*	*Nodes*
Senescence-Associated Secretory Phenotype (SASP)(R)	0.0048	52	2	1.37E-04	6.16E-03	CDK2, CDC16
APC/C-mediated degradation of cell cycle proteins(R)	0.0076	83	2	3.46E-04	6.16E-03	CDK2, CDC16
Progesterone-mediated oocyte maturation(K)	0.009	98	2	4.82E-04	6.16E-03	CDK2, CDC16
Ribosome biogenesis in eukaryotes(K)	0.0094	102	2	5.22E-04	6.16E-03	NOP56, WDR43

4.4 Summary

Whole transcriptomic differential expression analysis of the immune subgroups followed by enrichment and network analysis indicated that:

- The pathways enriched in the Low Immune Subgroup were associated with proliferation and cell cycling.
- The nodal gene identified in the Low Immune Subgroup, with the most central role in holding the network together was *MYC.*
- From the presented results based on LMC, TCGA and cell lines supported by the literature: *MYC* was considered as an immune evasive candidate, functioning via down regulation of *HLA-B* or possibly more MHC class I molecules.
- The pathways enriched in the High Immune Subgroup were associated with many crucial immune pathways.
- The nodal gene identified in the High Immune Subgroup was *NFKB1,* a transcription factor known to be expressed in all cell types and for its role in immune reactions.

4.5 Discussion

The three immune subgroups generated based on imputed immune cell scores (as described in the previous chapter) allowed the identification of key pathways and genes associated with the immune response to melanoma.

4.5.1 Methodological aspects

With the advent of large-scale gene expression data, many tools to analyse pathway enrichment have emerged, such as Database for Annotation, Visualization and Integrated Discovery (DAVID) [175], which is no longer updated, Kyoto Encyclopedia of genes and Genomes (KEGG) [176], Gene Set Enrichment Analysis (GSEA) [177], and Reactome [178]. These tools however are limited to only one type of data analysis, which is generally pathway enrichment. For the analysis of the genes upregulated in the Low and High Immune Subgroups I intended to use a tool used to analyse pathway enrichment and protein-protein interaction networks. The pathway enrichment

analyses give an overview of the biological characterisation of the input gene list, whereas network analysis enables to identify small subsets of those genes, which are functionally the most important. Amongst the most known bioinformatic software, that provide a possibly of these two types of analyse are: STRING ((https://string-db.org)) and Reactome FIViZ [160]. However, STRING does not support an analysis of more than 20,000 genes, hence I used Reactome FIViz. This tool has several additional advantages when compared to STRING, for example (as mentioned in the background section) it allows direct analysis of gene expression and mutational data, although I did not use this function during my analysis. As an input for Reactome FIViz, I have only used the gene names, not gene expression, because the gene list already came from differential expression analyses with upregulation in one group of tumours compared to another, which meant that all the genes in the list were positively correlated.

Irrespective of the software used, the biological networks predicted using curated databases can be affected by curation bias [179][180]. For example, more studied genes/protein are likely to have more identified annotations and network connections than those less studied. It is well known that transcription factors, whose function is to control other genes, are among the most well studied molecules due to their wide-ranging impact. It was therefore not entirely surprising that genes with the highest centrality in both the Low Immune and High Immune Subgroups were transcription factors (*MYC* and *NFKB1).* Validation of these genes in an independent experiment of melanoma cell cultures, in particular *MYC* for its role in immune evasion, was necessary and provided convincing results.

Additionally, in order to ensure validity of the results obtained on the gene expression level through the network analyses, they were compared to protein data from immunohistochemistry staining and to DNA copy number alteration data in the next chapter (Chapter 5).

4.5.2 Biological aspects

The genes upregulated in the Low Immune Subgroup were associated with proliferation and metabolic activity, suggesting that tumours in this subgroup were very aggressive. These agnostic genomic results are consistent with the results showed in Chapter 3, where the mitotic number assessed by dermatopathologists was

significantly higher in the Low Immune Subgroup, and the subgroup was associated with poorer survival. The most pivotal gene in the protein-protein interaction network in this subgroup was *MYC,* a widely studied proto-oncogene known to regulate the transcription of many genes involved in cell growth, apoptosis and metabolism by forming a heterodimer with its partner MAX (MYC-associated factor) and binding to E-box (enhancer box) of target genes [181].

In melanoma it has been previously reported, that MYC was involved in metastases and invasiveness [182]. However, in recent years researchers have examined MYC from a slightly different biological viewpoint, such as its involvement in immunosuppressive processes, and were able to provide some evidence for that in various cancer types [183][184][185], but not for melanoma. Given this evidence from literature and my observation of a negative correlation between MYC expression and immune pathways in the LMC dataset, I explored a possible immunosuppressive function of *MYC* in primary melanoma by further using cell line data from a collaboration (Sabrina Hogan, University of Zürich). The agnostic analysis of the cell line transcriptome revealed a similar trend as seen in patients of LMC dataset, i.e. that *MYC* was negatively associated with the immune function overall and the antigen processing machinery in particular. This was consistent with the report by Versteeg et al. [186], in which an inverse relationship between HLA class I and *MYC* expression was seen in melanoma cell lines. Evidence that *MYC* down-regulates the expression *HLA-B*, by directly binding to its proximal promoter independently of MAX has also been reported in melanoma cell lines [187]. These studies however were performed around 20 years ago, and up to nowadays this relationship of MYC and HLA-class I was not shown in human melanoma samples. More recently it was shown that *MYC* disrupts HLA class II-mediated immune recognition in human B cell tumours [188].

The comparisons between *MYC* and *CTNNB1* expression in the tumours indicated that both *MYC* and *CTNNB1* could contribute to the immune evasion simultaneously as the Low Immune Subgroup was comprised mostly of samples with both *MYC* and *CTNNB1* high expression. The whole genome comparison between the *MYC* and *CTNNB1* dominant groups revealed that there was no clear distinction seen in relation to the immunosuppressive pathways between these two groups in the whole dataset, however the genes upregulated in the *MYC* dominant group were associated with

keratinization. These observations might suggest that the *MYC* dominant group was contaminated by the expression of genes originating from keratinocytes. Similar results and conclusions were shown by Dr Jérémie Nsengimana in the analyses of β-catenin signalling as an immune evasion mechanisms in the LMC [76]. In this study, it was found that only *MYC* was upregulated (not *CTNNB1*) in the CIC6, which was called immune low/keratin high and it was thin group of samples. To make sure that *MYC* overall was not associated with keratin contamination I tested a correlation of *MYC* and *FLG* (filaggrin expressed by keratinocytes) using Spearman correlation, and the results confirmed that there was no correlation (R=0.007, P=0.8).

In summary the above evidences suggested that the immune evasive mechanisms driven by *MYC* and *CTNNB1* were not mutually exclusive and they co-occurred in a large proportion of the Low Immune Subgroup tumours. It is of note that, although *MYC* is downstream of β-catenin pathway, it does not eliminate the possibly that *MYC* is controlled by different mechanisms such as *MYC* amplifications, which was tested and this is reported in Chapter 5. Additionally, literature suggested that MYC might also control β-catenin signalling [149].

Overall, my observations and those from the literature manifest evidence that *MYC* is contributing to the immune evasion in primary melanoma by impairing the antigen presentation and processing machinery at several levels including by reducing the protein levels of *HLA-B* on the cell membrane.

Unsurprisingly the genes upregulated in the High Immune Subgroup were assigned overall to the immune function. Many immune checkpoint genes were also identified in this subgroup, as well as Tregs and MDSCs. Taking together upregulation of immunosuppressive and immuno-stimulating genes/immune cell scores in the high Immune Subgroup suggested a coordination of the immune cell populations as a whole or that the interpretation of immune cell subgroup infiltration from transcriptomic datasets may be insensitive to some subtle variations. However, it was recently reported that Treg numbers and checkpoint molecules increase as a results of a homeostatic mechanism when CD8+ T cell numbers increase in melanoma [189][76].

NFKB1 was the most nodal gene within the network created based on genes upregulated in the High Immune Subgroup. The majority of important *NFKB* family

genes such as Nuclear Factor Kappa B Subunit 2 (NFKB2), Proto-oncogene cRel (REL), Transcription factor p65 (RELA), Transcription factor RelB (RELB), and many more, were also upregulated in the High Immune Subgroup indicating activation of that pathway (Appendix A.2.2).

However, the expression of an important gene (*RELA*) coding for a protein that forms a heterodimer with NF-κB p105 (*NFKB1*) before its translocation to the nucleus and activation of transcription, did not vary across the immune 3 subgroups. This might reflect that this protein is constitutively expressed (meaning that its expression does not vary) in different tissues types and their biological states.

Nevertheless, the products of *NKFB1* such as NF-κB p105 precursor, which after phosphorylation becomes p50, might form different combination of heterodimers and homodimers with the rest of the NF-κB family members. The functions of each dimer vary, and they either might activate the transcription of target genes or repress it [190]. Using only gene expression data, we were not able to investigate this process deeper.

NF-κB singling is known for its immune regulatory functions as well as for regulation of apoptosis and cell survival [190]. Abnormal NF-κB singling and *NFKB1* gene expression particularly was found to have carcinogenic functions and promote tumour progression by inflammation (for example via STAT3) as well as to act as tumour suppressor [190][191][192]. Importantly, these functions depend on which components of tumour microenvironment (such as immune cells, cancer associated fibroblast, or tumour cells) have the abnormal NF-κB singling [193]. Assuming that the signal of NF-κB pathway of the LMC tumours was coming from the tumour cells, (which was tested in the next chapter), the results from this chapter suggested that this gene and overall NF-κB signalling was associated with higher anti-tumour immune responses. However, this observation did not explain the causality of the tumour NF-κB signalling activation. It could be that the immune cells attracted to the tumour microenvironment produced signals to activate NF-κB signalling by the tumour cells. Or on a contrary tumour cells prior being attracted by the immune cells, had NF-κB signalling active resulting in secretion of chemo attractant cytokines, which facilitated recognition of the tumour by the immune cells.

One study examined the expression of tumour NF-κB genes in lung cancer, and reported that this pathway promoted T cell- mediated immune surveillance [194]. Moreover, another study showed that hyper-activation of NF-κB signalling within the tumour microenvironment contributes to reprograming of chemokine microenvironment to enhance recruitment of cytolytic T cells in colorectal tumours [195]. Gastric cancer was also shown to be caused by loss of *NFKB1* particularly with aberrant inflammation within the tumour [196].

The literature supported the observations reported in this chapter such as positive association of NF-κB signalling and *NFKB1* with high immune signals within the tumour, however it is crucial to test where the NF-κB signalling gene expression originated from (see next chapter).

Chapter 5

Evaluation of the transcriptomic results on DNA and protein level

5.1 Aims

Herein the results from the previous chapters, which were based on the gene expression level, were evaluated by using protein immunohistochemistry (IHC) and DNA copy number alteration data (CNA). These analyses were performed only in a subset of the tumours due to technical and ethical reasons.

The aims were:

- To analyse CNA data for *MYC* and *NFKB1*, pathways and other genes from the top enriched pathways from the previous chapter aiming to observe some associations between CNAs and gene expression, the three immune subgroups and survival.
- To perform immunohistochemical staining for MYC, HLA-B and NFKB1 (NF-κB p105) and correlate the results with mRNA level of the corresponding genes aiming to observe the positive correlations.

5.2 Evaluation of CNA

5.2.1 Introduction

CNA data used in this thesis were generated based on next generation sequencing (NGS) of the FFPE tumours. The methods for analysis of structural variation of the genome in cancer have been identified previously using NGS data [197], however samples from fresh tissues were generally used, not FFPE. The performance of NGS highly depends on the quality of the DNA extracted from the analysed tissues [198] and it is well known that the genetic material extracted from FFPE tissues is degraded but in melanoma, it is very often the only available. In addition, melanomas are generally very small (compared to other tumours), and the quantity of the genetic material available is therefore limited. Moreover, the presence of melanin in melanoma tissue

further complicates the processing of genomic material as it has been shown to interact with polymerase in PCR manipulations, inhibiting its function [199] and affecting library production for sequencing.

Our group has developed methods which enabled to produce reliable CNA data from FFPE samples from relatively small quantities of DNA in the range of 25-1000ng (Filia et al., in press). Joey Mark Santiago Diaz (a colleague Marie Skłodowska-Curie PhD student in the group) has developed quality control and analysis pipelines for these data as part of his PhD project, and these data are now available.

5.2.2 Methods

Using the CNA data for 276 samples, I evaluated the amplifications and deletions of the nodal genes (*MYC* and *NFKB1)* identified in the network analyses in Chapter 4. Moreover, because the NF-κB and IFN-γ pathways were the most enriched pathways in the High Immune Subgroup, I subsequently asked if some of additional genes within these pathways were disrupted in the Low Immune Subgroup. For the NF-κB pathway, I chose these key genes involved in its signalling:

- Nuclear Factor Kappa B Subunit 2 (*NFKB2*)
- Proto-oncogene cRel (*REL*)
- Transcription factor p65 (*RELA*)
- Transcription factor RelB (*RELB*).

Furthermore, following a discussion with Prof. Ulf Klein (an immunologist from University of Leeds, specialising in NF-κB signalling), I included the following genes associated with positive regulation of this signalling pathway:

- Conserved Helix-Loop-Helix Ubiquitous Kinase (*CHUK*) [200]
- Interleukin-1 receptor-associated kinase-like 2 (*IRAK2*) [201]
- Mitogen-activated protein kinase kinase kinase 14/ NF-kB-inducing kinase (*MAP3K14/NIK*) [202]
- Myeloid differentiation primary response protein MyD88 (*MYD88*) [203]
- Mitogen-activated protein kinase kinase kinase 7 (*MAP3K7/TAK1*) [204]
- Inhibitor of nuclear factor kappa-B kinase subunit beta (*IKBKB*) [205]
- NF-kappa-B inhibitor alpha (*NFKBIA*) [206].

Interleukin-1 receptor-associated kinase-like 1 (*IRAK1*)[207] and Inhibitor of nuclear factor kappa-B kinase subunit gamma (*IKBKG*)[208] were also suggested, but the CNA data did not cover these genes, because they were located in the "black listed" (difficult to be sequenced) regions of the genome.

From the IFN-γ signalling pathway I have assessed copy number changes of the key genes involved in this signalling: tyrosine-protein kinase (*JAK1* and *JAK2*), signal transducer and activator of transcription 1-alpha/beta (*STAT1*), and interferon gamma (*IFNG*)[209]. In addition to these genes, I included β-catenin (*CTNNB1*) to test whether activation of β-catenin pathway is controlled by changes on DNA level.

The CNA data were generated at a continuous scale as normalised zero-centred read counts of DNA sequences per window of 10kb. Negative values indicate DNA fragment deletion and positive value indicate DNA fragment amplification. I used the median value of the ratio of the sample read count to normal control read count of the windows (10K) covering the gene of interest as the copy number estimate. Most of the analysed genes spanned more than one 10K window except *MYC,* which covered only one 10K window and initial analyses for *MYC* were not conclusive and therefore I chose to analyse a region commonly amplified in cancer around *MYC* (23091 bases) [210] rather than *MYC* alone.

To define amplifications and deletions, I developed an ad-hoc categorisation of continuous CNA data into amplification and deletion, based on visual examination of the CNA distribution for all genes of interest. An arbitrary cut-off, value of |0.3| (ratio of a read count of tumour to normal) was used. I defined deletion and amplification as cut-off point of windows (covering a gene region) median at below -0.3 or above 0.3 respectively for the majority of the genes and at below -0.45 or above 0.45 for one gene (*JAK2)*. These cut-offs however do not indicate whether the aberration is homo- or heterozygous although the assumption is that the deeper the aberration the higher chances for homozygous change or even polyploidy.

It was examined whether copy number alterations of the genes of interest were associated with the three immune subgroups using the Fisher's exact test. I analysed the expression level of corresponding genes among the aberrations using Kruskal-Wallis test for each gene. Moreover, I tested the survival of patients with and without

the aberrations in Cox proportional hazard regression models, unadjusted and adjusted for AJCC stage. I further tested whether the prediction of survival for patients with *MYC* and *NFKB1* aberrations was additive by including them jointly in the analysis.

For the CNA visualisation I used the ComplexHeatmp package in R [211] to create "Oncoprint" graph, which enables to visualise/detect mutual exclusivity of the aberrations.

As stated in Chapters 3 and 4, our group reported evidence that β-catenin signalling pathway is upregulated in 30% of LMC tumours overall and in 59% of the low immune tumours indicating its immunosuppressive function [76]. Apart from testing interaction of *MYC* and *CTNNB1* on gene expression level in previous chapter, I also examined the overlap between β-catenin gene expression and the possible immunosuppressive mechanisms driven by deletions of NF-κB pathway genes and their joint effect on survival using Cox-proportional hazard regression. I created and tested a NF-κB CNA score (*NFKB1*, *NFKB2*, *CHUK*, *MAP3K7*, *IRAK2*, *MYD88*) for which at least one alteration was observed within a tumour. The result from this kind of analysis could indicate whether these pathways were independent from each other.

5.3 Evaluation of Immunohistochemical staining

5.3.1 Methods

Stored unstained FFPE sections of available primary tumours from the LMC were used to assess the protein-level and mRNA level correlations. The terms of the ethical approval for this study were that it should be avoided to destruct of tumour blocks which might be required by the patient for further clinical testing, hence a limited number of samples was available for this analysis. The sections for IHC could only be cut from tumour blocks of deceased study participants both either from melanoma or non-melanoma causes. Dr Jonathan Laye and Tracey Mell sectioned the tumour blocks and mounted the 5µm sections on Superfrost Plus slides (Thermo Fisher Scientific, UK). Dr Jonathan Laye and I performed immunohistochemical staining of the slides using IntelliPath FLX detection reagents (MenaPath, A. Menarini Diagnostics, UK) by heat antigen retrieval, blocking, and Hematoxylin counter-staining according to

manufacturer's protocol. Both procedures were described in the Materials and Methods section.

The most influential genes (*MYC* and *NFKB1*) identified from the analysis of the protein-protein interaction networks, which were created using genes upregulated in the High and Low Immune Subgroup (see Chapter 4) were examined. Moreover, *HLA-B* was also analysed, which was negatively correlated with *MYC* in the LMC, TCGA (metastatic samples) and in melanoma cell lines, consistently with literature [187]. The antibodies (Abcam, UK) used for staining were: anti-MYC (ab32072; suggested antibody concentration: 5 μg/ml), anti-NF-κB p105 (ab32360; suggested antibody concentration: 1/250 - 1/500), and anti-HLA-B (ab193415; suggested antibody concentration: 1/20 - 1/200).

5.3.1.1 *Antibody optimisation*

The antibodies were optimised using the available tissues that are known to express the analysed proteins in high levels. The tissue type for the antibody staining was chosen based on the data at the Human Protein Atlas [143]. Moreover, the Human Protein Atlas and UniProt portal [212] provided the information of the cellular localisation of protein of interest.

The optimisation of an antibody is a process by which we can test whether the antibody is specific to the analysed protein (detected in the predicted cellular localisation) as well as to assess which concentration would be the most suitable to allow scoring of the staining of the tissue of interest – here melanoma tumours. Mainly the staining should not be too strong to overcall the positive staining and also should not be too weak to miss the specific staining.

The staining for the antibody optimisation was performed by Dr Jonathan Laye and myself using IntelliPath FLX detection reagents (MenaPath, A. Menarini Diagnostics, UK) according to manufacturer's protocol.

The anti-MYC antibody was optimised using healthy human tonsil tissue in four different concentrations (μg/μl): 1:50, 1:100, 1:150, and 1:200. The anti-HLA-B was optimised on healthy human skin tissue in concentrations (μg/μl): 1:50, 1:100, 1:150. Finally, the anti-NF-κB p105 was optimised on healthy human sentinel lymph node

biopsies in six concentrations (μg/μl): 1:100, 1:200, 1:250, 1:375, 1:500, 1:600. The concentration of the antibody for the subsequent tumour staining was chosen based on its optimal intensity to detect the signal.

5.3.1.2 *Scoring of the stained slides*

The scoring of the tumours after staining was developed by pairs of observers (by myself for all of the slides and by independent assessors: Prof. Julia Newton-Bishop, Dr Jonathan Laye and Sathya Muralidhar (Marie Skłodowska Curie PhD student) for selected antibodies). All the assessors were blinded to the transcriptomic data while scoring.

Anti-MYC staining was performed on 48 slides. Light microscopy at 10X magnification was used to evaluate the expression of MYC and HLA-B due to homogenous staining. Staining scores for these two proteins were recorded from the regions immediately surrounding the tumour core - the 'punch hole' (Figure 5.1). MYC was assessed by nuclear staining as this protein is known to function within the nucleus as it is a transcription factor [143][212] (Table 5.1).

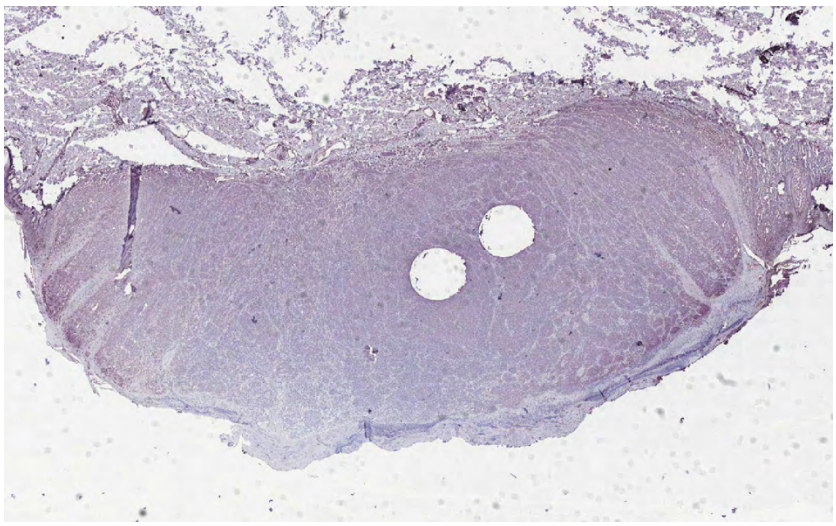


Figure 5.1: Example of a tumour with the two "punch holes" after the core sampling using 0.6mm microarray needle

Table 5.1: Scoring system for MYC IHC staining

IHC MYC Staining	
Cored region staining description	***Score***
Nuclear staining absent	0
Weak to moderate nuclear staining	1
Intense nuclear staining	2

The same scoring levels were utilised to assess the staining of HLA-B, however restricted to membranous staining, because it is the final localisation of this protein to present antigens to the immune cells [143][212]. (Table 5.2). In total 30 slides were stained for HLA-B.

Table 5.2: Scoring system for HLA-B IHC staining

IHC HLA-B Staining	
Cored region staining description	***Score***
Membranous staining absent	0
Weak to moderate membranous staining	1
Intense membranous staining	2

In the tumours that were sampled/cored more than once the staining scores were generated and if these were contradictory the slides were not used for subsequent analyses. Anti-NF-κB p105 was evaluated in cytoplasm as it is known to be detected in this location in inactive form and in the nucleus, for detection of its active transcription factor function [143][212]. The staining of NF-κB p105 for 29 tumours was evaluated using 20X magnification for cytoplasm around the core and 40X magnification for nuclei across the whole slide. Cytoplasmic staining was graded as in Table 5.3.

Table 5.3: Scoring system for cytoplasmic NF-kB p105 IHC staining

IHC NF-κB p105 Staining	
Tumour area staining description	***Score***
Cytoplasmic staining absent	0
Weak cytoplasmic staining	1
Moderate cytoplasmic staining	2
Intense cytoplasmic staining	3

Because the nuclear activity of NF-κB signalling is transient and not all the nuclei were positive at the same time, the scoring was assessed as follows: if more than 5 nuclei were positive in the tumour I recorded the staining as “positively stained”, otherwise as “negatively stained” (Table 5.4).

Table 5.4: Scoring system for nuclear NF-κB p105 IHC staining

IHC NF-κB p105 Staining	
Tumour area staining description	***Score***
<5 nuclei positively stained	Negative
≥5 nuclei positively stained	Positive

NF-κB p105 is known to be expressed by the immune cells, hence the infiltrated lymphocytes were scored as well for cytoplasmic and nuclear staining, either “positive” or “negative” (Table 5.5).

Table 5.5: Scoring system for TILs NF-kB p105 IHC staining

IHC NF-κB p105 Staining	
TILs staining description	**Score**
Cytoplasmic/nuclear staining absent	Negative
Cytoplasmic/nuclear staining present	Positive

Additional quantification of MYC and HLA-B staining was performed using light microscopy in conjunction with Nuance software v.3.0.1.2 (PerkinElmer, Inc.). Tissue sections were examined under 20X magnification, and consecutive MYC and HLA-B-stained images of the most representative part of the tumour were digitally scanned. Spectral analysis was performed to quantify the light signal derived from immunohistochemical chromagen, haematoxylin counterstain and signal co-localised from both chromagen and counterstain. Threshold levels were arbitrarily set to fix the strength of haematoxylin and chromagen signal deemed to be positive and for each tumour the signals were recorded as the percentage of positive pixels per scoring area.

The chromagen percentage score was utilized to quantify HLA-B staining. Because haematoxylin is a largely nuclear stain, MYC staining was quantified using the co-

localised signal derived from chromagen and counterstain to derive a MYC-nuclear expression score.

5.3.1.3 *Statistical analysis*

For the statistical analyses of MYC (tumour nuclear) and HLA-B (tumour membranous) staining, I tested the variation of the mRNA expression of the corresponding proteins, across tumours scored 0 to 2 for staining using Kruskal-Wallis test as well as between pooled scores: negative (0) and positive (1 and 2 pooled together), using Mann-Whitney U test in order to obtain higher statistical power. For NF-κB p105 staining I tested *NFKB1* mRNA expression change among the cytoplasmic four scores as well as pooled scores: negative (0) and positive (1,2 and 3 pooled) and nuclear scores (negatively vs positively stained) also in order to obtain higher statistical power. I moreover compared the nuclear staining of the tumour and nuclear staining of the TILs using the Chi-squared test. For the continuous scoring (using the Nuance software) of HLA-B and MYC I performed Spearman's rank correlation between mRNA and IHC staining excluding samples where MYC was detected in less than 1% of pixels, which I considered as absent.

5.4 Results

5.4.1 DNA copy number changes in primary melanoma in relation to immune response

Since up-regulation of *MYC* and down-regulation of *NFKB1* expression (identified as the nodal genes) were observed in the Low Immune Subgroup, I hypothesised that amplifications of *MYC* and deletions of *NFKB1* could be more common in these tumours. Using CNA from a subset (N=276) of the LMC tumours firstly the histograms for the CNA data per each gene were examined (Figure 5.2). Usually it was observed that for a particular gene, either deletion or amplifications were observed.

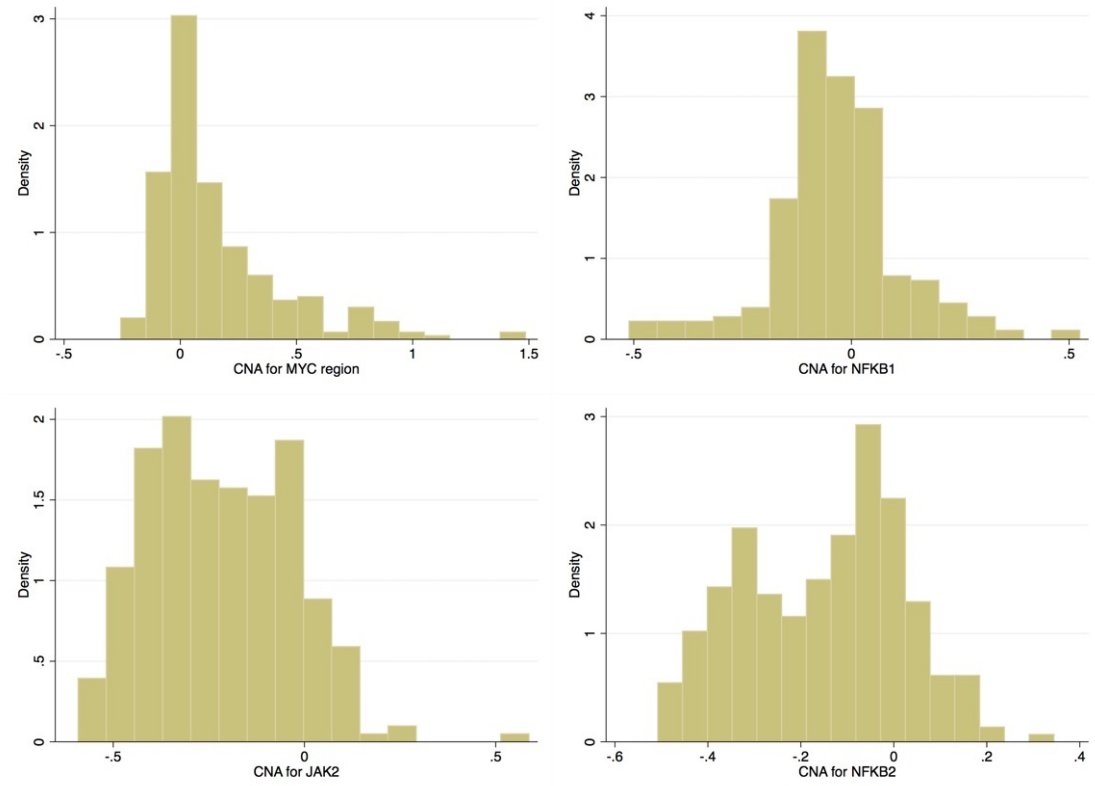


Figure 5.2: Examples of histograms representing the distribution of CNA for selected genes

Amplifications and deletions were generally exclusive per analysed gene.

MYC amplifications were found in 29% and *NFKB1* deletions in 14% of the Low Immune Subgroup, which was significantly higher than in the Intermediate or the High Immune Subgroup (P= 0.02 for *MYC*, P=0.0003 for *NFKB1*) (Table 5.6, Figure 5.3). Among the genes involved in the positive regulation of NF-κB signalling there were deletions in *NFKB2* (26%), *CHUK* (22%), *MYD88* (5%), *IRAK2* (5%), *MAP3K7* (17%), *JAK2* (10%), and *STAT1* (4%) in the whole dataset (Table 5.6, Figure 5.3). These copy number changes were not mutually exclusive (Figure 5.3) and were much more frequent in the Low Immune than in other subgroups (Table 5.6, Figure 5.3). The most significant distribution of the CNA across the three immune subgroups was observed for *JAK2* ($P=2x10^{-8}$), where 25% of samples had deletions in the Low Immune Subgroup, while only 1.6% in the High Immune Subgroup. Another gene with strong significant distribution of CNAs across the three immune subgroups was *NFKB2* and were

comparable to *NFKB1.* 38.6% of samples with NFKB2 deletions were identified in the Low Immune Subgroup, while 9.4% in the High Immune (P=0.0001).

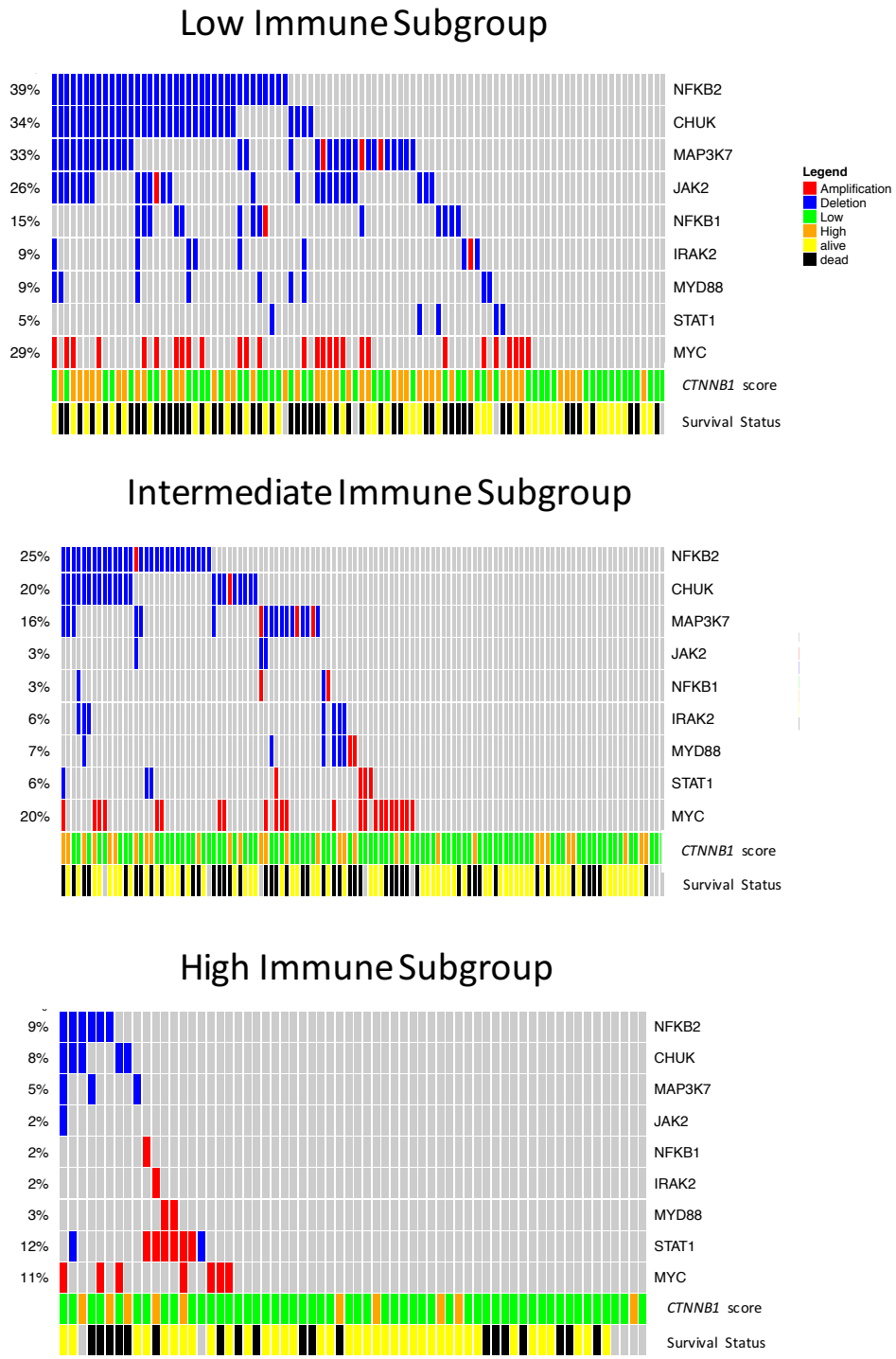


Figure 5.3: Graphical representation ("Oncoprint") of the significantly altered genes across the three immune subgroups

The two bottom rows represent *CTNNB1* gene expression and the survival status.

Table 5.6: Table representing CNA changes of all the analysed genes across the immune subgroups

	High Immune	***Inter. Immune***	***Low Immune***	***Total CNA***	***P***
TOTAL (N)	64	116	96		
MYC					
No change (N)	57	93	68	218	
No change (%)	89.06	80.17	70.83	78.99	
Amplification (N)	7	23	28	58	
Amplification (%)	10.94	19.83	29.17	21.01	**0.02**
NFKB1					
No change (N)	63	112	82	257	
No change (%)	98.44	96.55	85.42	93.12	
Amplification (N)	1	2	1	4	
Amplification (%)	1.56	1.72	1.04	1.45	
Deletion (N)	0	2	13	15	
Deletion (%)	0	1.72	13.54	5.43	**0.0003**
NFKB2					
No change (N)	58	87	59	204	
No change (%)	90.62	75	61.46	73.91	
Amplification (N)	0	1	0	1	
Amplification (%)	0	0.86	0	0.36	
Deletion (N)	6	28	37	71	
Deletion (%)	9.38	24.14	38.54	25.72	**0.0001**
CHUK					
No change (N)	59	93	63	215	
No change (%)	92.19	80.87	65.62	78.18	
Deletion (N)	5	22	33	60	
Deletion (%)	7.81	19.13	34.38	21.82	**0.0002**
IRAK2					
No change (N)	63	109	87	259	
No change (%)	98.44	93.97	90.62	93.84	
Amplification (N)	1	0	1	2	
Amplification (%)	1.56	0	1.04	0.72	
Deletion (N)	0	7	8	15	
Deletion (%)	0	6.03	8.33	5.43	**0.04**

	High Immune	***Inter. Immune***	***Low Immune***	***Total CNA***	***P***
MAP3K14					
No change (N)	63	106	86	255	
No change (%)	98.44	91.38	89.58	92.39	
Amplification (N)	1	4	3	8	
Amplification (%)	1.56	3.45	3.12	2.9	
Deletion (N)	0	6	7	13	
Deletion (%)	0	5.17	7.29	4.71	**0.2**
MYD88					
No change (N)	62	108	87	257	
No change (%)	96.88	93.1	90.62	93.12	
Amplification (N)	2	2	0	4	
Amplification (%)	3.12	1.72	0	1.45	
Deletion (N)	0	6	9	15	
Deletion (%)	0	5.17	9.38	5.43	**0.02**
MAP3K7					
No change (N)	61	98	64	223	
No change (%)	95.31	84.48	66.67	80.8	
Amplification (N)	0	3	3	6	
Amplification (%)	0	2.59	3.12	2.17	
Deletion (N)	3	15	29	47	
Deletion (%)	4.69	12.93	30.21	17.03	**0.00005**
REL					
No change (N)	62	103	85	250	
No change (%)	96.88	88.79	88.54	90.58	
Amplification (N)	2	11	8	21	
Amplification (%)	3.12	9.48	8.33	7.61	
Deletion (N)	0	2	3	5	
Deletion (%)	0	1.72	3.12	1.81	0.3
RELB					
No change (N)	61	108	87	256	
No change (%)	95.31	93.1	90.62	92.75	
Amplification (N)	3	8	8	19	
Amplification (%)	4.69	6.9	8.33	6.88	
Deletion (N)	0	0	1	1	

	High Immune	***Inter. Immune***	***Low Immune***	***Total CNA***	***P***
Deletion (%)	0	0	1.04	0.36	0.7
RELA					
No change (N)	58	103	80	241	
No change (%)	90.62	88.79	83.33	87.32	
Amplification (N)	4	3	3	10	
Amplification (%)	6.25	2.59	3.12	3.62	
Deletion (N)	2	10	13	25	
Deletion (%)	3.12	8.62	13.54	9.06	0.1
IKBKB					
No change (N)	63	103	73	239	
No change (%)	98.44	88.79	76.04	86.59	
Amplification (N)	1	8	13	22	
Amplification (%)	1.56	6.9	13.54	7.97	
Deletion (N)	0	5	10	15	
Deletion (%)	0	4.31	10.42	5.43	**0.001**
NFKBIA					
No change (N)	55	100	79	234	
No change (%)	85.94	86.21	82.29	84.78	
Amplification (N)	6	6	6	18	
Amplification (%)	9.38	5.17	6.25	6.52	
Deletion (N)	3	10	11	24	
Deletion (%)	4.69	8.62	11.46	8.7	0.5
JAK2					
No change (N)	63	113	71	247	
No change (%)	98.44	97.41	73.96	89.49	
Amplification (N)	0	0	1	1	
Amplification (%)	0	0	1.04	0.36	
Deletion (N)	1	3	24	28	
Deletion (%)	1.56	2.59	25	10.14	**$2x10^{-8}$**
STAT1					
No change (N)	56	109	91	256	
No change (%)	87.5	93.97	94.79	92.75	
Amplification (N)	6	4	0	10	
Amplification (%)	9.38	3.45	0	3.62	

	High Immune	***Inter. Immune***	***Low Immune***	***Total CNA***	***P***
Deletion (N)	2	3	5	10	
Deletion (%)	3.12	2.59	5.21	3.62	**0.02**
JAK1					
No change (N)	62	106	80	248	
No change (%)	96.88	91.38	83.33	89.86	
Amplification (N)	2	5	11	18	
Amplification (%)	3.12	4.31	11.46	6.52	
Deletion (N)	0	5	5	10	
Deletion (%)	0	4.31	5.21	3.62	0.05
IFNG					
No change (N)	60	107	83	250	
No change (%)	93.75	92.24	86.46	90.58	
Amplification (N)	1	5	4	10	
Amplification (%)	1.56	4.31	4.17	3.62	
Deletion (N)	3	4	9	16	
Deletion (%)	4.69	3.45	9.38	5.8	0.4
CTNNB1					
No change (N)	63	102	85	250	
No change (%)	98.44	87.93	88.54	90.58	
Amplification (N)	1	6	1	8	
Amplification (%)	1.56	5.17	1.04	2.9	
Deletion (N)	0	8	10	18	
Deletion (%)	0	6.9	10.42	6.52	**0.02**

Deletions rather than amplifications of *CTNNB1* were seen in the Low Immune Subgroup (10.4% in Low immune and none in High immune, P=0.02) (Table 5.6), although the gene expression was unaffected by these deletions (P=0.2) (Table 5.7). For the vast majority of other genes, the copy number changes were highly correlated with mRNA expression of the corresponding gene (Table 5.7). For example, the results were the most prominent for *NFKB2* (FC=1.6, P=$1.5x10^{-11}$).

Table 5.7: Variation in gene expression between the groups of tumours with and without CNA (for the genes significantly varying across the three immune subgroups)

Gene name	*Fold change*	*P*
MYC	0.8	0.0006
NFKB1	1.2	0.008
NFKB2	1.6	$1.5x10^{-11}$
CHUK	1.2	0.004
IRAK2	1.7	0.003
MYD88	1.2	0.02
MAP3K7	1.2	$4.8x10^{-10}$
STAT1	1.1	0.7

Importantly, copy number changes of *MYC* (amplifications) and *NFKB1 (*deletions*)* were strongly predictive of poor prognosis overall in univariable analysis and when AJCC stage was adjusted (*MYC* amplifications: adjusted HR=1.8 (95% CI 1.8-2.6, P=0.006; *NFKB1* deletions: adjusted HR=1.5 (95% CI 1.1-2.1, P=0.007) (Table 5.8, Table 5.9). Seven patients who had both amplifications of *MYC* and deletion of *NFKB1* had an even worsened prognosis (adjusted HR=3.7, 95% CI 1.6-8.5, P=0.002), (Table 5.8, Table 5.9), suggesting an additive detrimental effect of these copy number changes. The unadjusted survival analyses of *MYC* and *NFKB1* CNAs are plotted on Figure 5.4. It is of note, that all seven patients with alterations in both genes have died.

Other deletions associated with melanoma specific survival were observed in *NFKB2*, *CHUK*, *IRAK2*, *MYD88*, *MAP3K7*, *JAK2*, *STAT1* (Table 5.8), and *CHUK, MYD88, IRAK2 or JAK2* remained significant when adjusted for AJCC stage (Table 5.9).

Table 5.8: Univariable melanoma specific survival anaysis of CNAs

Gene name	***HR***	***P***	***95% Conf. Interval***	
MYC	1.79	**0.004**	1.20	2.66
NFKB1	1.67	**0.001**	1.24	2.26
NFKB2	1.22	**0.044**	1.01	1.48
CHUK	1.32	**0.007**	1.08	1.61
IRAK2	1.40	**0.044**	1.01	1.93
MYD88	1.57	**0.004**	1.15	2.15
MAP3K7	1.24	**0.054**	1.00	1.54
JAK2	1.51	**0.001**	1.18	1.95
STAT1	1.04	0.871	0.63	1.72

Table 5.9: Melanoma specific survival anaysis of CNAs adjusting AJCC stage

Gene name	*HR*	*P*	*95% CI*	
MYC	1.76	**0.006**	1.18	2.64
NFKB1	1.52	**0.007**	1.12	2.05
Amp of MYC + Del of NFKB1	3.70	**0.002**	1.60	8.52
NFKB2	1.16	0.136	0.95	1.42
CHUK	1.31	**0.009**	1.07	1.60
IRAK2	1.38	**0.053**	1.00	1.91
MYD88	1.49	**0.013**	1.09	2.04
MAP3K7	1.14	**0.252**	0.91	1.43
JAK2	1.40	**0.009**	1.09	1.81
STAT1	0.97	**0.908**	0.59	1.60

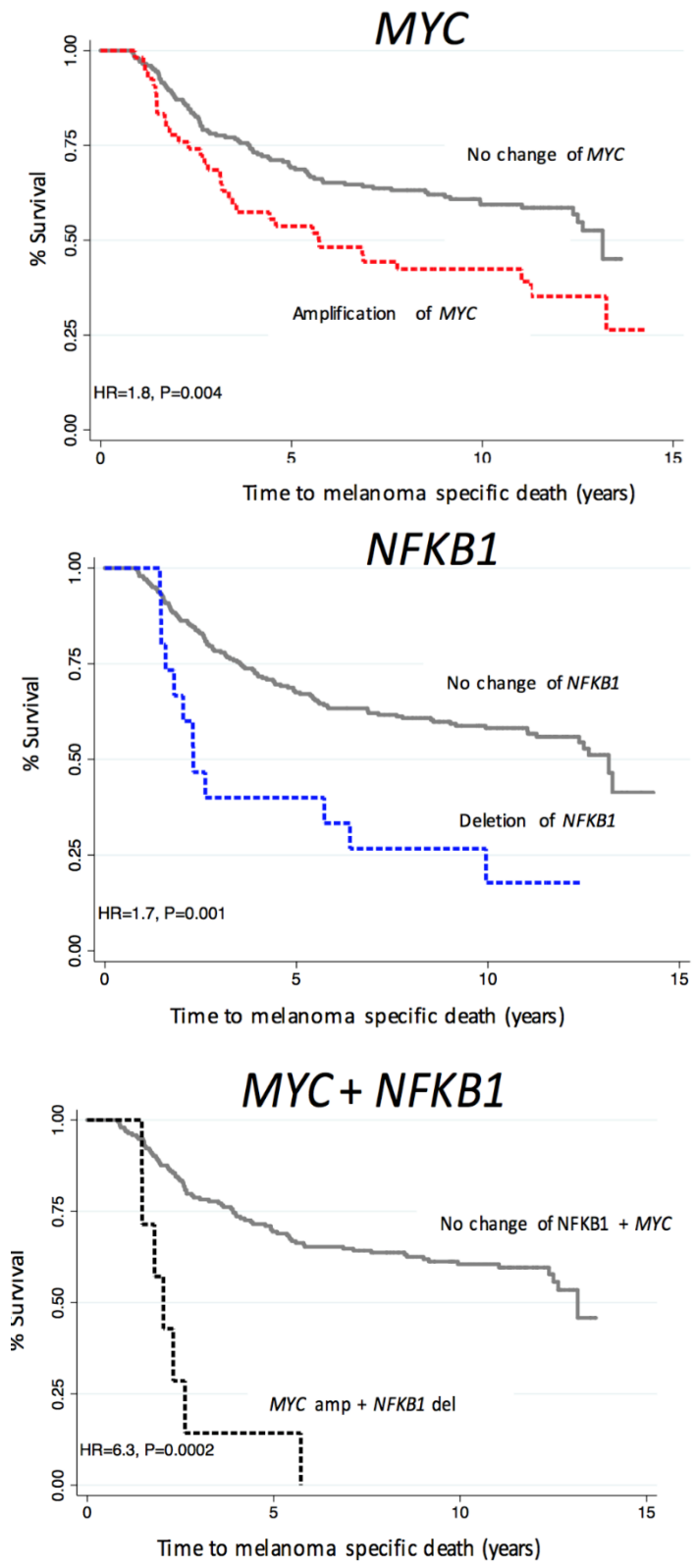


Figure 5.4: Kaplan Meier plots for amplifications of *MYC* vs no change, deletion of *NFKB1* vs no change, and amplification of *MYC* + deletion of *NFKB1* vs no change

As our group previously reported that β-catenin pathway is involved in immune evasion, I tested its relation with deletions in the NF-κB pathway to see if the 2 mechanisms are interrelated or independent. I compared the CNAs of genes from the NF-κB pathway and *CTNNB1* expression dichotomised into high (30%) and low (70%) using the software Xtile [213] by Dr Jérémie Nsengimana [76]. This analysis showed some overlap between the two immune evasion mechanisms but also some heterogeneity. More specifically, 15% of tumours had increased *CTNNB1* expression alone, 32% had a deletion in at least one gene of the NF-κB pathway without *CTNNB1* overexpression, whilst 31% had both (i.e. increased *CTNNB1* and a deletion in at least one gene). Figure 5.3 showed that tumours of the Low immune group were more likely to have *NFKB1* deletions and high *CTNNB1* overexpression, as well as more deaths. In prognostic terms, in the whole LMC dataset the effect of *CTNNB1* upregulation was HR=2.2, $P=5x10^{-5}$, 95%CI 1.5-3.1; the effect of any deletion in the NF-κB pathway was HR=2.03, $P=2x10^{-4}$, 95%CI 1.4-3; and the effect of a combination of these two pathways was HR=3.4, $P=5x10^{-5}$, 95% CI 2.2-5.5 (Figure 5.5, Table 5.10).

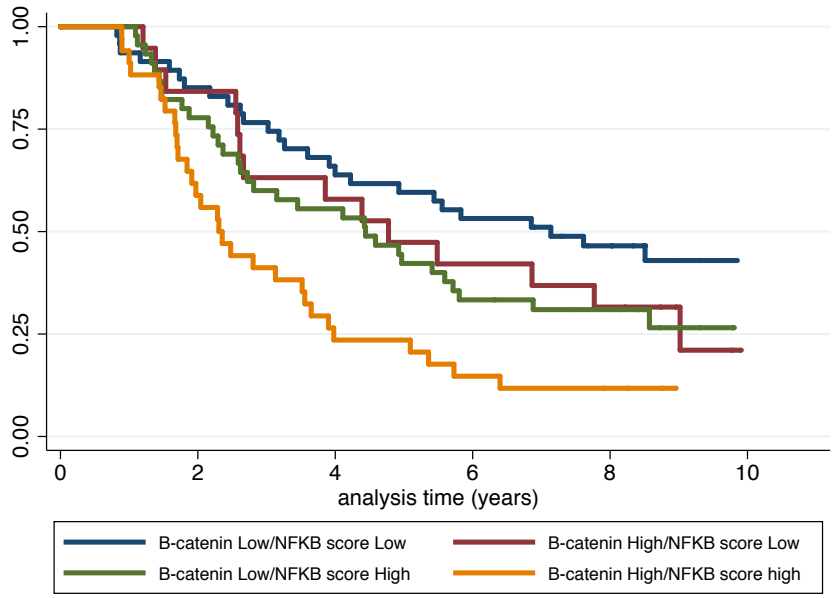


Figure 5.5: Kaplan Meier plot for combination of Beta-catenin and NFKB CNA scores

Table 5.10: Melanoma specific survival analysis of combined Beta-catenin expression and *NFKB* CNA scores

Combinations of the scores	*HR*	*P*	*95% CI*	
B-catenin Low/NFKB score Low	-	-	-	
B-catenin High/NFKB score Low	2.03	0.019	1.12	3.67
B-catenin Low/NFKB score High	2.02	0.005	1.24	3.29
B-catenin High/NFKB score High	3.35	$5x10^{-5}$	2.02	5.55

5.4.2 Immunohistochemical staining

5.4.2.1 *Antibody optimisation*

The concentrations of the antibodies were chosen based on visual examination of the intensity of staining. For the anti-MYC antibody optimised on healthy human tonsil tissue I used the concentration of 1:100 (Appendix A.3.1). The same concentration was used for anti-HLA-B antibody optimised on healthy human skin (Appendix A.3.2). A concentration of 1:250 was selected for the anti-NF-κB1 antibody optimised on healthy human sentinel lymph node biopsies. (Appendix A.3.3).

5.4.2.2 *Scoring of the stained slides and statistical analysis*

The anti-MYC staining in the melanoma tumours was primarily observed in the nuclei and the staining was mostly homogenous, when assessing the staining by eye under light microscope. There was a trend to stronger staining for MYC protein in association with gene expression across the three staining groups 0, 1 and 2 but this did not reach statistical significance (P=0.14). There was a significant correlation when comparing high staining (scores 1 and 2 pooled) and absence of staining (0) (P=0.06) (Figure 5.6).

The staining of HLA-B was predominantly observed on the cellular membrane (as expected given the HLAs function) and similarly to MYC the staining was mostly homogenous. The transcriptomic expression for *HLA-B* was incremental across the three ascending categories of membranous scores (P=0.006) (Figure 5.7). I observed similar results when I reduced the number of categories to 2 (negative and positive staining), however with increased significance (P=0.002) (Figure 5.7).

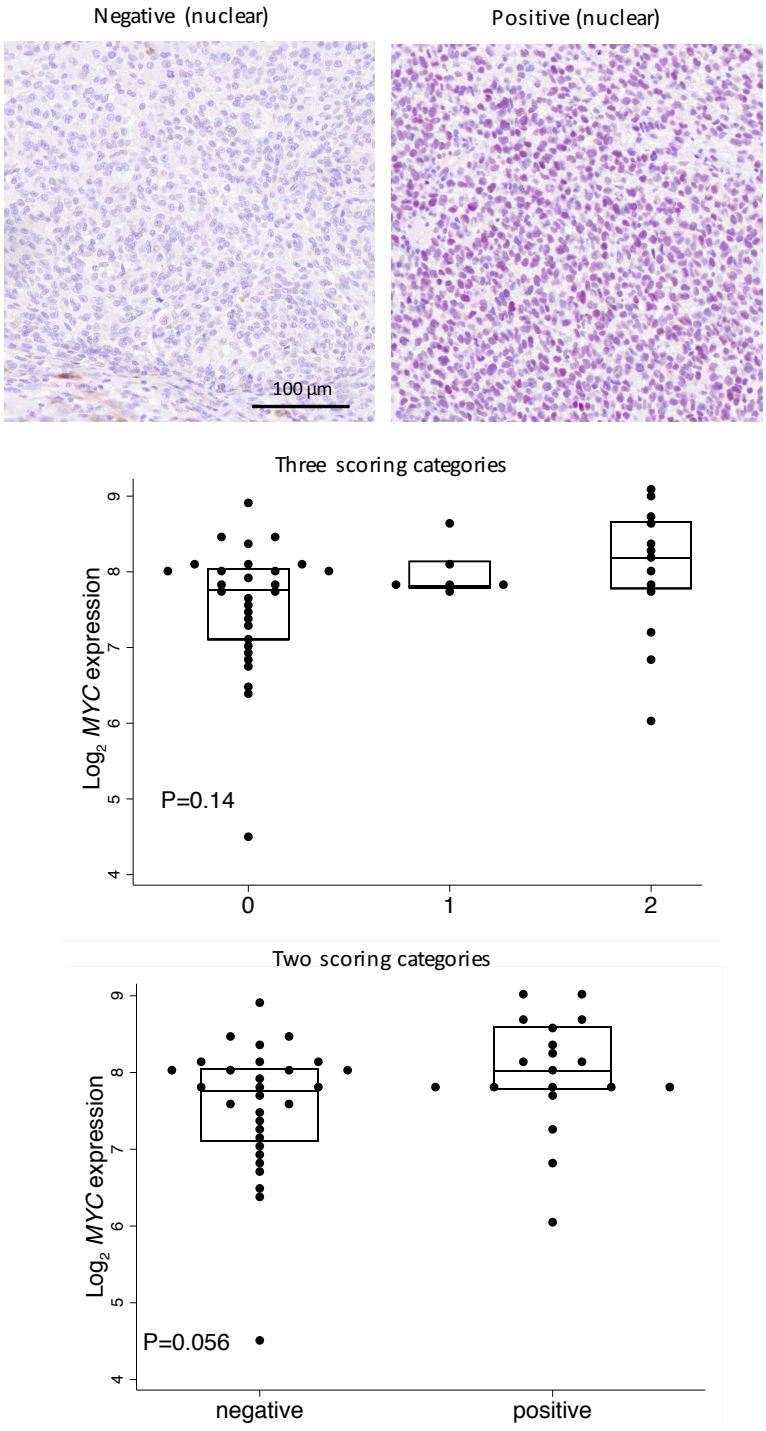


Figure 5.6: Example photograph of negative (top left) and positive (top right) nuclear staining of MYC (magnification 20x) in tumour cells

The box and dot plots represent mRNA expression of *MYC* across the three protein scoring categories (top) and two scoring categories (bottom).

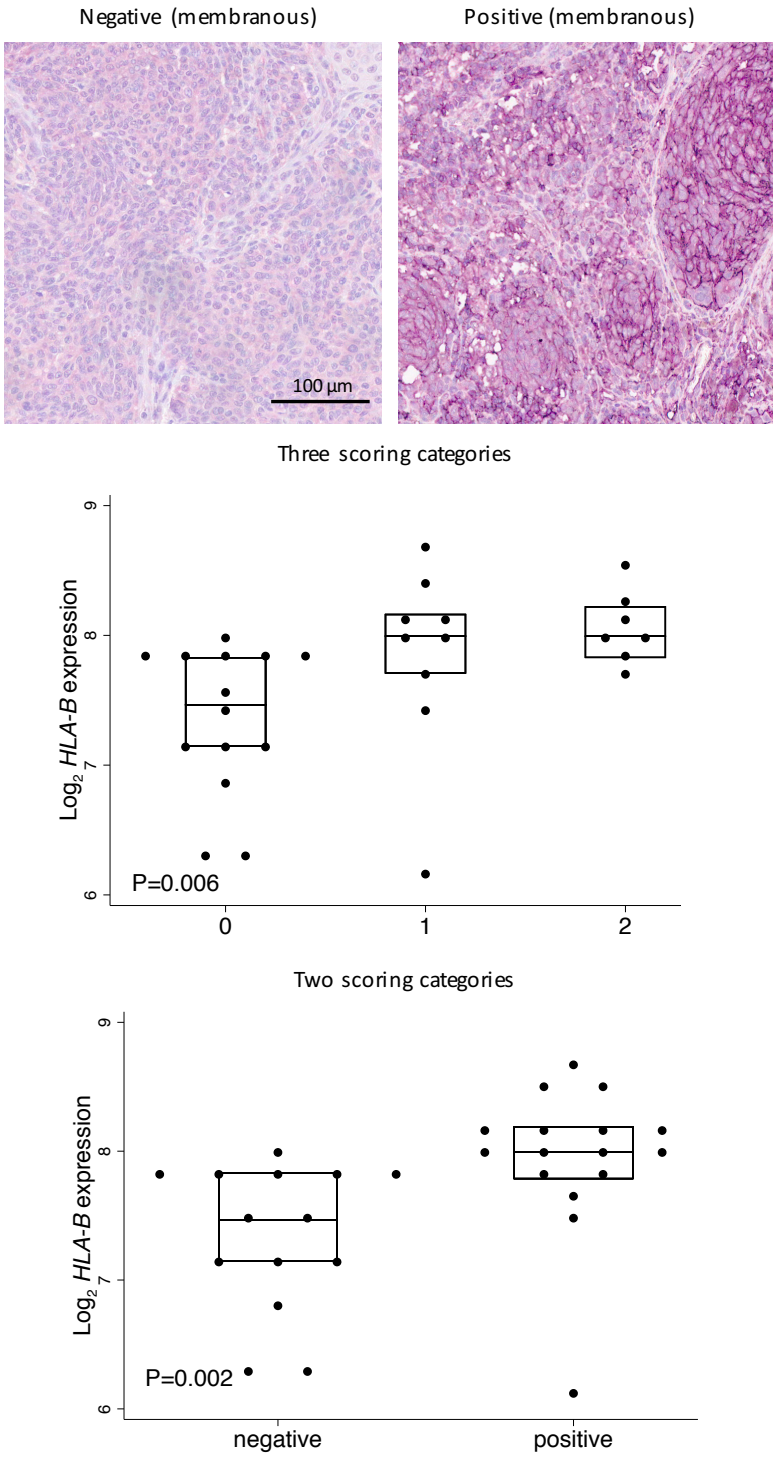


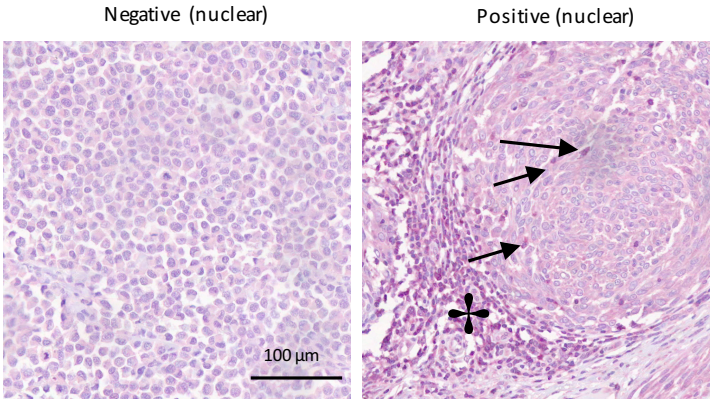
Figure 5.7: Example photograph of negative (top left) and positive (top right) membranous staining of HLA-B (magnification 20x) of tumour cells

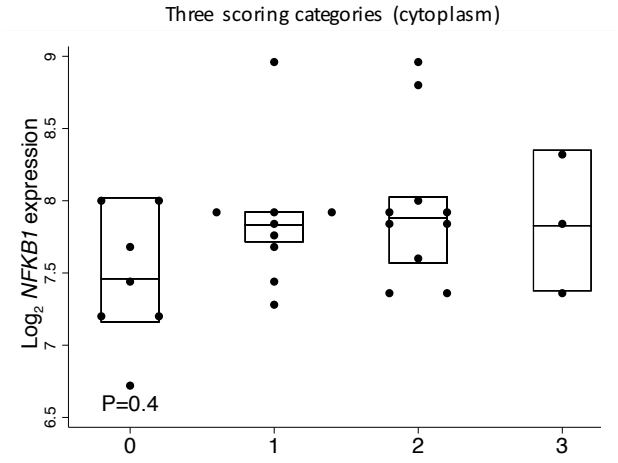
The box and dot plots represent gene expression of *HLA-B* across the three scoring categories (top) and two scoring categories (bottom).

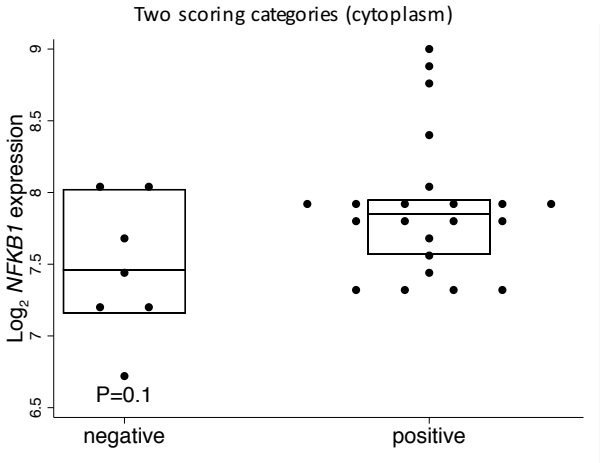
The staining for NF-κB p105 in 29 tumours was observed in the tumour cell nuclei and in the cytoplasm, although the cytoplasmic staining was always seen as a pink blush, and it was problematic to assess whether it was target specific. *NKFB1* gene expression did not vary significantly across the cytoplasmic staining scores of NF-κB p105 (P=0.4, P=0.1, Figure 5.8). The staining of the nuclei was clearer and more distinct, but the positive cells were so rare that looking only around the core would result in too many negative scores. Therefore, I counted the positive nuclei across the whole tumour (Figure 5.8). The *NFKB1* expression was significantly higher in positively stained tumours (more than 5 positive nuclei) than in negatively stained (less than 5 nuclei stained) (P=0.04). The nuclear staining of NF-κB p105 in the tumour strongly correlated with the nuclear staining of the TILs, $P=3x10^{-5}$ (Table 5.11).

Figure 5.8 (following page): Example photograph of negative (top left) and positive (top right) nuclear staining of NF-κB p105 (magnification 20x) in both tumour cells and TILs

The star on the right picture shows the positive staining of the TILs. The box and dot plots represent gene expression of *NFKB1* across the three scoring categories (top) and two scoring categories (middle) of the membranous staining. The bottom graph shows the gene expression of *NFKB1* between two categories of nuclear staining.


Negative (nuclear)
Positive (nuclear)
100 µm


Three scoring categories (cytoplasm)
Log2 NFKB1 expression
6.5
7
7.5
8
8.5
9
P=0.4
0
1
2
3


Two scoring categories (cytoplasm)
Log2 NFKB1 expression
6.5
7
7.5
8
8.5
9
P=0.1
negative
positive

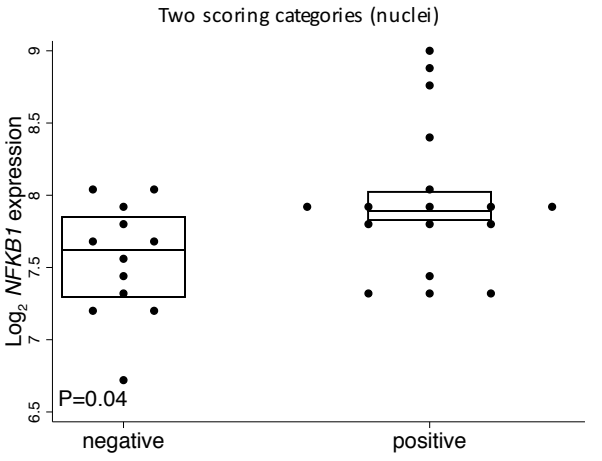

Two scoring categories (nuclei)
Log2 NFKB1 expression
6.5
7
7.5
8
8.5
9
P=0.04
negative
positive

Table 5.11: Cross tabulation of samples with nuclear postive and negative staining of NF-κB p105 of the tumour and TILs. Fisher's exact P=3x10^{-5}

TILs nuclear staining	*Tumour cell nuclear staining*		
	Score 0	*Score 1*	*Total*
Score 0	**10**	**1**	11
Score 1	**2**	**16**	18
Total	**12**	**17**	29

The IHC scoring at the continuous scale using Nuance software could be analysed using 3 types of metrics: a) percentage of haematoxylin staining (non-antibody specific nuclear staining), b) percentage of chromagen staining (overall antibody signal), and c) co-localised staining of haematoxylin and chromagen (antibody specific nuclear staining). Figure 5.9 below shows examples of such signals.

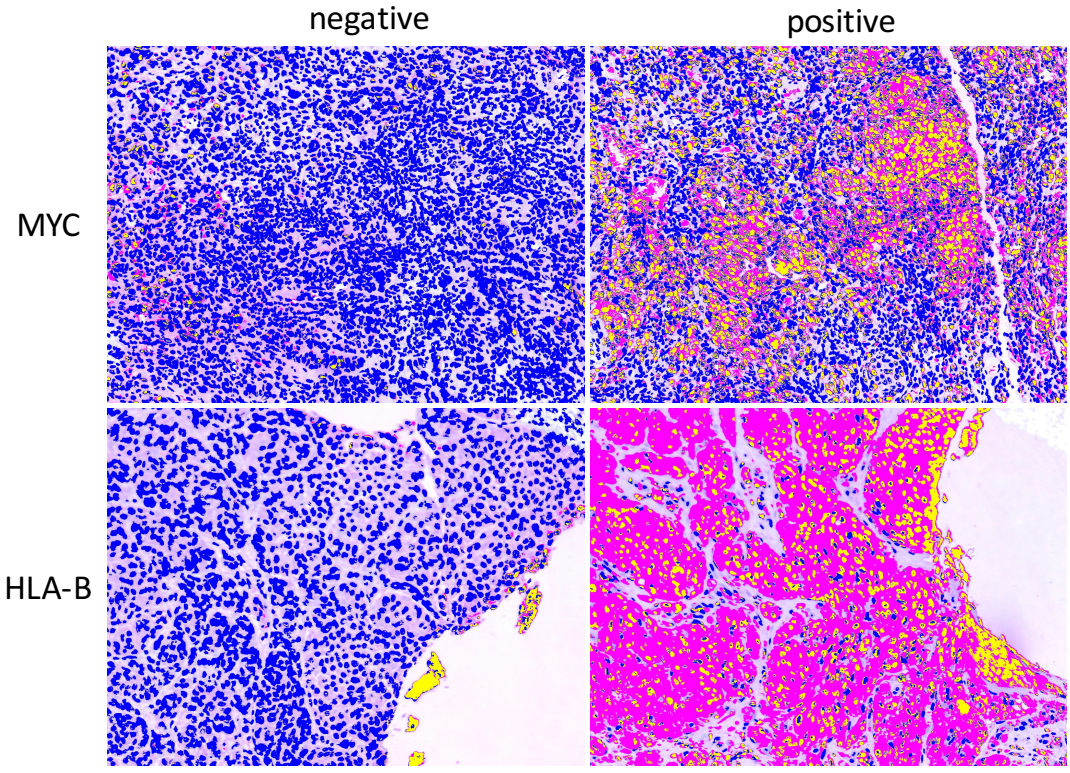


Figure 5.9: Representative images of negative and positive co-localised signal for MYC (top panel) and negative and positive total chromagen signal for HLA-B (bottom panel)

Blue colour indicates haematoxylin, pink is chromagen and yellow show co-localised signal of haematoxylin and chromagen.

The correlation analyses showed that haematoxylin staining for MYC and HLA-B were positively correlated (R=0.6, P=0.002), which indicated consistency in staining (Figure 5.10).

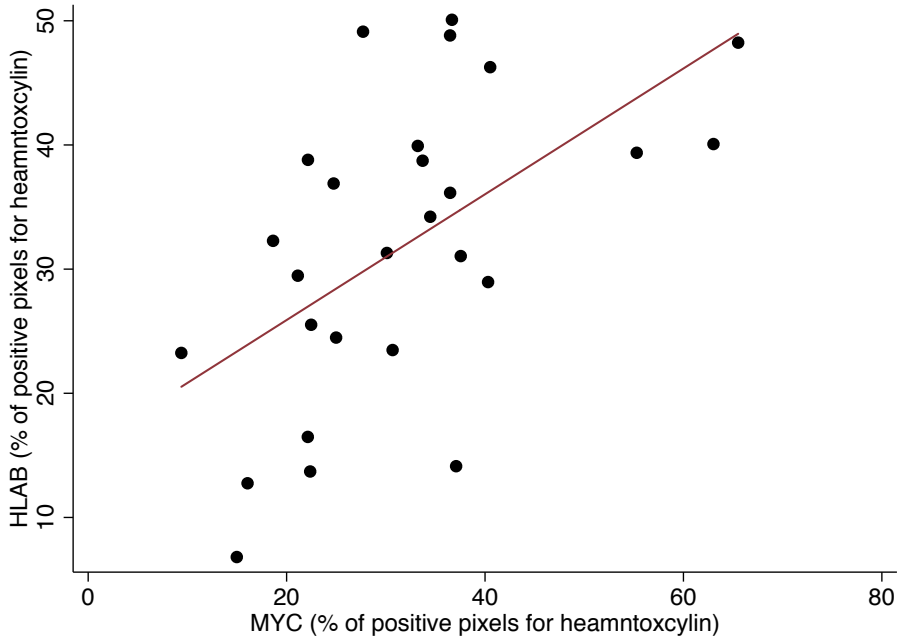


Figure 5.10: Scatterplot representing the HLA-B scoring (percentage of positive pixels for haematoxylin) on the y-axis and MYC (percentage of positive pixels for haematoxylin) on the x-axis

The correlation of HLA-B (chromagen) and MYC (haematoxylin + chromagen) on the continuous scale was negative (Spearman's rank correlation: R=-0.6, P=0.02). However, this analysis was only the case where more than 1% of pixels were detected for MYC (Figure 5.11). For less than 1%, there does not seem to be an association.

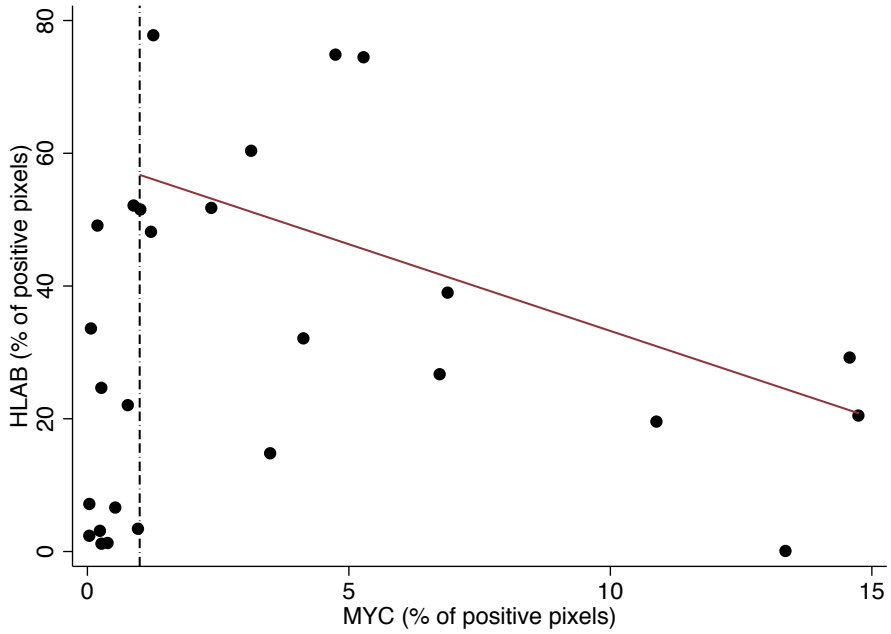


Figure 5.11: Scatterplot representing the HLA-B scoring (percentage of positive pixels for chromagen) on the y-axis and MYC (percentage of positive pixels for both haematoxylin and chromagen) on the x-axis

The vertical dashed line indicates MYC detection at less than 1%, which was ignored in deriving the best fit (red line).

5.5 Summary

In this chapter I showed that:

- Tumours classified in the Low Immune Subgroup had more CNAs, many of which were deletions of genes from NF-κB and IFN-γ signalling.
- In the subset of genes tested, tumour gene expression (mRNA), as determined from the transcriptomes correlated with DNA CNAs: lower expression was associated with genomic deletion, whilst increased expression was associated with amplifications.
- Majority of observed CNAs predicted poor survival such as amplification of MYC and deletions of NF-κB and IFN-γ signalling genes.
- The tumour gene expression levels correlated with protein levels assessed by IHC.
- MYC and HLA-B were negatively correlated on protein level as they were at mRNA level.

5.6 Discussion

The aim of this chapter was to evaluate the observations made on the gene expression level using CNA data and protein scores obtained by using IHC.

5.6.1 CNA

The main question stated regarding the CNA data was whether the observation made on gene expression level in relation to immune responses can be also seen on the DNA level. Indeed, it was observed that there was a strong agreement for the results obtained from the transcriptomic and CNAs data.

From the technical perspective, it is important to note that the CNA data was obtained from the DNA extracted from tumours that were thicker than average due to insufficient material for extraction of RNA and DNA from smaller tissue. Moreover, the approach I took to categorise the CNA data could be considered as arbitrary, because the categorisation of the continuous CNA data to define amplifications and deletions of a given gene can be challenging and there is no set threshold. There are existing methods to estimate the relative copy number within tumour data, such as GISTIC [214] or ABSOLUTE [215] algorithms. However, at the time I was analysing the CNA data none of them had been yet applied to these data.

However, none of these limitations suggest a bias towards the observed results and the concordance between transcriptomic expression and CNA indicates a good robustness of the approaches used.

I observed the variation of CNAs in key genes across the three immune subgroups. *MYC* was identified to be overexpressed in the Low Immune Subgroup on the gene expression level and the CNA data showed that the amplifications of *MYC* were more frequently present in that subgroup. Likewise, genes involved in NF-κB and IFN-γ signalling were downregulated in the Low Immune Subgroup and I hypothesised that it might be due to deletions of these genes, which was confirmed by CNA data: many of these genes indeed were deleted much more frequently in the Low than in the High Immune Subgroup. As expected, the majority of observed DNA structural aberrations strongly predicted poor prognosis, and this observation confirm the importance of these results. The gene expression level was significantly correlated with CNAs for the majority of tested genes. One explanation for the few instances where there was no

correlation (for example *CTNNB1*) could be that gene expression is controlled by many other factors (such as promoter methylations, which was confirmed in TCGA data by Nsengimana et al. [76]) among which structural variation has only a minor role. Moreover, the gene expression might be suppressed or enhanced depending on the biological states of the cells irrespective of the DNA copy number.

The landscape of CNA was already revealed in melanoma and the most frequently altered genes were identified [46][50][48]. For example MYC is known to be commonly amplified in melanoma [50] and as well as in other cancers [216]. Some of these studies were investigating the overall landscape of CNAs across the all melanoma sample types analysed together [50][48], and other was focused on categorised samples by driver mutations such as *BRAF*, *NRAS* and *NF1* [46]. None of these studies however directed their questions towards effectiveness of the immune responses affected by these CNAs.

The results from this chapter indicate that tumours which manifest low immune cell infiltration might have an immune evasion process induced by various mechanisms, including deletions of key immune genes (*NFKB2*, *CHUK*, *IRAK2*, *MYD88*, *MAP3K7*, *JAK2*, *STAT1*), amplifications of oncogenes (*MYC*), and up-regulation of β-catenin signalling. Moreover, the fact that these mechanisms were not mutually exclusive suggested that some of the tumours could be heterogenic, i.e. different tumour clones acquired diverse mechanism of immune evasion.

These are impactful observations, because if indeed the aberrations induce impairment of immune responses to melanoma, they should be further investigated regarding immunotherapy resistance. For example mutations in *JAK2* gene has already been reported to be involved in acquired and primary resistance to anti PD-1 therapy [217][218]. In summary, these results might suggest that a significant proportion of melanoma tumours in the Low Immune Subgroup may have an intrinsic resistance to immunotherapies.

5.6.2 IHC

The results from the transcriptomic data were also evaluated using the immunohistochemical staining. This approach enables to assess the protein level within the cell as well as its cellular localisation. It is a standard method for protein

level assessment in FFPE samples, although with some limitations. Firstly, scoring of staining of melanoma tumours is challenging due to the strong pigmentation (melanin) within some of the tumours, which occasionally cannot be distinguished from the actual staining of the protein of interest. To mitigate this problem, we used purple instead of brown chromagen, nevertheless the problem still occurred for very pigmented tumours. For such tumours it was difficult to differentiate positive purple staining among very dark pigment, hence were excluded from the analyses. Secondly, staining can sometimes be irregular putatively as a reflection of tumour heterogeneity and this becomes a problem when a region surrounding one core is positive and another core negative. In this instance the sample was dropped, because it was not known which core was used for nucleic acid extraction. Moreover, assessing the staining for multiple proteins performed on separate slides for each protein is time consuming. There are methods available for simultaneous multi-antigen detection, but, usually they require usage of fluorochrome rather than chromagen [105][106] and are much more expensive than single-antigen detection. Finally, due to ethical issues tumour blocks were available only for samples from deceased patients, which restricted the statistical power and also may have potentially biased the results.

Despite these caveats, I was able to demonstrate a positive association of IHC scores for the three proteins (MYC, HLA-B, and NF-κB p105) with measured gene expression in the tumours. However, in case if the gene expression did not correlate with IHC staining, the explanation could be that the protein of interest was posttranslationally modified, which could affect the epitope structure and the interaction with the antibody.

Importantly staining was seen in the cellular components that were expected, for MYC and NF-κB p105 in the nucleus and HLA-B on the cellular membrane. There was a similar negative correlation between HLA-B and MYC on the protein level as on gene expression level in both the LMC dataset and patient-derived melanoma cell lines. However, the MYC-HLA-B correlation at protein level was negative only when MYC was detectable in the slides in more than 1% (Figure 6). Under this threshold, which is very low and could reflect the absence of MYC protein, HLA-B protein took a wide range of values, suggesting the existence of other regulators of HLA-B in the absence of MYC. One example of a cause of low cell surface expression of HLA-B could be

posttranslational modifications of this protein as reported by Dellgren et al. [219]. Nevertheless, this result was remarkable given the challenge of MYC and HLA-B protein scoring on different slides from the same FFPE tumour block, which was very challenging. Ideally the experiment should be designed to performed dual (multi-antigen) staining for both MYC and HLA-B with different colours of chromagen per antigen.

Furthermore, IHC staining analysis showed that tumour nuclear localization of NF-κB p105 significantly positively correlated with TILs and with the gene expression of *NFKB1*, suggesting a reciprocal interaction of NF-κB signalling between tumour and immune cells. This fact and the observation that the deletions of the genes involved in the NF-κB signalling pathway were more frequent in the Low Immune Subgroup and were associated poor survival confirmed that this signalling pathway indeed might play a prominent role in tumour immune surveillance and progression, as discussed in Chapter 4.

Chapter 6

Survival analysis of prognosis predictors in melanoma in relation to immune response

6.1 Aims

In Chapter 3, I reported the identification of three immune subgroups of primary melanomas. In this chapter I describe an investigation of the determinants of survival within the immune subgroups. The aims were:

- To analyse the association between reported (smoking) or measured environmental factors (vitamin D levels) and melanoma specific survival (analysis adjusted for known prognosis predictors) in the whole dataset and in each of the three immune subgroups.

To perform survival analyses taking under consideration total mutational load:

- To test the association between a measure of mutational load and melanoma specific survival (analysis adjusted for known prognosis predictors).
- To test this association within each of the immune subgroups.

6.2 Background

In Chapter 1, the known predictors of prognosis, such as AJCC stage, sex, age, mitotic rate, and site of melanoma were already described. As recently reviewed by Chen and Mellman et al. [87], effective cancer cell killing by immune cells is a multistep process which can be rendered ineffective at multiple points along the way. Chen and Mellman et al. described the potential effects of variation in host genetics, differences in the microbiome, environmental exposures, therapeutic agents and cancer cells themselves leading e.g. to inability to present tumour antigen/neoantigens, or suppression of immune responses by checkpoint molecule expression. I have already described evidence (Chapter 4 and 5) that tumour variation e.g. deletion of genes coding for the NF-κB signaling pathway or expression of oncogenic *MYC* might play a role in suppressing the immune responses in melanoma.

In this chapter I present an analysis of two environmental exposures, smoking and vitamin D levels for which the Leeds Melanoma Research Group has already identified a role in melanoma specific survival. Moreover, I present the initial analysis of LMC mutational load data (with the assumption that it might represent the neoantigen load). Overall, I report here an investigation of possible interaction effects on survival from these variables and immune responses.

6.2.1 Environmental factors

The Leeds Melanoma Research Group has previously explored the biological significance of microscopic ulceration and identified some evidence for a role for environmental exposures in melanoma progression. The group first performed an immunohistochemical study of ulcerated tumours and reported them to have higher vascularity and more macrophages [220]. Subsequent transcriptomic studies identified a “chronic wound healing” - chronic inflammatory gene expression phenotype [120]. The group then argued that if there was chronic inflammation in tumours, then ulceration might be more frequent in people with systemic inflammation [221]. Ulceration was shown to be more common in the obese patients, the vitamin D deficient, smokers and diabetics in univariable analysis but only smoking and vitamin D deficiency were associated independently with ulceration and poorer melanoma specific survival [221].

I chose then to look at smoking and vitamin D levels as predictive of survival in the three immune subgroups. Since smoking and vitamin D levels (Chapter 3) did not vary across the three immune subgroups, I hypothesised that they might interact with these groups in terms of survival.

The Leeds Melanoma Research Group has had a long-standing interest in the role of vitamin D in melanoma progression since participants in the Leeds Melanoma Cohort displayed an inverse relationship between vitamin D levels and tumour thickness [114]. Melanoma specific survival was also superior in those with higher levels of vitamin D independently of tumour thickness [221]. Four subsequent studies from three continents also reported an inverse relationship between vitamin D levels at diagnosis and Breslow thickness [222][223][224][225]. Vitamin D has been shown to inhibit cancer cell proliferation *in vitro* [226], to be negatively correlated with blood levels of

C Reactive Protein (an inflammatory marker) [227] and that low levels are associated with autoimmunity and increased risk of infection [89]. These observations led to the hypothesis that vitamin D may be antiproliferative for melanoma cells but that it may also have a beneficial effect via reduction of systemic inflammation [228].

In my thesis I have used vitamin D levels in the survival analysis: I have not attempted to learn anything more about the role of vitamin D in melanoma progression. An in depth biological investigation of this environmental factor was a PhD project (simultaneous to my project) of Sathya Muralidhar, Marie Skłodowska-Curie PhD student.

Cigarette smoking is a globally detrimental environmental factor and it is known to be the preventable cause of deaths from many different cancer types [229] and this factor, as well as deficiency of vitamin D was of interest to the Leeds Melanoma Research Group. The contribution of smoking and risk of melanoma development is uncertain and some studies even proposed a controversial protective effect of tobacco smoking on melanoma development [230][229][231]. As mentioned above, the Leeds Melanoma Research Group has reported that cigarette smoking was associated with poor melanoma specific survival and with ulceration in the LMC [221]. A later (2017) study performed by an independent research group confirmed that smoking was predicting poor outcome and additionally revealed that that it was associated with a greater likelihood of melanoma metastasis to sentinel lymph nodes [232]. This study and one more additionally reported that higher tumour thickness was associated with smoking [232][233].

These evidences indicate that a biological interaction between smoking and melanoma progression might occur, however the exact mechanisms behind this have not yet been elucidated.

Some studies reported that cigarette smoking affects immune responses systemically: that it might induce systemic inflammation or cause immunosuppression [88]. A study analysing systemic inflammatory markers from participants suffering from prostate, lung, colorectal, and ovarian cancer, demonstrated that smokers had higher levels of some of these inflammatory markers than non-smokers [234]. Smoking associated tissue inflammation was mostly reported for lung cancer [235], as this tissue is directly

exposed to inhaled chemicals from the cigarettes. In other types of cancer this phenomenon was not widely studied, yet smoking was described as a risk factor for various cancer types [88].

The Bauer laboratory compared measures of the circulating immune cells between smokers and non-smokers and were able to identify associated methylation of some genes, which were occasionally followed by changes in their expression [236][237]. One of these genes was *GRP15*, which is known to be hypo-methylated in CD3+ T cells in smokers and this alteration in CpG sites was assumed to mediate enhanced gene expression of *GPR15*. However the causative effect of smoking on methylation of *GPR15* was not evidenced [237]. *GPR15* is non-the-less considered to be a biomarker of smoking being overexpressed in circulating immune cells in smokers in comparison to non-smokers [236][237] but its full biological function and significance is not yet understood. It was however proposed that the high expression of *GPR15* might be involved in inflammatory states in the intestine and skin epithelial cells in smokers [238].

In this chapter I report an investigation of the relationship between immune status, smoking and *GPR15* expression in the LMC transcriptomic data set.

6.2.2 Mutational load

Adaptive immunity requires that tumour cells express antigenic peptides which might serve as neoantigens or self-antigens to which T cell tolerance is incomplete e.g. as a result of restricted tissue expression [239]. Neoantigens are peptides generated by the degradation of abnormal proteins, which are detected by the cell as intruders potentially harmful for the cell functions. Tumour antigens/neoantigens are presented via MHC class I to the immune cells (as described in Chapter 1). In order to trigger effective immune responses against melanoma the neoantigens must be broken down into smaller molecules that can be successfully presented to the effector immune cells present within the tumour microenvironment [87][239].

In principle it is expected that melanoma tumours having a high mutational load (and hence putatively more neoantigens) will have a more favourable survival, due to the greater likelihood of activation of adaptive anti-tumour immune responses. It is of note however, that the survival effect might be moderated by other immune factors that

play an important role in anti-tumour immunity and do not require antigen presentation, such as innate immune responses. For example, if a tumour of a patient does not produce/present neoantigens but generates signals that activate e.g. NK cells, which effectively kill the tumour cells, the patient might still manifest relatively good survival.

For primary melanoma, there has not yet been any reported association between mutational load and survival. However, many studies concentrating on immunotherapy treatment responses in metastatic melanoma have shown survival benefit for patients having higher mutational load [145][146][147] (as described in Chapter 3). As I defined in Chapter 3, I found no significant difference in mutational load across the three immune subgroups. Although, as discussed earlier, since the data were generated for candidate cancer genes rather than on the whole genome, these mutation data might not represent the true neoantigen load that triggers immune responses. Nevertheless, these data were used to test whether an association with survival in relation to the three immune subgroups could be observed.

6.3 Methods

6.3.3 Survival analysis

The survival analyses were firstly performed for environmental together with clinico-histopathological factors. The mutational load variable was afterwards analysed adjusting the factors significantly predicting prognosis in the first analysis. This order was reasoned by the fact that the data for mutational load was available only for 319 of the 703 tumours, which would significantly reduce the power of the overall model if all variables were tested together.

6.3.3.1 *Environmental factors*

Firstly, using a univariable Cox proportional hazard model, the association between the clinical and environmental variables (AJCC staging, age at diagnosis, sex (females vs males), site of melanoma (limbs vs the rest), smoking (ever vs never) and vitamin D levels at recruitment (season-adjusted) with MSS were tested in the whole 703 LMC dataset. I used the variable “ever/never” smokers as this gave the most power for the

whole survival model but repeated all analyses using 2 other variables describing smoking habits: duration of smoking and packs of cigarettes smoked per year.

Subsequently, the significant clinical and environmental predictors from the univariable model were tested in a multivariable model adjusting the immune subgroups.

Finally, a multivariable Cox proportional hazard analysis was conducted within each immune subgroup using the predictors that showed the strongest degree of independence from the previous analysis (Figure 6.1).

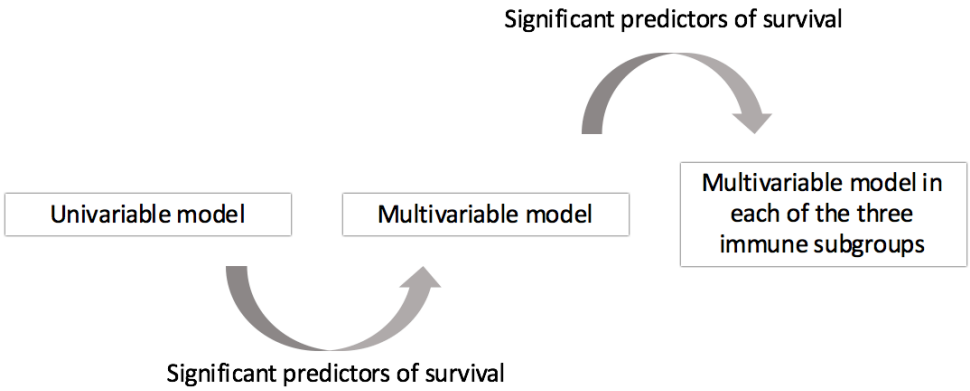


Figure 6.1: Schematic work-flow of the survival analyses

6.3.3.2 *Mutational load*

The derivation of the mutation count variable was explained in Chapter 2. Briefly, it represented the mutation count per megabase for a panel of 555 genes, which was split into tertiles (high, low, intermediate) (Figure 6.2).

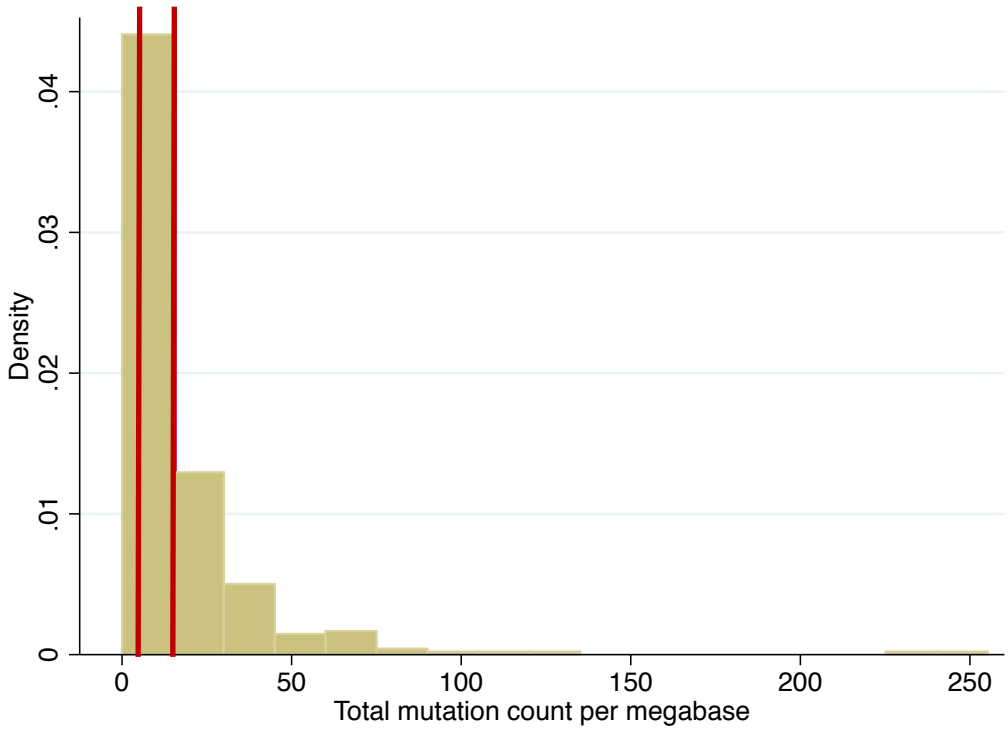


Figure 6.2: Histogram representing the density of samples for the total mutation count per megabase for 555 gene panel

The red lines show the points where the data were divided into three groups (high, low, intermediate).

Association between the three groups of total mutational load in the whole data set (N=521) and survival was tested using a univariable Cox proportional hazard model. Subsequently a multivariable Cox proportional hazard model for mutational load was applied adjusting other prognostic factors. Finally taking the predictors with the strongest degree of independence from the multivariable analysis in the whole dataset I tested their association with survival within each of the three immune subgroups using a multivariable model.

Taken together, the variables used for the survival analyses were:

- Age at diagnosis (continuous, years)
- Site of melanoma (%) - Rest (Head, Trunk, Rare) vs limbs, (see details in Chapter 2)
- Sex (females vs males)
- Mitotic rate (continuous, count/mm2)

- AJCC stage (%) (I vs II vs III trend)
- Smoking (ever vs never)
- Duration of smoking (years)
- Reported packs of cigarettes smoked per year
- Season-adjusted serum vitamin D at recruitment (continuous, nmol/L),
- Categorised mutational load (high vs intermediate vs low).

All these variables were described in detail in Chapter 2.

6.3.4 Comparative analysis of the tumours by smoking status

In order to explore the biological role of smoking, I then took an agnostic approach to analysis of the transcriptomic data. I hypothesised that if smoking was mediating an effect on host tumour interaction, then associated biological pathways would be identified in differential gene analyses comparing smokers with non-smokers. I also considered a candidate gene approach: I asked if the expression of genes coding for cytokines and their receptors, and inferred immune cell subgroups were associated with reported smoking. I listed the genes for 133 cytokines and their receptors from the Human Genome Organisation (HUGO) database (https://www.genenames.org). These and the immune cell scores were compared between smokers and never smokers using the Kruskal Wallis test and those with Benjamin-Hochberg FDR<0.05 were considered significant. The analyses were repeated however with the restriction to the immune subgroup for which the smoking status predicted prognosis in the strongest manner.

I sought to replicate a published finding that *GPR15* expression is a marker of smoking in the blood [236][237] by testing the variation of this gene by smoking status in the tumours with the Kruskal Wallis test. To test whether *GPR15* was associated with inflammatory states in the LMC tumour samples, as suggested in the literature [238] I tested the expression correlation with the list of genes coding for 133 cytokines and their receptors from the HUGO database (https://www.genenames.org) as above in the High Immune Subgroup.

6.4 Results

6.4.1 Environmental factors

In a univariable Cox proportional hazard model, AJCC staging, smoking (ever/never), duration of smoking, average packs of cigarettes smoked per year, site of melanoma, age at diagnosis, sex and tumour mitotic rate were significantly predictive of MSS while vitamin D levels was not statistically significant (Table 6.1).

Table 6.1: Table representing results from univariable Cox proportional hazard model

AJCC stage was categorized as stage I, II or stage III. Mitotic rate is the count of mitoses per mm^2. Duration of smoking was measured in years. Sex was tested for males vs females. Site of melanoma was tested for rest vs limbs. Smoking was tested as ever vs never. HR means melanoma specific death hazards ratio.

Predictor	***HR***	***P***	***95% CI***	
AJCC stage	2.46	$1.2x10^{-16}$	1.99	3.05
Age at diagnosis	1.03	$2.4 x10^{-8}$	1.02	1.05
Mitotic rate	1.03	$2.2 x10^{-}7$	1.02	1.04
Sex (males)	1.5	0.008	1.1	1.9
Duration of smoking	1.02	$3.4 x10^{-6}$	1.01	1.03
Site (rest)	1.91	$1.5 x10^{-5}$	1.41	2.59
Packs per year	1.02	$6.5 x10^{-5}$	1.01	1.03
Smoking (ever)	1.63	0.001	1.21	2.20
Vitamin D	0.99	0.089	0.99	1.00

In the multivariable Cox model when testing the independence factors found significant in the univariable model, the AJCC, smoking (ever vs never), site of melanoma, age at diagnosis, and mitotic rate remained significant (Table 6.2). When the significant predictors of MSS were tested within each of the immune subgroups, using the multivariable model, different variables were found to be significant in each of the subgroups. Interestingly, smoking (ever vs never) remained the most striking predictor of prognosis in the High Immune Subgroup (Table 6.2). The hazard ratio was significantly different among the subgroups $P<0.03$ (Test for HR equality).

Table 6.2: Multivariable Cox proportional hazard model for the environmental and clinico-histopathological variables significantly predicting prognosis (including "smoking": never/ever) in the univariable model, in the whole dataset and each immune subgroup

AJCC stage was categorized as stage I, II or stage III. Site of melanoma was tested for rest vs limbs. Age at diagnosis was reported in years. Mitotic rate is the count of mitoses per mm^2. HR means melanoma specific death hazard ratio.

Predictor	*HR*	*SE*	*z*	*P*	*95% CI*	
OVERALL (N=666)						
AJCC stage	2.05	0.26	5.58	$<10^{-7}$	1.59	2.64
Smoking (ever)	1.45	0.25	2.14	0.03	1.03	2.04
Site (rest)	1.64	0.32	2.51	0.01	1.11	2.41
Age at diagnosis	1.03	0.01	4.26	$<10^{-4}$	1.02	1.05
Sex (males)	0.97	0.17	-0.14	0.89	0.69	1.38
Mitotic rate	1.02	0.01	2.67	0.01	1.00	1.03
Inter. cluster vs Low	0.76	0.14	-1.52	0.13	0.53	1.08
High cluster vs Low	0.63	0.16	-1.86	0.06	0.39	1.03
LOW IMMUNE (N=202)						
AJCC stage	2.01	0.37	3.82	$<10^{-3}$	1.40	2.87
Smoking (ever)	0.92	0.23	-0.34	0.73	0.56	1.51
Site (rest)	1.98	0.56	2.42	0.02	1.14	3.43
Age at diagnosis	1.03	0.01	3.1	0.002	1.01	1.06
Mitotic rate	1.01	0.01	1.11	0.27	0.99	1.03
INTERMEDIATE (N=207)						
AJCC stage	1.75	0.39	2.51	0.01	1.13	2.72
Smoking (ever)	1.78	0.51	1.99	0.05	1.01	3.13
Site (rest)	1.36	0.41	1.02	0.31	0.75	2.46
Age at diagnosis	1.03	0.01	2.32	0.02	1.00	1.06
Mitotic rate	1.04	0.01	3.55	$<10^{-3}$	1.02	1.06
HIGH IMMUNE (N=122)						
AJCC stage	3.99	1.47	3.74	$<10^{-3}$	1.93	8.23
Smoking (ever)	4.59	2.35	2.97	0.003	1.68	12.53
Site (rest)	2.52	1.31	1.78	0.08	0.91	6.99
Age at diagnosis	1.05	0.02	2.24	0.03	1.01	1.10
Mitotic rate	1.02	0.04	0.37	0.71	0.93	1.11

The Kaplan Meier plot in Figure 6.3 is a graphical representation of the smoking effect in the whole LMC cohort. Ever smoking was associated with a significantly worsened survival.

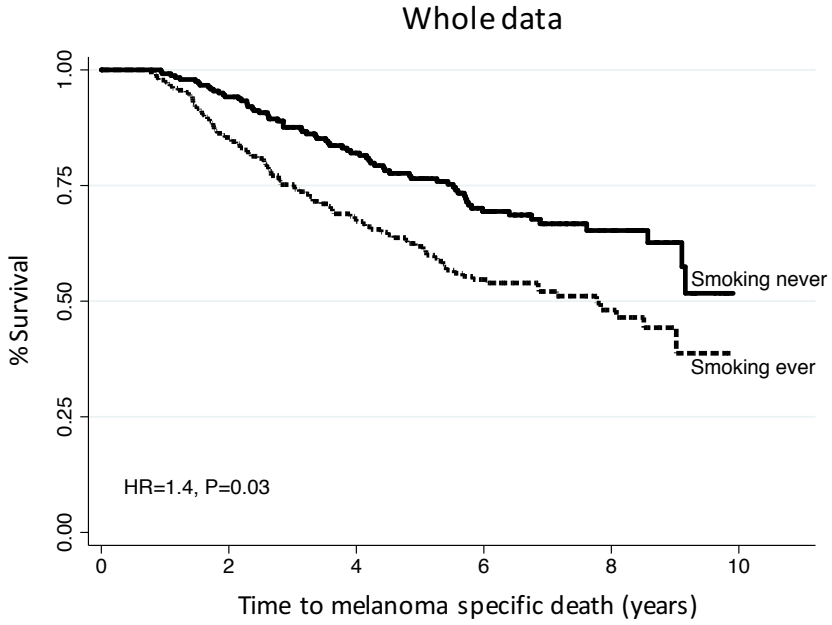


Figure 6.3: Kaplan Meier plot for ever vs never smoking in the whole dataset

P from multivariable analysis adjusting for prognostic factors.

In Figure 6.4 to Figure 6.6, Kaplan Meier plots show the different survival profiles associated with smoking in Low, Intermediate and High Immune Subgroups respectively. It can be seen that the effect of having ever smoked was the strongest in participants with tumours assigned to the High Immune Subgroup. There is a clear trend with the increasing strength of immune responses: no smoking effect (P=0.73) in low immune tumours, moderate effect (HR=1.8, P=0.05) in intermediate immune tumours and very strong effect (HR=4.6, P=0.003) in high immune tumours.

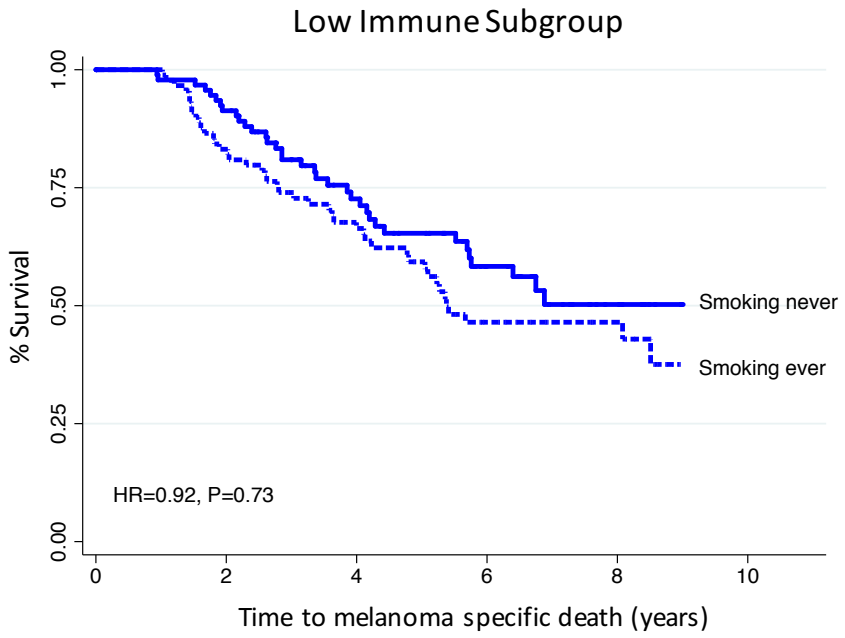


Figure 6.4: Kaplan Meier plot for ever vs never smoking in the Low Immune Subgroup

P from multivariable analysis adjusting for prognostic factors.

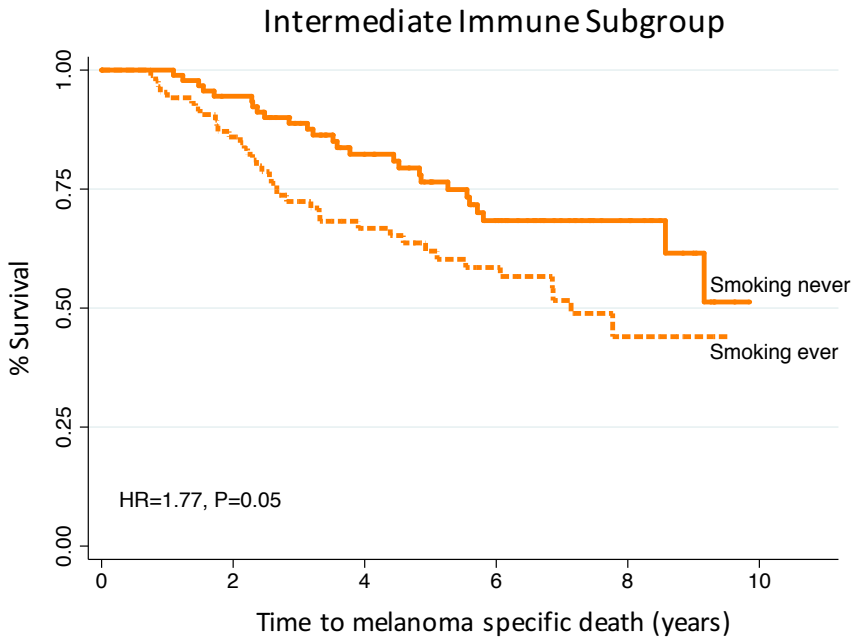


Figure 6.5: Kaplan Meier plot for ever vs never smoking in the Intermediate Immune Subgroup

P from multivariable analysis adjusting for prognostic factors.

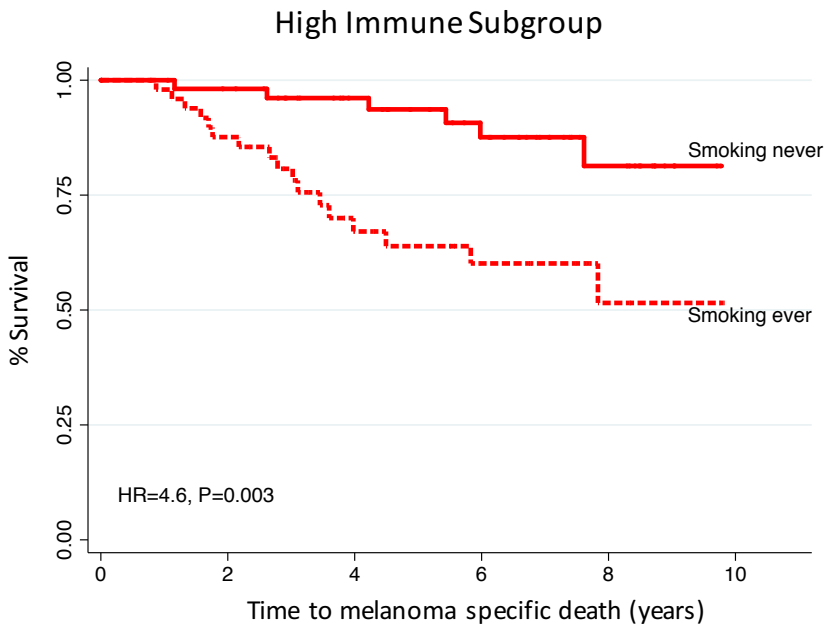


Figure 6.6: Kaplan Meier plot for ever vs never smoking in the High Immune Subgroup

P from multivariable analysis adjusting for prognostic factors.

I replicated the above observation using two more variables describing reported habits of smoking: duration of smoking and average packs smoked per year. All these variables were strong hazard factors for melanoma death in the High Immune Subgroup (HR=1.04, P=<0.001; HR=1.03, P=0.01, respectively) and less so in the Intermediate Immune Subgroup (HR=1.01, P=0.19; HR=1.02, P=0.09, respectively), with the weakest effect seen for average cigarette packs smoked per year in the Low Immune Subgroup (HR=0.01, P=0.23; HR=1.01, P=0.02, respectively). The difference in strengths of these HRs within one immune subgroup reflects the different units they correspond to: for example the effect of smoking for one year is not the same as smoking one pack per year.

AJCC stage and age at diagnosis strongly predicted MSS in each of the immune subgroups (Table 6.2, Table 6.3, Table 6.4). Site of melanoma was the strongest predictor of MSS in the Low Immune Subgroup (Table 6.2, Table 6.3, Table 6.4). Mitotic rate had the strongest effect in the Intermediate Immune Subgroup (Table 6.2, Table 6.3, Table 6.4).

Table 6.3: Multivariable Cox proportional hazard model for the environmental and clinico-histopathological variables significantly predicting prognosis (including the "duration of smoking" variable) in the univariable model, in the whole dataset and each immune subgroup

AJCC stage was categorized as stage I, II or stage III. Site of melanoma was tested for rest vs limbs. Age at diagnosis was reported in years. Mitotic rate is the count of mitoses per mm^2. HR means melanoma specific death hazard ratio.

Predictor	*HR*	*SE*	*z*	*P*	*95% CI*	
OVERALL (N=527)						
AJCC stage	2.03	0.26	5.49	$<10^{-3}$	1.58	2.61
Duration of smoking	1.02	0.01	2.9	0.004	1.00	1.03
Site (rest)	1.65	0.32	2.56	0.01	1.12	2.43
Age at diagnosis	1.03	0.01	4.04	$<10^{-3}$	1.02	1.05
Sex (males)	0.91	0.17	-0.49	0.62	0.64	1.30
Mitotic rate	1.02	0.01	2.88	0.004	1.01	1.03
Inter. cluster vs Low	0.81	0.15	-1.14	0.26	0.56	1.17
High cluster vs Low	0.65	0.16	-1.75	0.08	0.40	1.05
LOW IMMUNE (N=200)						
AJCC stage	1.95	0.36	3.64	$<10^{-3}$	1.36	2.79
Duration of smoking	1.01	0.01	1.2	0.23	0.99	1.02
Site (rest)	1.78	0.49	2.07	0.04	1.03	3.06
Age at diagnosis	1.03	0.01	2.79	0.005	1.01	1.05
Mitotic rate	1.01	0.01	1.13	0.26	0.99	1.03
INTERMEDIATE (N=207)						
AJCC stage	1.74	0.38	2.5	0.01	1.13	2.68
Duration of smoking	1.01	0.01	1.3	0.19	0.99	1.03
Site (rest)	1.43	0.43	1.19	0.23	0.79	2.60
Age at diagnosis	1.03	0.01	2.25	0.03	1.00	1.06
Mitotic rate	1.04	0.01	3.61	$<10^{-3}$	1.02	1.06
HIGH IMMUNE (N=120)						
AJCC stage	3.14	1.13	3.19	0.001	1.55	6.34
Duration of smoking	1.04	0.01	3.14	0.002	1.01	1.06
Site (rest)	1.97	1.04	1.29	0.20	0.70	5.53
Age at diagnosis	1.04	0.02	1.9	0.06	1.00	1.09
Mitotic rate	1.04	0.05	0.93	0.35	0.96	1.14

Table 6.4: Multivariable Cox proportional hazard model for the environmental and clinico-histopathological variables significantly predicting prognosis (including the "packs cigarettes per year" variable) in the univariable model, in the whole dataset and each immune subgroup

AJCC stage was categorized as stage I, II or stage III. Site of melanoma was tested for rest vs limbs. Age at diagnosis was reported in years. Mitotic rate is the count of mitoses per mm^2. HR means melanoma specific death hazards ratio.

Predictor	*HR*	*SE*	*z*	*P*	*95% CI*	
OVERALL (N=507)						
AJCC stage	2.13	0.28	5.75	$<10^{-3}$	1.65	2.76
Packs per year	1.02	0.00	3.48	0.001	1.01	1.03
Site (rest)	1.65	0.32	2.58	0.01	1.13	2.40
Age at diagnosis	1.03	0.01	3.24	0.001	1.01	1.04
Sex (males)	0.97	0.18	-0.14	0.89	0.68	1.40
Mitotic rate	1.01	0.01	2.11	0.04	1.00	1.03
Inter. cluster vs Low	0.79	0.15	-1.24	0.21	0.55	1.14
High cluster vs Low	0.57	0.15	-2.16	0.03	0.35	0.95
LOW IMMUNE (N=190)						
AJCC stage	1.98	0.36	3.75	$<10^{-3}$	1.38	2.82
Packs per year	1.01	0.01	2.38	0.02	1.00	1.03
Site (rest)	1.98	0.56	2.43	0.02	1.14	3.44
Age at diagnosis	1.02	0.01	2.13	0.03	1.00	1.05
Mitotic rate	1.00	0.01	0.34	0.73	0.98	1.02
INTERMEDIATE (N=202)						
AJCC stage	1.89	0.44	2.72	0.01	1.19	2.98
Packs per year	1.02	0.01	1.71	0.09	1.00	1.04
Site (rest)	1.44	0.42	1.23	0.22	0.81	2.57
Age at diagnosis	1.03	0.01	2.05	0.04	1.00	1.05
Mitotic rate	1.03	0.01	3.22	0.001	1.01	1.06
HIGH IMMUNE (N=116)						
AJCC stage	4.28	1.74	3.58	$<10^{-3}$	1.93	9.51
Packs per year	1.03	0.01	2.49	0.01	1.01	1.05
Site (rest)	1.50	0.76	0.8	0.43	0.55	4.07
Age at diagnosis	1.04	0.02	1.63	0.10	0.99	1.08
Mitotic rate	1.03	0.05	0.66	0.51	0.94	1.12

Smoking is known to be associated with deprivation [240]. The deprivation index in the LMC cohort was recorded as a Townsend score, based on the residence area (post code). Therefore, I took advantage of the information included in LMC and tested if the smoking effect (ever vs never) on survival was confounded by deprivation status.

The results showed that the hazard ratio and the P value for the detrimental effect of smoking on MSS was the same before and after adjustment for Townsend score: HR=1.6 P=0.001, indicating that the smoking effect was unlikely to be confounded by deprivation.

Taken together, reported smoking was shown to be associated with worse outcome independently to other known melanoma specific survival predictors and socio-economic status, with the strongest effect being observed in the High Immune Subgroup.

6.4.2 Comparative analysis of the tumours by smoking status

6.4.2.1 *Agnostic analysis of the transcriptomic data*

I tested whole transcriptome differences in the tumours excised from smokers compared with non-smokers. No gene was significantly differentially expressed by tumours removed from participants who smoked and those who did not, after multiple testing correction (Benjamini-Hochberg FDR<0.05) in the whole dataset (see the top 20 genes in Appendix A.4.1) and in the High Immune Subgroup (top 20 genes in Appendix A.4.2).

6.4.2.2 *Clinico-histopathological factors and immune cell scores*

Subsequently, in order to understand the biological differences in the tumours between smokers and never smokers I tested whether the clinico-histopathological characteristics differed between these two groups. In the whole dataset, smoking was associated with thicker tumours (Breslow Thickness) (P=0.008), and with a higher mitotic rate at a borderline significant level (P=0.09) (Table 6.5). However, smoking status did not differ with ulceration status as previously reported in LMC [221], but this is likely to reflect weaker statistical power compared for the previous study which was based on more than 2000 patients [221]. When the analysis was restricted to the High Immune Subgroup, only mitotic rate came close significance at P=0.08 (Table 6.6).

Table 6.5: Variation of clinico-histopathological features of the tumours between ever and never smokers in the whole dataset

Characteristic	***Ever smokers***	***Never Smokers***	***P-value (N)***
Number of participants (658)	334	324	-
Ulcerated (%)	35.1	35.2	0.2 (658)
Breslow thickness (median, mm)	2.4	2.1	0.008 (648)
Mitotic rate (median, count/mm^2)	4	3	0.09 (559)
TILs (%) (as determined by clinic dermatopathologists) *Brisk* *Non-brisk* *Unclassified* *No TILs*	 13.9 63.9 8.7 13.5	 15.8 60.1 9.8 14.3	0.9 (518)
TILs (%) – (single observer, S O'S) *Brisk* *Non-brisk* *No TILs*	 11.7 81.7 6.6	 9.9 81.2 8.9	0.5(567)

Table 6.6: Variation of clinico-histopathological features of the tumours between ever and never smokers in the High Immune Subgroup

Characteristic	***Ever smokers***	***Never Smokers***	***P-value (N)***
Number of participants (152)	74	78	
Ulcerated (%)	34	26	0.27 (152)
Breslow thickness (median, mm)	2.2	1.95	0.51 (149)
Mitotic rate (median, count/mm^2)	3.5	2	0.08 (129)
TILs (%) (clinic dermatopathologists) *Brisk* *Non-brisk* *Unclassified* *No TILs*	 23 65.6 9.8 1.6	 32.8 46.3 13.4 7.5	0.1 (128)
TILs (%) – (single observer, S O'S) *Brisk* *Non-brisk* *No TILs*	 20.6 76.5 2.9	 25.3 70.4 4.2	0.7(139)

As there was an interaction effect on survival between immune subgroups and smoking, I further asked if there were differences in immune cell scores by smoking. I tested the differential expression of 27 immune cell scores between ever and never smokers in the whole dataset and in the High Immune Subgroup. However, this analysis did not show any significant results even before adjusting for multiple testing (Table 6.7). The score for plasmacytoid dendritic cells was only close to being significant (higher in smokers) prior to multiple testing correction in the High Immune Subgroup (P=0.07).

Table 6.7: Differences in immune cell scores between ever and never smokers in the whole dataset and the High Immune Subgroup

Negative Z score indicates higher score in ever smokers.

	Whole data		High Immune Subgroup	
Cell Type	**Z score**	**P value**	**Z score**	**P value**
Activated B cells	-0.74	0.46	0.36	0.72
Central memory CD4	-0.78	0.44	-0.31	0.76
Central memory CD8	-0.51	0.61	-0.72	0.47
Cytotoxic cells	-0.94	0.35	1.52	0.13
DC	-0.04	0.97	-0.87	0.38
Effector memory CD8	-0.80	0.42	0.54	0.59
Eosinophil	-0.48	0.63	-0.13	0.89
iDC	-1.21	0.23	-0.28	0.78
Immature B cells	-1.28	0.20	-0.01	0.99
Macrophages	-0.61	0.54	0.52	0.6
Mast cells	-0.89	0.37	-1.04	0.3
MDSC	-1.28	0.20	-0.22	0.82
Memory B cells	-1.52	0.13	-0.87	0.39
Monocytes	-1.66	0.10	-0.62	0.53
Neutrophils	-0.09	0.93	0.61	0.54
NK	-0.94	0.35	0.52	0.61
NK56 bright	-0.46	0.65	1.6	0.11
NK56 dim	1.02	0.31	0.04	0.97
NKT	-0.01	1.00	0.9	0.37
pDC	-1.09	0.28	-1.8	0.07
T cells	-0.60	0.55	-0.37	0.71
TFH	-0.17	0.86	1.23	0.22
TGD	-1.01	0.31	0.63	0.53
Th1	-0.70	0.48	0.28	0.78
Th17	0.90	0.37	0.78	0.43
Th2	-0.87	0.38	-0.98	0.33
Treg	-0.60	0.55	0.09	0.93

6.4.2.3 ***GPR15***

Since *GPR15* is known as a marker of smoking in circulating immune cells [236][237], I tested whether its tumour expression was associated with smoking status in the LMC. Surprisingly, the data showed that *GPR15* expression was not significantly higher in the tumours of ever compared to never smokers in the whole dataset of 703 tumours (P=0.12), but *GPR15* expression was significantly higher in tumours of ever smokers compared to never smokers of the High Immune Subgroup, even though the statistical power was lower in this subset (P=0.02) (Table 6.8). Using still smoker vs non-smoker as an alternative definition of the smoking variable, *GPR15* expression was slightly higher in tumours of still smokers in both the whole dataset (P=0.01) and the High Immune Subgroup (P=0.002) (Table 6.9).

Table 6.8: Association of *GPR15 gene expression* between ever and never smoking in the whole dataset and the High Immune Subgroup

GPR15	***Ever (mean of log2 gene expression)***	***Never (mean of log2 gene expression)***	***Fold change***	***P***
Whole dataset	8.0	7.9	1.07	0.12
High Immune	8.5	8.1	1.32	0.02

Table 6.9: Association of *GPR15* gene expression between non and still smokers in the whole dataset and the High Immune Subgroup

GPR15	***Non-smokers (mean of log2 gene expression)***	***Still smokers (mean of log2 gene expression)***	***Fold change***	***P***
Whole dataset	7.9	8.3	1.32	0.01
High Immune	8.1	9.0	1.87	0.002

Moreover, when I tested the expression of *GRP15* across the three immune subgroups restricting the analysis only to “ever smokers” or “never smokers”, I observed that *GRP15* was significantly higher in the High Immune Subgroup only in ever smokers (Figure 6.7).

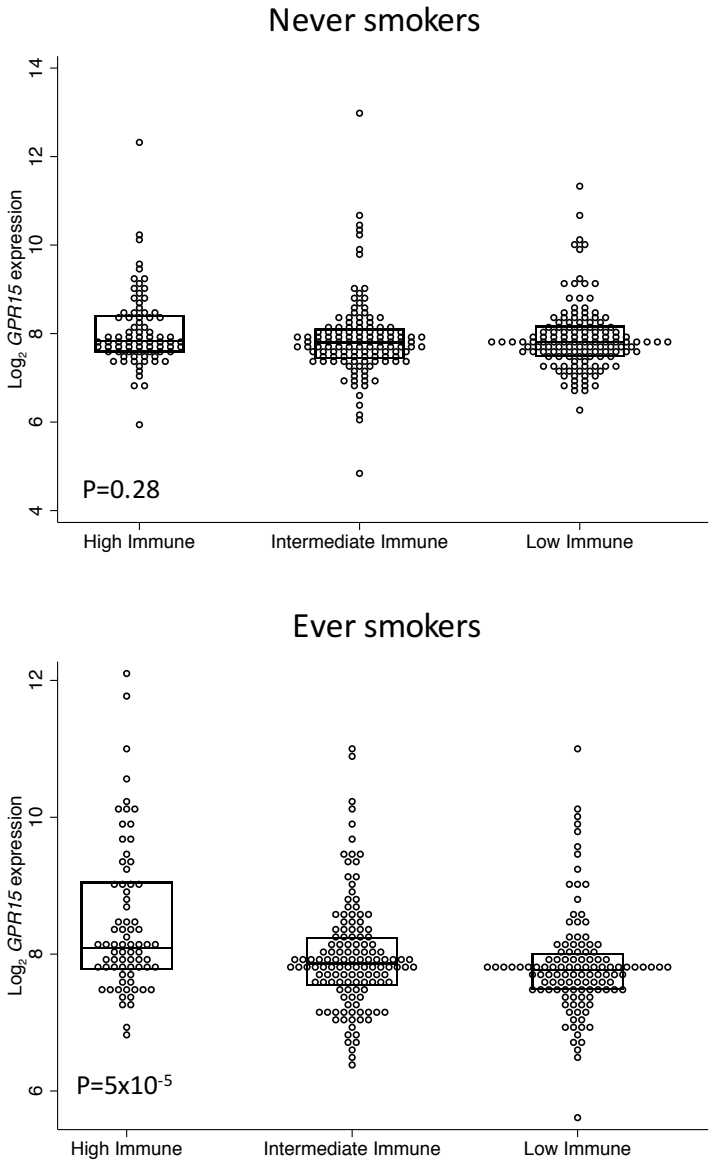


Figure 6.7: Expression (log2 scale) of *GPR15* across the three immune subgroups restricted to never (top) and ever smokers (bottom)

6.4.2.4 *Analysis of GPR15 expression in peripheral blood transcriptomes*

As our group have also generated gene expression data from peripheral blood from melanoma patients in an independent patient cohort, Dr John Davies (Senior statistician in the group) tested the whole transcriptome differences between non-smokers and still smokers. Importantly the most significantly differentially expressed gene after multiple correction was *GPR15* ($P=1.5x10^{-5}$, data not shown), which corroborated the existing literature, suggesting that *GPR15* is higher in peripheral blood in smokers than non-smokers.

6.4.2.5 *GPR15 and cytokines*

Following the literature suggestions that *GPR15* might be involved in inflammatory conditions and the above results that *GPR15* correlated with smoking status particularly in the High Immune Subgroup I further tested the correlation between *GPR15* tumour expression and the expression genes coding for cytokines and their receptors.

The top positively correlating gene with *GPR15* was Chemokine (C-X-C motif) ligand 3 - *CXCL3* (R=0.27, P=0.001). Among the other strong positively correlating genes (coding for cytokines and their receptors) with *GPR15* in the High Immune Subgroup were: Interleukin 6 - *IL6* (R=0.14, P=0.09), Interleukin 6 Receptor - *IL6R* (R=0.19, P=0.016), and Interleukin 6 Signal Transducer: - *IL6ST* (R=16, P=0.01) genes. It is of note that the correlations with *GPR15* expression values were weak overall and the results were not corrected for multiple testing. The 0.05 alpha threshold would be 0.0004 after multiple testing correction (Bonferroni, 133 tests) and none of the correlations reaches this level. However, it is of interest that 12 genes reached significance (unadjusted 0.05 level, Table 7) when no more than 7 were expected by chance.

Table 6.10: Cytokines and their receptors most strongly correlated with *GRP15* in the High Immune Subgroup, Spearman's rank correlation

Gene name	***P-value***	***R***
CXCL3	0.001	0.27
IL18R1	0.003	0.24
IL7R	0.01	0.22
CCL21	0.01	0.21
IL6ST	**0.01**	**0.20**
CCL19	0.02	0.19
IL15	0.02	0.19
IL6R	**0.02**	**0.19**
IL32	0.04	0.17
CXCL12	0.03	0.17
CXCR6	0.05	0.16
CCR4	0.05	0.16
IL11RA	0.07	0.15
IL6	**0.09**	**0.14**
CCR2	0.08	0.14
IL9	0.08	0.14
CXCR4	0.08	0.14

A similar analysis conducted in TCGA dataset (metastatic tumours) confirmed that the expression of *IL6ST* and *IL6R* were positively correlated with *GPR15* in the High Immune Subgroup (R=0.4, P=0.0003; R=0.3, P=0.01, respectively), while IL6 (R=-0.2, P=0.9) and *CXCL3* (R=-0.2, P=0.1) were not.

In the whole LMC dataset, none of these three genes was correlated with *GPR15*: *IL6* (R=0.04, P=0.3), *IL6R* (R=0.05, P=0.2), *IL6ST* (R=-0.03, P=0.5). In the whole TCGA dataset *IL6R* (R=0.13, P=0.02) and *IL6ST* (R=0.2, P=0.0004) were positively correlated with *GPR15*, but less so than in the High immune Subgroup. To add a layer of complexity, *IL6* (R=0.2, P=0.0004) was positively correlated with *GPR15* in the whole TCGA dataset, although it was not in the High Immune Subgroup.

Summarising, the correlation between *IL6R* and *IL6ST* with *GPR15* were consistent in the LMC and TCGA High Immune Subgroup.

6.4.2.6 ***GPR15 and ulceration***

To further explore the relationship between *GPR15* and inflammation, I tested whether the expression of *GPR15, CXCL3, IL6, IL6R, and IL6ST* varied with ulceration status since the Leeds Melanoma Group has reported that tumour ulceration was associated with transcriptomic evidence of wound healing inflammation in a subset of this cohort (200 tumours) [120]. While *GPR15*, *CXCL3*, *IL6R*, and *IL6ST* expression did not significantly vary between ulcerated and non-ulcerated tumours (P=0.5, 0.9, 0.8, 1.0, respectively), *IL6* itself was significantly more expressed in ulcerated than in non-ulcerated tumours ($P=1.2x10^{-5}$) (Figure 6.8), consistently with the earlier report [120].

Taken together these data suggest that *GPR15* might be associated with inflammatory *IL6* signalling in the High Immune Subgroup as the results from the LMC were consistent with TCGA. Moreover, *GPR15* correlated positively with *CXCL3*, however only in the LMC which could potentially be the result of biological differences between primary and metastatic tumours. Or possibly this observation was a result of chance.

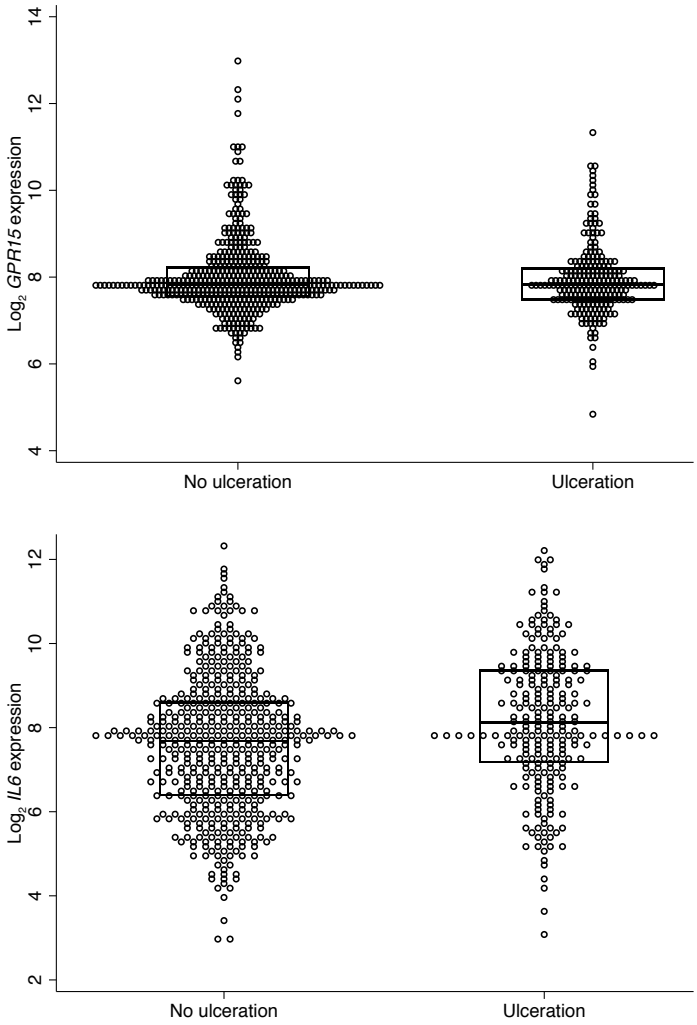
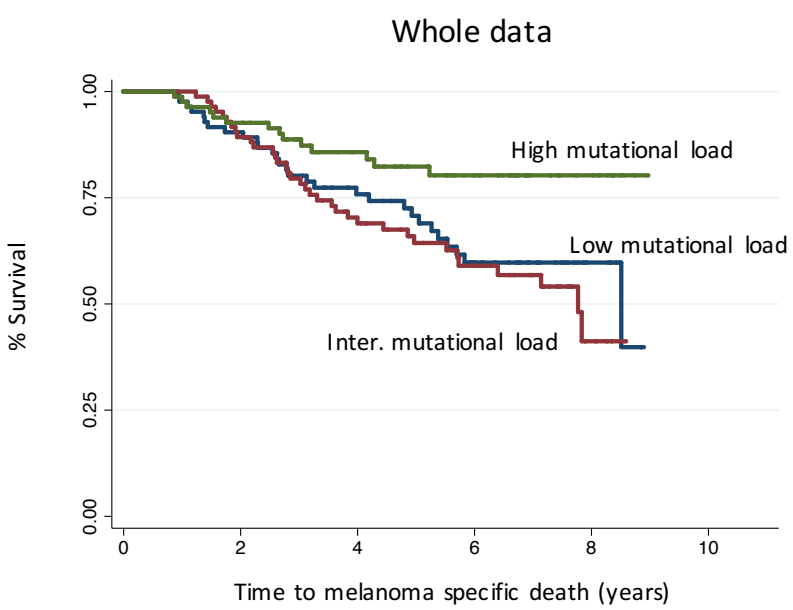


Figure 6.8: Expression of *GPR15* (top) and *IL6* (bottom) between non-ulcerated and ulcerated tumours

6.4.3 Mutational load

The survival analysis of the three immune subgroups by mutational load in the whole dataset showed that a high mutational load predicted the most favourable prognosis (HR=0.6, P=0.06) and no difference was seen between low and intermediate mutational loads (HR=1.3, P=0.3) (Figure 6.9). Similar results were observed in the Low Immune Subgroup (Figure 6.10), while in the Intermediate Subgroup none of the mutational load groups predicted prognosis (Figure 6.11. In the High Immune Subgroup, the results showed that high mutational load predicted even better prognosis compared to other subgroups (in fact all 21 patients who were classified in the High Immune Subgroup and had high mutational load survived), but intriguingly

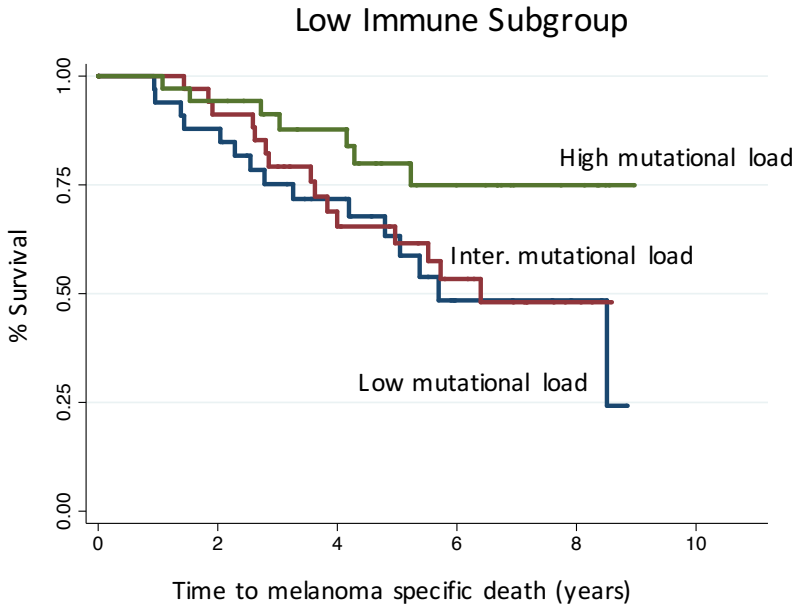
the low mutational load predicted better prognosis than the intermediate (Figure 6.12).



Mutational load	*HR*	*P*	*95% CI*	
Low	-	-	-	-
Intermediate	1.3	0.3	0.79	2.1
High	0.6	0.06	0.3	1.02

Figure 6.9: Kaplan Meier plot for three mutational load groups in the whole dataset (N=301)

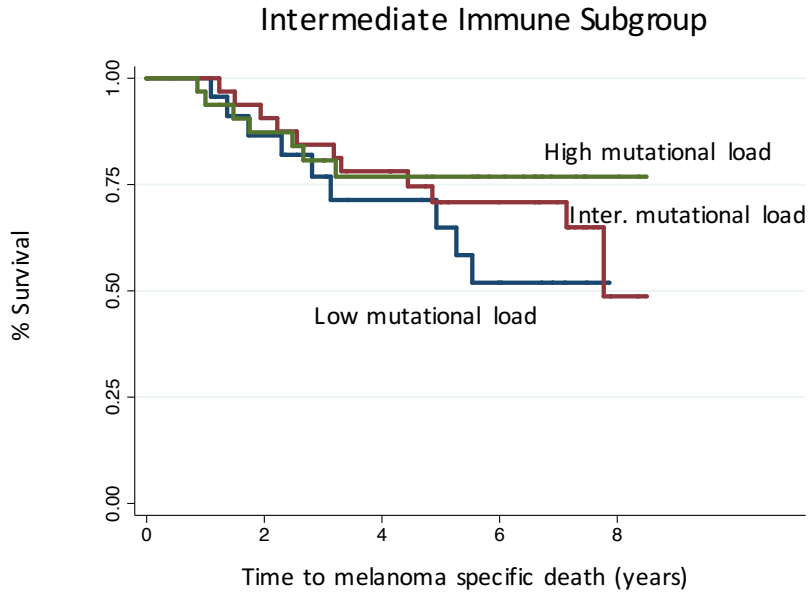
Table represents the results from a univariable Cox proportional hazard model.



Mutational load	*HR*	*P*	*95% CI*	
Low	-	-	-	-
Intermediate	1.0	1.0	0.5	2.0
High	0.4	0.03	0.2	1.8

Figure 6.10: Kaplan Meier plot for three mutational load groups in the Low Immune Subgroup (N=117)

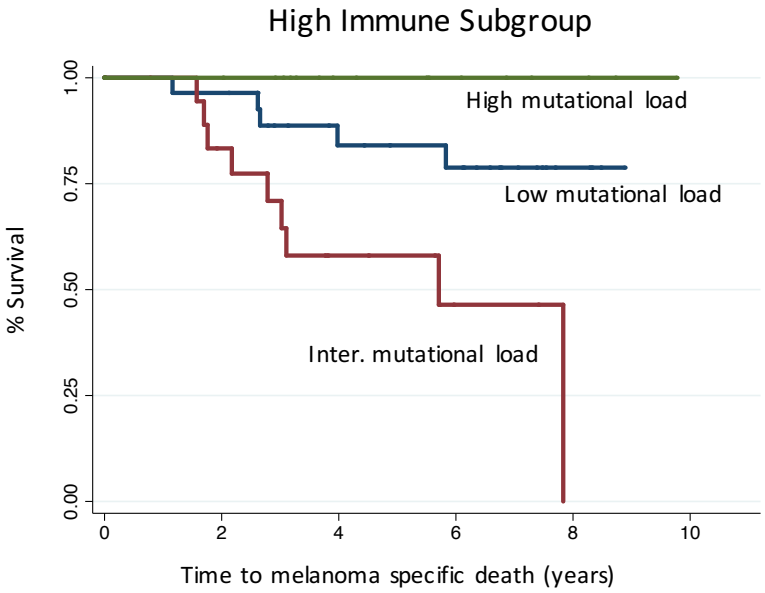
Table shows the results from univariable Cox proportional hazard model.



Mutational load	*HR*	*P*	*95% CI*	
Low	-	-	-	-
Intermediate	0.8	0.6	0.3	1.9
High	0.6	0.3	0.2	1.7

Figure 6.11: Kaplan Meier plot for three mutational load groups in the Intermediate Immune Subgroup (N=107)

Table represents the results from a univariable Cox proportional hazard model.



Mutational load	*HR*	*P*	*95% CI*	
Low	-	-	-	-
Intermediate	3.7	0.02	1.2	11.08
High	-	-	-	-

Figure 6.12: Kaplan Meier plot for three mutational load groups in the High Immune Subgroup (N=77)

Table represents the results from univariable Cox proportional hazard model.

In the whole dataset, when mutational load survival analysis was adjusted for known melanoma prognosis predictors it remained significantly protective of melanoma death (trend test, HR=0.65, P=0.02), (Table 6.11). In the stratified adjusted analysis, the mutational load significantly predicted prognosis in the Low and High Immune Subgroups (HR=0.5, P=0.01; HR=0.4, P=0.05, respectively) and the HRs were comparable. Importantly smoking (ever vs never) still predicted prognosis significantly in the High and Intermediate Immune Subgroups (HR=5.8, P=0.03; HR=5.1, $P<10^{-3}$) in these analyses including mutation load, but was not significant in the Low Immune Subgroup (HR=1.0, P=0.1) (Table 6.11).

Table 6.11: Multivariable Cox proportional hazards model for the variables significantly predicting prognosis in the univariable model including smoking (never/ever) and mutational load in the whole dataset and within immune subgroups

AJCC stage was categorized as stage I, II or stage III. Site of melanoma was tested for rest vs limbs. Age at diagnosis was reported in years. Mitotic rate is the count of mitoses per mm^2. Mutational load was categorised as high, intermediate or low. HR means melanoma specific death hazards ratio.

Predictor	*HR*	*SE*	*z*	*P*	*95% CI*	
OVERALL (N=240)						
AJCC stage	1.66	0.35	**2.39**	**0.02**	1.10	2.52
Smoking (ever)	1.94	0.54	**2.35**	**0.02**	1.12	3.36
Site (rest)	1.26	0.39	0.74	0.46	0.69	2.30
Age at diagnosis	1.05	0.01	**3.44**	**0.00**	1.02	1.07
Mitotic rate	1.02	0.01	**1.86**	**0.06**	1.00	1.03
Inter. cluster vs Low	1.06	0.33	0.17	0.86	0.57	1.94
High cluster vs Low	0.64	0.23	-1.25	0.21	0.31	1.30
Mutational load	0.65	0.12	**-2.42**	**0.02**	0.46	0.92
LOW IMMUNE (N=90)						
AJCC stage	1.49	0.46	1.28	0.20	0.81	2.73
Smoking (ever)	0.99	0.38	-0.03	0.98	0.46	2.12
Age at diagnosis	1.05	0.02	**2.64**	**0.01**	1.01	1.09
Mitotic rate	1.01	0.01	0.74	0.46	0.99	1.03
Mutational load	0.50	0.13	**-2.60**	**0.01**	0.30	0.84

Predictor	*HR*	*SE*	*z*	*P*	*95% CI*	
INTERMEDIATE (N=82)						
AJCC stage	1.59	0.66	1.10	0.27	0.70	3.60
Smoking (ever)	5.12	2.85	**2.93**	**0.00**	1.72	15.24
Age at diagnosis	1.04	0.02	1.64	0.10	0.99	1.09
Mitotic rate	1.06	0.01	**3.98**	**0.00**	1.03	1.09
Mutational load	0.74	0.24	-0.92	0.36	0.39	1.41
HIGH IMMUNE(N=68)						
AJCC stage	1.89	1.00	1.21	0.23	0.67	5.31
Smoking (ever)	5.76	4.56	**2.21**	**0.03**	1.22	27.16
Age at diagnosis	1.08	0.04	**2.11**	**0.04**	1.01	1.17
Mitotic rate	1.09	0.09	1.14	0.25	0.94	1.27
Mutational load	0.44	0.18	**-1.99**	**0.05**	0.20	0.99

6.5 Summary

- Vitamin D did not predict prognosis in the whole data, nor in immune subgroups.
- Smoking strongly predicted death from melanoma in an immune dependent manner: the strongest in the High Immune Subgroup (in the analyses including and excluding the mutational load variable).
- *GPR15* was associated with smoking status and correlated with expression of genes in the IL6 signalling pathway in the High Immune Subgroup, with replication in metastatic tumours (TCGA dataset).
- *GPR15* was equally expressed in non-ulcerated and ulcerated tumours while *IL6* was more highly expressed in ulcerated tumours.
- High mutational load predicted the most favourable prognosis, particularly in the High Immune Subgroup in the univariable analysis but the effect was comparable between High and Low immune subgroups in multivariable analyses.

6.6 Discussion

6.6.1 Effect of clinico-histopathological and environmental factors on prognosis in different immunological contexts

6.6.1.1 *Clinico-histopathological factors*

AJCC stage was one of the strongest independent survival predictors in primary melanoma in the whole dataset and it remained consistently significant in stratified analysis by the immune subgroups. This observation indicated that irrespective of the level of immune response within melanoma AJCC stage is still a strong prognostic marker. Higher mitotic rate was borderline significant for melanoma death overall and subgroup analysis suggested that the strongest association was within the intermediate immune subgroup.

Site of melanoma significantly predicted prognosis mainly in the Low Immune Subgroup, however this effect was driven by the fact that the rare (sun protected) sites were mostly classified into this group, and as mentioned in Chapter 1 these are the tumours that have the worse prognosis for the melanoma patients [14][15][17][18] possibly due to low C>T mutation rates [19], or late diagnosis [16].

6.6.1.2 *Smoking*

The environmental exposures were analysed: vitamin D levels and smoking, however only smoking showed to be an important, very strong predictor of MSS in an immune dependent manner. The effect of cigarette smoking was the strongest in the High Immune Subgroup even though this was the smallest subset and the analysis had therefore the lowest statistical power. Higher cancer specific death associated with smoking was reported in breast cancer [241], head and neck [242], ovarian [243], prostate [244], and colorectal cancer [245]. However, none of these studies has tested the dependency on the immune responses.

The absence of differentially expressed genes in tumours in relation to cigarette smoking in the agnostic analysis, despite a significant association with melanoma death was surprising. If overall survival was used it might have been hypothesised that the detrimental effect of smoking was related to non-melanoma deaths but our study has

used accurate measures of cause of death and therefore we were confident that the increased hazard of death in smokers was related to melanoma itself. If I had more time during my PhD I would repeat the survival analyses testing the overall survival in order to explore whether the smoking effect was also related to non-melanoma specific death, although the number of non-melanoma death is limited.

Given the well reported association of smoking with markers of systemic inflammation e.g. C reactive protein levels [246], we explored the association of smoking with immune cell scores but no significant associations were seen either for immune cells or for genes coding for immune checkpoint molecules. It is commonly accepted that smoking might induce either pro- or anti-inflammatory systemic states [247][248] implying complex effects on host/tumour interaction.

There are a number of hypotheses which might explain the lack of detectable immune differences in the tumours of smokers vs non-smokers. The first hypothesis is that smoking may impair tumour cell killing functionally even when immune cells are present in the tumour, by downregulating the functions of adaptive and innate immune cells [88]. The second hypothesis is that inference of immune cell subpopulation from transcriptomic data might be insensitive to subtle changes in immune function. Or this changes were not related to the gene expression level but for example to protein modifications (so called non-genetic modifications). Third, the immune cells inferred within melanomas could be a reflection a chronic inflammatory state (from smoking or other causes) rather than specific anti-melanoma inflammation [88][248]. That is that the inferred T cells might not be targeted to tumour cells: rather that they are present within tumour cells as they may be present in other tissues in smokers.

Furthermore, some studies showed that smoking might affect patient microbiome in the gut, which in turn might affect systemic immune responses [249]. Interestingly, recently researchers studying responses to immunotherapy in melanoma patients have shown that microbiome might indicate the outcome of this therapy. One study reported that higher diversity and relative abundance of bacteria of the Ruminococcaceae family [250] were associated with better responses to immunotherapy. Of note, another study analysing changes in microbiome of side-stream smoking mice (irrespective of cancer) reported that the relative abundance of

Ruminococcus albus (Ruminococcaceae family) in these animals was decreased [251]. Another study reported that microbial diversity increases after smoking cessation in human samples [252]. In conclusion, these reports suggest that smoking might have an impact on anti-tumour immune response and immunotherapy outcomes by affecting patients' microbiome, followed by changes in systemic responses.

Cigarettes are composed of 4.500 compounds five of which are cancerogenic [253]. The major compound of cigarettes is nicotine, but it is not clear whether it acts as carcinogen, however it has been proposed to cause immunosuppression [253][254][255]. It is of note, that e-cigarettes, a relatively new method of "safe" smoking electronic nicotine delivery system (ENDS), might also be harmful and affect immune responses. It has been recently shown that the e-cigarette liquid induces inflammatory states of alveolar macrophages *in vitro* [256]. On the other hand, e-cigarette vaping was reported to induce downregulation of immune genes in nasal epithelial cells [257].

Taken together, it is clear from the literature review that smoking does affect the systemic or tissue-specific immune responses as it might also be for e-cigarettes, however in melanoma it remains to be elucidated. In our study, cigarette smoking reduced patients survival, more so in those with strong immune infiltration, indicating that smoking impairs immune cell function rather than limit their numbers.

In summary the results obtained from the survival analyses stress the importance of smoking cessation, for both classical cigarettes and e-cigarettes for melanoma patients and those at risk of developing melanoma. Additionally, because the strongest risk was seen in the High Immune Subgroup it also reinforces the view that smoking might generate a "false positive" immune responses within the tumour, which do not kill the tumour. Therefore, smoking habits of the patients should be taken under consideration in immunotherapy clinical trials.

6.6.1.3 ***GPR15***

The observation of *GPR15* over-expression in smokers corroborated the literature, which suggested this gene as smoking biomarker. However, it was not proven that smoking directly increases *GPR15* expression. It is rather proposed that smoking might induce the excess of the *GPR15* positive immune cells [237]. *GPR15* is an orphan

chemokine receptor and with as yet unidentified natural ligand [238] and it is known to be expressed in immune cells, colon mucosa, but also in skin and bladder [238]. The knowledge about biological functions of *GPR15* within each tissue type is limited, however it is proposed that it might be involved in their chronic inflammation, due to attraction of homing T cells into skin and colon [238][258].

In this chapter, I reported a positive correlation between *GPR15* and genes involved in IL6 signalling in the High Immune Subgroup, from 2 datasets (the LMC and TCGA). IL6 signalling is perceived as being involved in chronic and systemic inflammation [259][260]. Furthermore, IL6 was shown to regulate almost all hallmarks of cancer in various cancer types, including inhibition of apoptosis, promotion of survival, proliferation, angiogenesis, invasiveness and metastasis [261]. However, the exact role of IL6 signalling in melanoma is still not fully understood [262]. The Leeds Melanoma Research Group has reported evidence of its involvement in microscopic ulceration of primary melanoma, a marker of a chronically inflamed tumour, associated with systemic inflammation [220][221]. *GPR15* did not significantly vary between ulcerated and non-ulcerated tumours, however *IL6* itself was more highly expressed in ulcerated tumours. It would be interesting to follow up the correlation of *GPR15* with the cytokines in more details, however the time did not allow me to do so.

Summarising, the *GPR15* expression in the melanoma tumours was significantly associated with smoking status and borderline with IL6 signalling in the High Immune Subgroup, which could support the hypothesis that the presence of the immune cells within melanoma tumour of the smokers was a non-melanoma specific (chronic) inflammation phenomenon. Clearly this is an important issue which encourages for further investigation, which should include IHC staining for GPR15 to define the origin of its transcript signal.

6.6.2 Mutational load and prognosis

The survival analyses including the mutational load variable had reduced sample size than the analyses excluding this variable, and therefore reduced statistical power.

The high mutational load was associated with better outcome for melanoma patients overall and in the High and Low Immune Subgroups, but not in the Intermediate Subgroup. The results remained significant even after adjusting for other strong

survival predictors. The high mutational load was associated with the most favourable outcome, but the poor outcome was related to intermediate mutational load, which is counterintuitive and challenging to understand. It may be explained by small sample size and calls follow up. In Chapter 3 it was shown that the mutational load did not vary across the three immune subgroups, which could be a true observation, or simply a result from the limitations of our dataset: the mutational data used for these analyses were not ideal as discussed in Chapter 3, since they covered a small number of genes. Nevertheless, assuming that the data truly reflected the unmeasured whole genome mutational load and that the mutational load could reflect the neoantigen load, it could be suggested that in LMC primary tumours the immune responses were not driven by it. All the results taken together suggested that melanoma cells need to have an effective antigen presentation machinery (as it is the case in the tumours from the High Immune Subgroup) in order to present the neoantigens to the immune system and only then the survival will improve (no death was observed among 21 High Immune patients with high mutation load up to 10 years after diagnosis). This observation was not surprising, as the literature proposes that either having downregulation and/or absent neoantigens are known immune evasive mechanisms in cancer [87] and that high mutational load was associated with better immunotherapy responses in melanoma [145][146][147] and other cancers [263], however lack of association between mutational lad and level of T cell infiltration across TCGA tumours was shown [264]. Therefore, even on a limited scale (i.e. based on a candidate gene panel), mutation data have the potential to be informative for the efficient use of immunotherapies. It would be interesting to conduct a comparative analysis of this economical approach and the costlier whole genome sequencing.

Chapter 7

Final Discussion

7.1 Methodology

7.1.1 LMC

The Leeds Melanoma Cohort data set contains a collection of rich clinical, histopathological, exposure/environmental, and transcriptomic data from primary cutaneous melanoma tumours.

The transcriptomic data were the basis of the analyses in this thesis (703 samples), which were replicated in independent datasets (TCGA and patient derived melanoma cell lines), validated by immunohistochemical staining (IHC) and further expanded to analysis of copy number alterations (CNAs) and mutational data, albeit for consequent reduced sample sizes.

The development of checkpoint therapies and their positive impact on patient survival has reduced the focus on understanding the biological processes underlying the immune responses to the primary tumours [87], hence studies focusing on primary untreated melanoma, are in need. Therefore, the LMC is considered as a unique collection of the data, which allows us to address this research question.

The LMC is one of the biggest population ascertained cohorts worldwide and this unique collection of data allowed me to perform agnostic genetic and survival analyses with statistical power.

7.1.1.1 *Limitations of the data originating from the LMC*

LMC patients were predominantly treatment naive, in that only 16 of the 703 patients have been known to have received checkpoint therapies, as the period of recruitment (2000-2012) preceded frequent usage. However, as time passes more of them will receive such treatment at late relapse. The Leeds Melanoma Research Group is therefore in the process of seeking national data from Systemic Anti-Cancer Therapy (SACT) dataset to update the treatment information. I have reduced the chance of treatment responses influencing my findings by removing the 16 treated participants

and by repeating the survival of the three immune subgroups and the results were unchanged.

The transcriptomic and genomic LMC data were generated from FFPE tumour blocks. The storage of the tumours in formalin means that the tumour samples remain intact with the original and intratumoural shapes, allowing histopathological examination beyond the original clinical examination. The fixation process however affects the protein and particularly genetic material structures. DNA and RNA extracted from FFPE samples are usually degraded, and their strands are shortened [265] or may form crosslinks with proteins [266][267]. These changes might impact on the quantity of the extracted genetic material, introducing noise into the final data and/or meaning that some of the genes or their transcripts might not be detected by the probes or sequencing. Researchers have established procedures of extracting DNA and RNA from FFPE samples to reach higher efficiency and developed methods of gene expression analysis in order to overcome these problems and these were used by Leeds Melanoma Research Group. For example, the gene expression was measured using the Whole-Genome DASL HT12v4 assay (Illumina®), which was particularly intended for the FFPE samples, by designing short probes (50bases) to enable the detection of degraded RNA. Note that this platform is not on the market any longer. Tumour samples might also be stored fresh and frozen in liquid nitrogen and this method affects the genetic material to a lesser degree than formalin fixation, however as mentioned above storage of samples as FFPE is preferred due to the preservation of the tumour structures, which are important for histopathological and immunohistochemical assessments. The studies benefitting from fresh frozen samples derive their samples from tumours which are large enough at excision for the pathologist to be satisfied to sacrifice a small proportion for research. Therefore, the sample sets built using this approach are inevitably biased towards thicker tumours. The TCGA data set was therefore predominantly composed of metastases and, even so, required the pooling of samples from international research centres. I have listed the use of formalin fixed tumours here as a weakness because of degradation of RNA but the Leeds Melanoma Research Group regard the design as a strength as the sample set is derived from population ascertained participants with primary disease and although there is some bias towards the use of thicker tumours, the bias is much less extreme than for example in TCGA.

Importantly, because this is a population-based cohort, the design allows extrapolation to findings to the general population of melanoma patients, not just for those for whom a metastatic biopsy can be obtained.

Heterogeneity of tumours is a well-documented feature of all cancers. The Leeds Melanoma Research Group chose to derive RNA from small tissue micro array (TMA) needles as when tumour sampling began, the quantity of RNA required was considerably larger than is now required. Using a 0.6mm diameter core allowed the generation of sufficient RNA without damaging the tumour block which the patient might require in the future for clinical testing. The group consistently sampled a region of the invasive tumour with the least stroma and inflammation order to make the samples comparable within the cohort. However, the gene expression data were usually analysed only from one cored region and these bulk tumour data could not therefore delineate the distinct tumour gene expression patterns, for example in case of heterogeneous tumours.

This sampling method does not allow the distinction of the origin of the gene expression signal from the distinct components of the tumour microenvironment, such as fibroblasts, immune cells, blood vessels etc. However, it was the intent to sample the tumour in its microenvironment and indeed my project was dependent on this approach. The use of laser microdessection [268][269], can be adopted if the intent is to study specific cell populations or single cells from the tissue on the glass slide however it is very expensive and not logistically feasible for a large number of samples. Another method to address the heterogeneity issue could be single cell RNA sequencing from different parts of the tumour, however for this a fresh tissue (just after excision) is needed. Non-the-less single cell analyses are being carried out increasingly and these will complement the approach described in my thesis.

The use of a TMA needle to sample the blocks generates relatively generous quantities of RNA but requires histopathological expertise in sampling. The haematoxylin and eosin stained slides were reviewed by Prof. Julia Newton-Bishop and Dr Jonathan Laye and were marked using a microfine felt tip pen. The intent was to sample the deepest part of the tumour but at a depth where a core would remain predominantly within tumour. In fact ,using the package Estimate [270], Nsengimana et al. reported [76] that the tumour purity was comparable to that reported in the literature for most sampling

techniques [46][117] and can therefore be judged as successful. In some tumours tongues of epidermis can be found even deep within the tumour. Moreover, some of tumour samples were relatively thin and the needle used for the core extraction could reach the keratin components and then the gene expression could be diluted by the gene signals from the "healthy" skin surrounding the tumour. Some of the results confirmed this concern, such as the "keratin" TCGA class in LMC (Chapter 3; and MYC dominant group (Chapter 4). However it is an inherent issue for transcriptomic studies devised to investigate host-tumour interaction as keratin tumour classes were reported by TCGA consortium (keratin group was predominantly (74%) comprised of primary tumours in TCGA) [46] and by Cirenajwis et al. ("normal-like" group was comprised of 62% of primary tumours and 31% of metastases of which 23% were local metastases) [136].

Finally, the samples for which transcriptomic data could be combined with IHC, CNA or mutational data were available, were on average thicker (as generating sufficient RNA and DNA for all approaches requires bigger tumours) and the sample size was therefore smaller. This could be considered as biased/selected data, not representing the LMC as a whole.

In summary, although some of the data originating from LMC could be potentially biased this dataset was nevertheless more representative of the population ascertained cohort compared to TCGA data, which was selected and from different centres and populations.

7.1.2 Bioinformatic analyses to infer immune microenvironment of melanoma tumours

Immunotherapies, especially checkpoint blockades showed a great success in melanoma treatment and this achievement is also emerging for other cancers [271][272]. For melanoma only a proportion (~50%) of patients benefit from these therapies [95] and on average similar response rates are observed in other cancers [272]. Therefore, understanding of the interaction of the tumour and immune microenvironment is critical to improving the impact of these therapies. Immunotherapies principally act by unleashing/increasing anti-tumour immune responses. The failure of the response to these therapies might be due to the absence

of the immune cells within the microenvironment. The lack of immune responses within the tumour is likely to be mediated by primary or acquired resistance mechanisms, such as immune evasion (discussed later on) or immune editing [273]. Another explanation of ineffective immunotherapies might be insufficiency of usage of the anti-checkpoint molecules antibodies (anti-PD1, anti-PD-L1, anti-CTLA4) blocking the immune responses, as these are not the only existing checkpoint molecules.

The studies to recognise the patients' response to immunotherapies are designed to understand different tumour types on the molecular level, the host immune responses and their interactions.

Immunological research has traditionally used the reductive experiments such as *in vitro* examination of the functioning of the immune cells. By contrast, the bioinformatic approaches enable the study of more complex interactions on a greater scale (while requiring validation by *in vitro* and *in vivo*). This is what I have attempted to explore during this PhD work.

I have used the inferred immune cell scores derived from relatively small areas of the tumours and categorised the tumours based on these scores, examining the whole transcriptome and clinico-histopathological characteristics resulting in the identification of three different immune microenvironments.

Bioinformatic characterisation of the tumour immune microenvironment using gene expression data is rather challenging. At the beginning of my PhD project, few bioinformatic methods to infer immune cell scores were available, but over time new approached emerged and many of them are based only on RNA-seq data [274]. In my thesis I have explored CIBERSORT and developed a method based on Angelova et al.'s Immunome, because these methods were available at the time of my initial analyses, were the most suitable for the microarray data and allowed inference of large numbers of different immune cell scores.

Although CIBERSORT is an attractive algorithm for deconvolution of the immune cells from bulk tumour data, I observed that it had some limitations. For example, it failed to estimate the immune cell proportions from the tumours that manifested weak/moderate immune signals and more than a half of the samples could not be characterised. Therefore, this method was not used for analysis of the LMC samples.

Application of the Angelova et al.'s Immunome for the characterisation of the immune cell scores within the LMC tumours, leaves uncertainties as to the origin of the measured gene expression signals (it was not clear from the literature or my examination of transcriptomes derived from melanoma cell lines) whether the expression of some genes originated from the melanoma or immune cells and specific adjustments were necessary. I found that almost half of the "Immunome" genes were highly expressed by the melanocytes and melanoma cell lines.

Nonetheless, the successful use of Angelova et al.' Immunome to colon cancer was reported [1]. In the LMC the immune cell scores were created after application of the gene filtrations steps (based upon gene expression by melanoma/melanocyte cell lines and excluding negatively correlating genes per cell type specific-score) and the results related to these scores were replicated in the fresh frozen NGS TCGA data.

Clustering of tumour samples is commonly performed based on gene expression data [132]. In studies designed to find global tumour subtypes within a given dataset information on a large numbers of genes are used [46][136]. However, to address specific questions, only the genes of interest might be utilised for clustering. For example, Dr Jérémie Nsengimana applied the Bindea et al.'s Immunome (immune genes) to cluster LMC tumours and identified 6 immunologically and histologically different subgroups [76]. In my work I reasoned that using the 27 immune cell scores instead of hundreds of genes could reduce the dimensionality for clustering and result in more stable clusters, which would be less affected by the variation caused by histological or technical differences. Indeed, three immunologically different stable clusters (with distinct survival) were obtained, which were replicated in metastatic tumours from TCGA. The clustering analysis of the immune cell scores revealed that the majority of the immune cell scores were strongly correlated. Overall it was observed that the immune genes were highly correlated with each other and high expression of these cells was associated with good survival. This fact could mask putative subtle differences of the immune cell scores, which were highly related to each other at least on the gene expression level.

My findings are in some ways similar to those reported in a recent publication, in which Pan-Cancer clustering of TCGA tumours was performed based on 6 immune signatures: Wound healing, IFN-gamma dominant, Inflammatory, Lymphocyte depleted,

Immunologically quiet, and TGF-beta dominant [275]. Although a direct comparison between Thorsson's clustering and mine is slightly problematic as Thorsson et al. developed clusters based on 6 immune signatures, which were dissimilar from the 27 immune cell scores I generated in the LMC. Moreover, the Thorsson's signatures were not adjusted to cancer genes before the generation of immune signatures as it was performed in the LMC by removal of "melanoma" genes from the list of genes used to create the immune cell scores. The Pan-Cancer research group showed that melanoma samples were mainly dominant in: Wound healing, IFN-gamma dominant, Inflammatory, and Lymphocyte depleted clusters [275]. The number of immune clusters to which melanoma samples were assigned was comparable to the number of clusters obtained in LMC (four vs three), but the immune clusters were slightly dissimilar phenotypically from the ones identified in LMC. However, the Thorsson's Inflammatory cluster could be compared to the LMC High Immune Subgroup, and both of these immune clusters were associated with good survival. The Thorsson's Lymphocyte depleted cluster could be compared to the LMC Low Immune Subgroup, for which the survival was poor both in LMC and Pan-Cancer. The Thorsson's Wound healing, IFN-gamma clusters were not identified in the LMC. However, the survival for these clusters could be compared to the LMC Intermediate Immune Subgroup, for which it was moderate in comparison to the rest of the immune subgroups. Additionally, another similarity that was observed between the Pan-Cancer and mine analysis was for example, higher expression of *CTNNB1* associated with "Lymphocyte Depleted" cluster associated with the worse survival in the Thorsson et al.'s study [275] and alike observation was made in the LMC.

Summarising, the three immune subgroups displayed a biologically informative subgrouping of the tumours, which was productively investigated from the transcriptomic, genomic and environmental perspective. These groups allowed us to discover new, and confirm already known, immunosuppressive pathways in melanoma. Moreover, the impact of smoking could only be discovered in this immune dependent manner, facilitated by having sufficient information to test the survival of the non- and smokers in each of the immune subgroup with adequate statistical power.

7.2 Host immune responses

Understanding the immune responses to melanoma and to cancer generally has become one of the main focuses in cancer research as discussed above.

Currently the understanding of host-tumour interaction is focused on studies investigating the primary and acquired resistance to immunotherapies. However, as mentioned above, it is important to uncover the initial/basic mechanisms of immune evasion/editing in the primary tumours. It is crucial to study the primary tumours as the understanding at the level of metastatic disease might be too late for the patient to be cured. Understanding the immune responses to early stage primary melanomas might result in discovery of prognostic biomarkers, identification of more effective adjuvant therapies and lifestyle factors which affect the overall immune response.

The anti-cancer immune responses can be disrupted in each step of the path leading to cancer killing: from the release of the cancer antigens through priming, activation and trafficking of the T cells into the tumour, to the final cancer killing phase (described by Chen and Mellman [87]).

Firstly, the release of cancer antigens/neoantigens is one of the initial mechanisms for tumours to be recognised by the immune system. Higher neoantigen load was proposed to have an impact on responses to immunotherapies and patients' survival in various cancers, particularly in melanoma [145][146][147]. Alexandrov et al. described several mutational signatures and one of them was "signature 7" (rich in C>T mutations, the UV exposure signature), which was particularly highly enriched in melanoma. This signature was further shown to predict better survival and high immune cell scores in the TCGA dataset [148]. All the evidence suggests that neoantigen load/signature might trigger the immune responses. However, the whole machinery of neoantigen presentation is essential for effective antigen presentation. Loss of any major component of this process might impact on immunogenicity of the cancer cell (as discussed next). In this thesis, it was shown that in primary melanoma high mutational load was associated with better outcome. However, there was no overt relationship between the level of immune infiltration and the neoantigen load. Similar observation was made by Spranger et al. when analysing the TCGA data for various tumour types including melanoma, where it was shown no significant

difference of mutational load between inflamed and non-inflamed tumours [264]. One possible explanation for this could be incomplete data on the mutational/neoantigen load for the LMC, since it was represented by only a 555 gene panel. Another explanation could be that it is not the load itself but quality of the presented antigens which is the most important factor [150][151].

Next, in order to present the neoantigens to the immune cells, tumours are required to have the active antigen processing and presentation machinery. The down-regulation of HLA class I signalling and losses in *B2M* were shown to be inversely correlated with two gene cytolytic score (*GMZA*, *PRF1*) in various cancers, including melanoma [276] and to be involved in affecting the immunotherapy responses [277].

The gene expression of antigen processing and presentation via HLA class I and II were significantly lower in the Low Immune Subgroup, which was not surprising. However, in this work the direct inverse relationship with HLA class I genes and the oncogene *MYC* was proposed as a novel finding. It suggested that this oncogene is not only involved in tumour progression by stimulating proliferation, but also by contributing to immune evasion. The negative association of MYC and HLA-B was already proposed in 1994 by observation of melanoma cell lines [187]. However up to now it was not shown in human melanoma samples.

A recent study by Jerby-Arnon et al. (November 2018) analysing single cell RNA sequencing from treated and non-treated patients by checkpoint blockade, reported that non responding tumours manifested up-regulation of MYC targets and down-regulation of HLA class I molecules [278], which is partially consistent with the results from this thesis.

Other mechanisms of immune evasion related to tumour intrinsic oncogenic pathways, including MYC were reported in various cancers [279]. For example, MYC was reported to up-regulate PD-L1 and CD47 molecules on human leukaemia and lymphoma cells [280], and to down-regulate HLA class II in human B cells tumours [188]. If I had additional time during my PhD, I would perform an experiment to prove this observation by knocking down, or overexpressing MYC in melanoma cell lines, and testing the expression of *HLA-B* and antigen processing and presentation genes using (quantitative Reverse Transcription Chain Reaction) RT-qPCR.

Moreover another oncogenic pathway such as loss of PTEN (a tumour suppressor gene) was demonstrated to impact on the reduction of T cell infiltration and T cell killing in melanoma mouse models and human samples treated by anti-PD1 therapy [281]. The work of this thesis did not uncover specifically PTEN involvement in immune responses on the gene expression level, however if I had more time I would test the CNA of this gene across the three immune subgroups.

Summarising all the evidence of MYC functioning as a tumour immune suppressor, suggests that patients who do not respond to immunotherapy could be given MYC inhibitors to sensitize non responding tumours. MYC has already been considered as drug target as it is well known oncogene, but translation research has been difficult to target MYC directly in solid tumours [282]. The *MYC* induced tumours might require other targeting approaches such as for example an oncotic viral activity, which was reported to be permissive to tumours driven by *MYC* [283]. Alternatively, one study based on xenograft melanoma mice models, has suggested antisense oligonucleotides as viable treatment option in human patients after observing tumour shrinkage and expanded lifetime in mice [284].

Even though tumours that manifest generation of neoantigens and have an effective antigen processing and presentation machinery need to provide a suitable niche for the effective infiltration of the immune cells into the tumour microenvironment. One of the results from this work, which could potentially explain the regulation of this process in melanoma was the observation of the down-regulation and losses of the components of tumour NF-κB signalling associated with lower immune infiltration and worse melanoma specific survival. These results could be considered as an immune evasion mechanism acquired by the tumour. Abnormal NF-κB signalling and *NFKB1* gene expression was particularly found to have carcinogenic functions and promote tumour progression as well as to act as tumour suppressor [190][191]. However, the deletions, and down-regulation of genes belonging to this pathway were not shown to be involved in immune evasion in melanoma and other solid cancers according to my literature search. However, one study described that loss of *NFKB1* was causal for gastric cancer development specifically with aberrant inflammation within the tumour [196]. Moreover tumour NF-κB signalling was reported to be associated with T cell infiltration and not with proliferation in mice and human lung cancer [285].

Additionally, a hyper-activation of NF-κB signalling within the tumour microenvironment was shown to contribute to reprograming of chemokine microenvironment to enhance recruitment of cytotoxic T cells into colorectal tumours [195].

In the LMC, the immunohistochemical staining analysis showed that tumour nuclear localization of NF-κB significantly correlated with lymphocytic infiltration, suggesting a reciprocal NF-κB-driven phenotype generated between the tumour and its surrounding microenvironment as described in the two studies mentioned above [285][195]. These results and the inferior survival for the patients with losses of the NF-κB pathway might highlight a potential warning for the implementation of targeted therapeutic inhibition of the NF-κB pathway in melanoma (Beleyer et al. [286][287]), because it could have adverse immune effects if in fact tumour NF-κB activated pathways are actively attracting T-cells to the melanoma tumour microenvironment.

NF-κB signalling is a complicated process, particularly in cancer. Understanding its functioning in melanoma using only gene expression data is challenging, since it can only explain the level of association not causation. However, the CNA data could partially delineate the causative effect of the deleted genes. The observation concerning losses of NF-κB pathway genes could be further validated by performing a knock out e.g. for *NFKB1* from the human tumour cells and test how does this affect the gene expression of downstream targets of NF-κB signalling. Most importantly it could be tested whether these cells were indeed less immunogenic by conducting immune cell killing experiments.

β-catenin signalling is a pathway already shown to affect the immune cells trafficking (mainly dendritic cells) into the tumours [75] – the middle stage of the melanoma killing path. This pathway was revealed to be upregulated in a great proportion of primary melanoma tumours in LMC [76], and to regulate immune responses in other cancers [288] and this thesis reports analyses supporting these results. β-catenin is known to regulate MYC, however the reciprocal regulation was also proposed [149]. Moreover, it was recently reported by Prof. David Fisher from Harvard University (oral presentation at the 15th International Congress of the Society for Melanoma Research in Manchester, October 2018) that gain of function amplifications in methyltransferase

G9A/EHMT2 (Euchromatic Histone Lysine Methyltransferase 2) might be involved in cancerogenic processes in melanoma and other cancers. Moreover, G9A was shown to increase the WNT/β-catenin pathway, which subsequently contributed to decreasing the immune responses (data not published). His research group proposed that G9A suppresses DKK1 a repressor of WNT/β-catenin pathway in melanoma mouse models and *in vitro*. In our collaborative analyses we have confirmed some of these results by using the LMC data. We performed differential gene expression between *G9A* high and low copy number changes and observed that among top pathways associated with *G9A* high tumours was WNT signalling and among the genes were *MYC* and *DKK1*. Moreover, we observed that high *G9A* copy number was significantly higher in the CIC 4 (Low Immune/β-catenin high) from Nsengimana et al.'s consensus immune clusters.

Once the tumour is infiltrated and recognised by the immune cells it may evolve and affect the immune cells so that they become inefficient in their killing functions (mechanism described as "immune suppression"). One of the most recognised immune suppressive mechanisms in cancer including melanoma is persistent activation of regulatory immune cells within the tumour microenvironment, such as Tregs [289][81][290] and MDSCs [291][292]. Moreover, up regulation of checkpoint molecules such as PD-L1, PD1 and IDO by the tumour or immune cells within the tumour microenvironment, are considered to be involved in suppressing the activation of T cells [87]. However, some studies including the one from the Leeds Melanoma Research Group, reported that checkpoint molecules and regulatory immune cells correlate with the overall immune cells abundance and improved survival, and this observation is considered as homeostatic feedback [189][76][293]. Overall, the immune homeostatic feedback is important in terms of prevention of the development of excessive inflammation i.e. by checks and balances. The results of this thesis corroborated these observations, as all the immune cell scores were highly correlated and the immune checkpoint molecules were significantly upregulated in the High Immune Subgroup.

Another mechanism affecting responsiveness of the immune system in different phases of the melanoma killing path is impaired interferon gamma signalling [294][295]. The disruption of this signalling might result in down-regulation of antigen processing and presentation as well as in lower production of cytokines that might

attract anti-tumour immune cells [218][296]. I observed that the genes involved in interferon signalling were downregulated in the Low Immune Subgroup and *JAK2* deletions were frequently observed in this subgroup. Mutations/deletions of *JAK2* have recently been reported to be involved in primary and acquired resistance to anti PD-1 therapy in melanoma metastatic tumours [217][218]. The results from this thesis indicate an early onset of these aberrations in melanoma progression, which might further impact on patients' responsiveness to immunotherapies.

Summarising the overview acquired from all the results from transcriptomic, and CNA data together allowed me to conclude that immune evasive pathways were not mutually exclusive, and moreover their co-existence was shown to worsen the patient's survival. It could be concluded that the tumours in which many immune evasive pathways were observed could imply higher tumour heterogeneity, and different immune evasive mechanisms being active. Moreover, these observations suggest that a great proportion of melanoma tumours in the Low Immune Subgroup could have an intrinsic resistance to immunotherapies.

Finally, environmental factors might influence systemic immune responses, which might have a further effect on recognition or killing of the tumour by the immune cells. One of the well-studied determinant of inflammatory/suppressor effect on systemic immunity is cigarette smoking [247][248].

In this thesis, it was shown that cigarette smoking decreased the protective effects of immune cells within the tumours in the High Immune Subgroup. Importantly, these results were consistent when using three different variables describing habits of smoking and it was shown that it was not cofounded by deprivation status. It was also observed that *GPR15* expression, a known biomarker for smoking in the circulating blood [237], was significantly associated with smoking in the LMC tumours. Furthermore, *GPR15* was suggested to be associated with chronic inflammatory pathologies [258] and in LMC and TCGA it was observed that *GPR15* suggestively correlated IL6 signalling, an inflammatory pathway. However this observation would need a further validation on the protein level and testing for whether it was related to systemic or *in situ* inflammation.

The literature is rich in describing the studies regarding cancer specific survival and cigarette smoking. For example smoking was shown to shorten survival for breast [241], head and neck [242], ovarian [243], prostate [244], and colorectal cancer [245] patients. Previously in the LMC it was demonstrated that smoking was associated with worse outcome [221]. However, none of these studies showed the interaction between smoking and the immune responses. Hence our results might be considered as novel and clinically relevant, stressing the importance of smoking cessation and incorporation into the studies analysing immunotherapy responses. Moreover, since it is not clear whether the survival effect of smoking in melanoma patients was due to different cigarette compounds, or mostly nicotine, it is important to stress that e-cigarettes might be potentially as harmful as cigarettes due to nicotine. As a result of relatively short existence of e-cigarette (vaping) there is not much evidence that the nicotine delivery system affects the immune responses systemically. However the system were shown to affect immune responses *in vitro* [256] and in the human nasal tissue [257].

Summarising, the strength of this study was the ability to discover new and confirm already reported immune evasive mechanisms in primary melanoma using gene expression, copy number and protein level data. Furthermore, I made the novel observation that the survival associated with a well-known environmental factor such as smoking interacted with immune responses measured at the gene expression level. All the novel results from this thesis might be further investigated in relation to immunotherapy responses in melanoma and other cancers.

Chapter 8

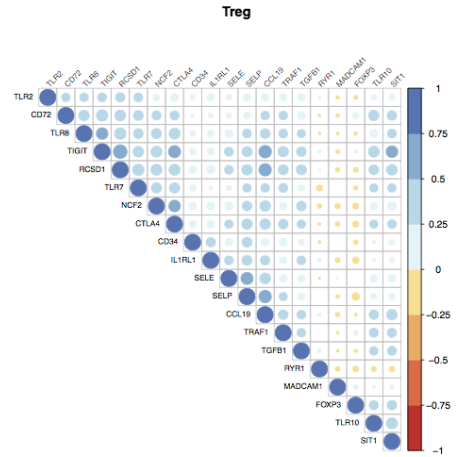
Appendix A

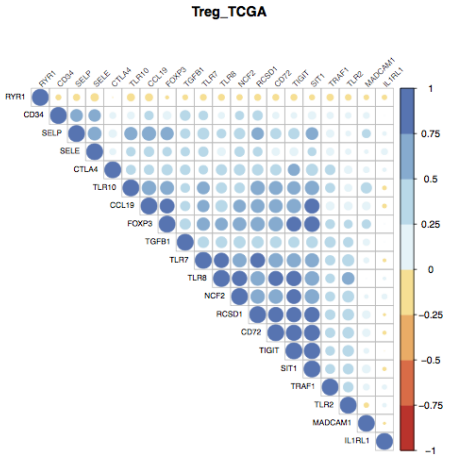
A.1 Creation of the immune cell scores

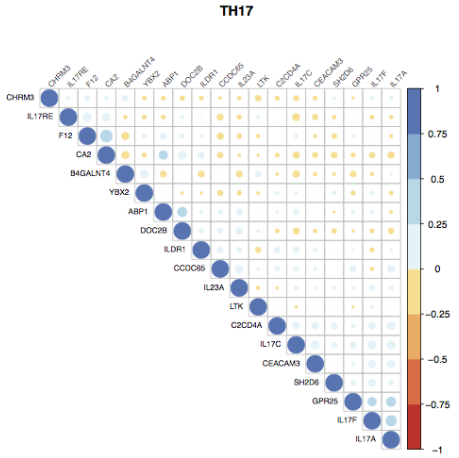
A.1.1 Percentage calculation of the remained genes per immune cell score after the first filtration step

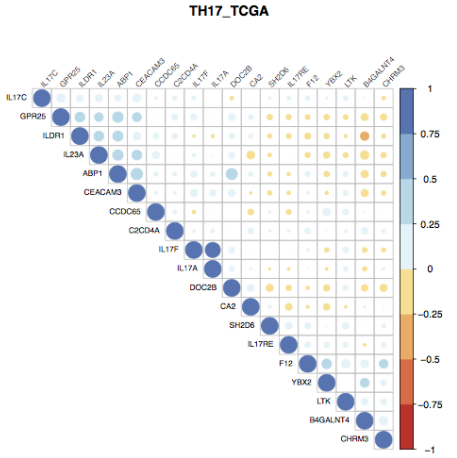
Cell type	*Before (number of genes)*	*After (number of genes)*	*% of genes remained*	*Removed*
mDC	21	0	0.0	x
pDC	18	1	5.6	
Effector_memory_CD4	14	1	7.1	x
Activated_CD8	24	2	8.3	x
Activated_CD4	35	3	8.6	x
Monocytes	25	5	20.0	
Memory_B_cells	14	5	35.7	
NK56_bright	8	3	37.5	
MDSC	85	37	43.5	
Th2	25	11	44.0	
NK56_dim	17	8	47.1	
Central_memory_CD4	21	10	47.6	
iDC	21	11	52.4	
Eosinophils	15	8	53.3	
Central_memory_CD8	18	10	55.6	
Effector_memory_CD8	42	25	59.5	
TFH	10	6	60.0	
Immature_B_cells	24	15	62.5	
Neutrophils	16	10	62.5	
NK	17	11	64.7	
Cytotoxic_cell	9	6	66.7	
NKT	12	8	66.7	
TGD	43	30	69.8	
TH17	26	19	73.1	
Mast_cells	42	31	73.8	
Treg	26	20	76.9	
DC	34	27	79.4	
T_cells	86	69	80.2	
Macrophages	22	18	81.8	
Th1	37	31	83.8	
Activated_B_cell	18	17	94.4	

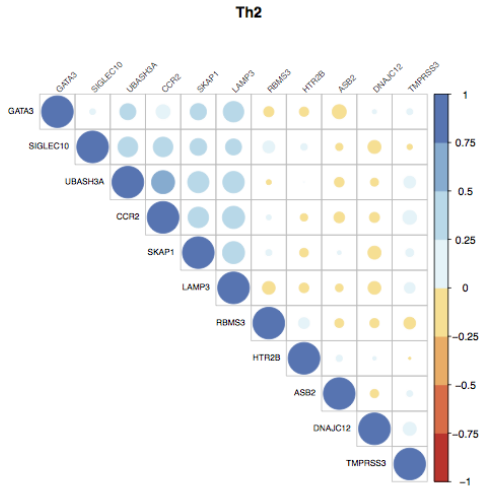
A.1.2 Correlation matrices of the genes per immune cell scores

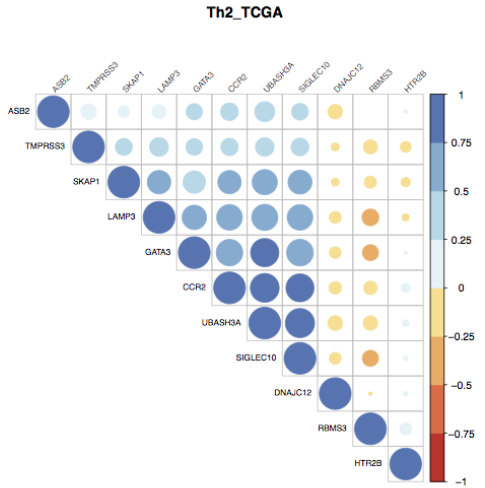


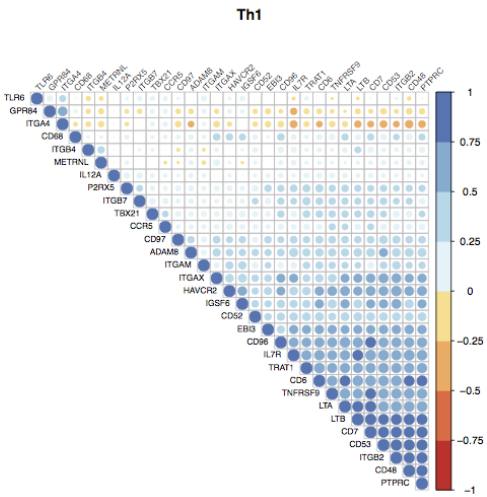


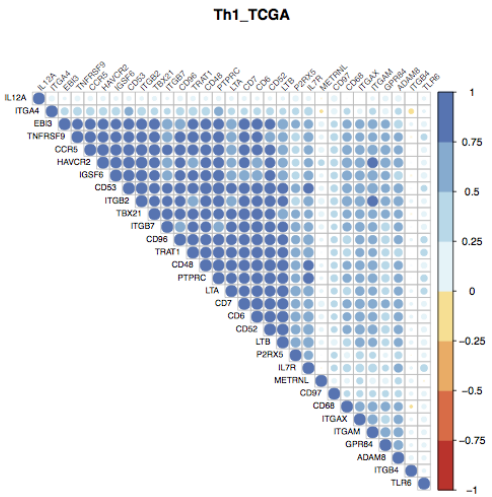




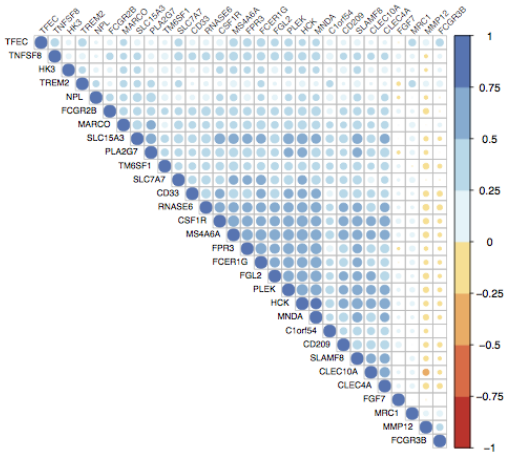

Th2
GATA3
SIGLEC10
UBASH3A
CCR2
SKAP1
LAMP3
RBMS3
HTR2B
ASB2
DNAJC12
TMPRSS3
1
0.75
0.5
0.25
0
−0.25
−0.5
−0.75
−1


Th2_TCGA
ASB2
TMPRSS3
SKAP1
LAMP3
GATA3
CCR2
UBASH3A
SIGLEC10
DNAJC12
RBMS3
HTR2B
1
0.75
0.5
0.25
0
−0.25
−0.5
−0.75
−1

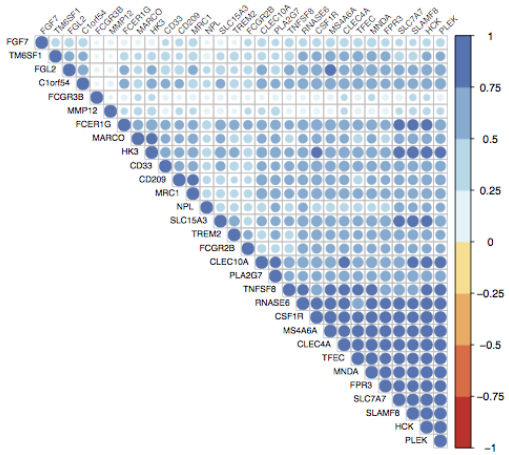

Th1
TLR6
GPR84
ITGA4
CD68
ITGB4
METRNL
IL12A
P2RX5
ITGB7
TBX21
CCR5
CD97
ADAM8
ITGAM
ITGAX
HAVCR2
IGSF6
CD52
EBI3
CD96
IL7R
TRAT1
CD6
TNFRSF9
LTA
LTB
CD7
CD53
ITGB2
CD48
PTPRC
1
0.75
0.5
0.25
0
−0.25
−0.5
−0.75
−1


Th1_TCGA
IL12A
ITGA4
EBI3
TNFRSF9
CCR5
HAVCR2
IGSF6
CD53
ITGB2
TBX21
ITGB7
CD96
TRAT1
CD48
PTPRC
LTA
CD7
CD6
CD52
LTB
P2RX5
IL7R
METRNL
CD97
CD68
ITGAX
ITGAM
GPR84
ADAM8
ITGB4
TLR6
1
0.75
0.5
0.25
0
−0.25
−0.5
−0.75
−1

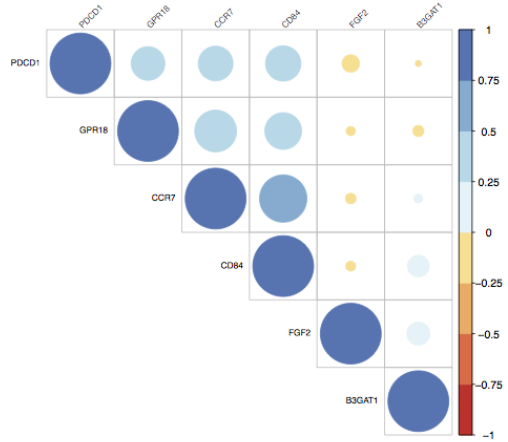
TGD



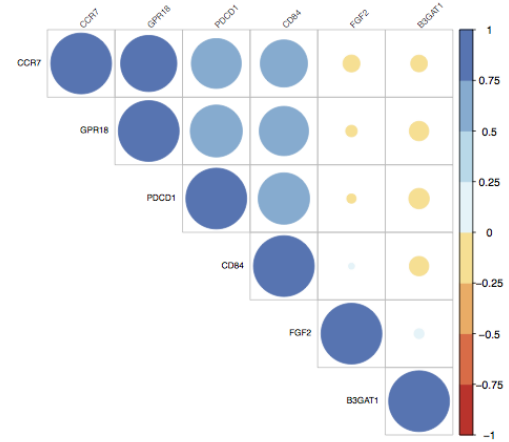
TGD_TCGA



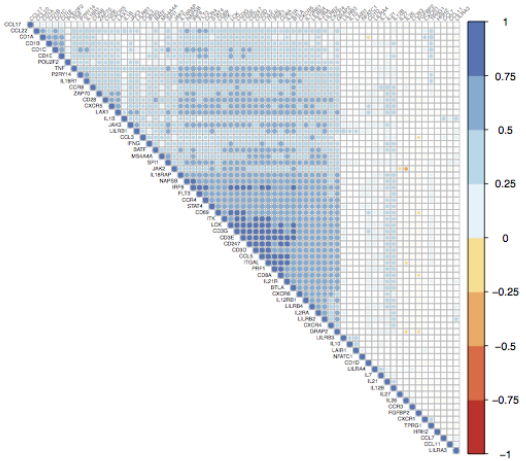
TFH



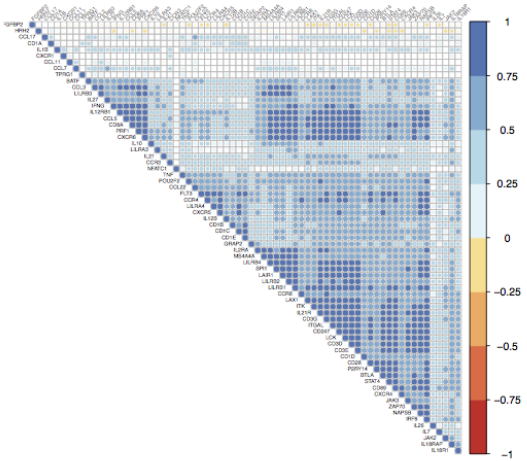
TFH_TCGA



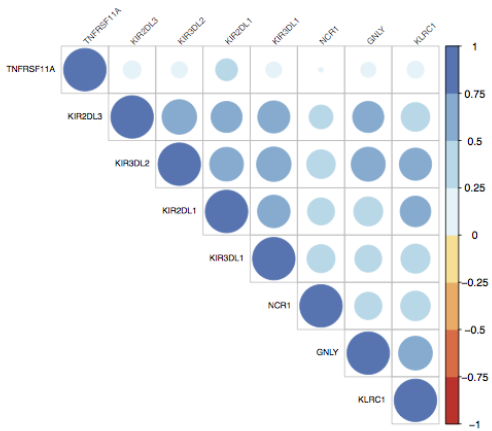
T_cells



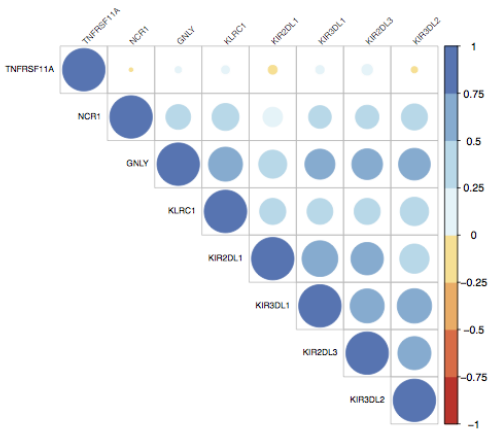
T_cells_TCGA

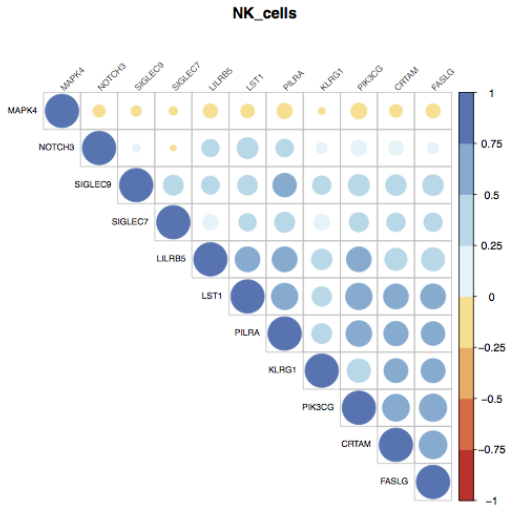


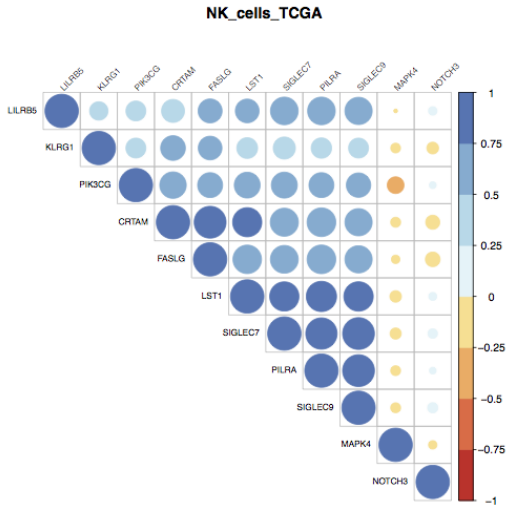
NKT

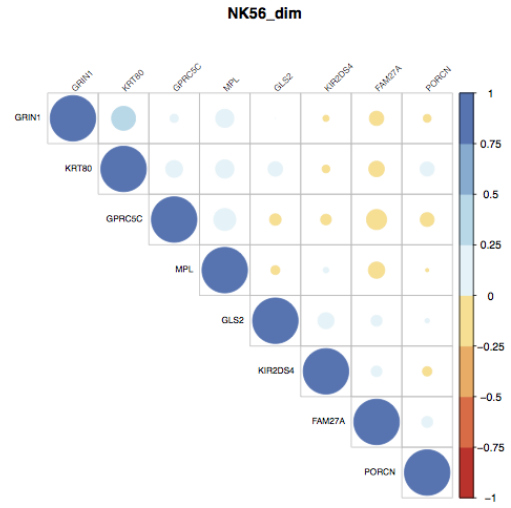


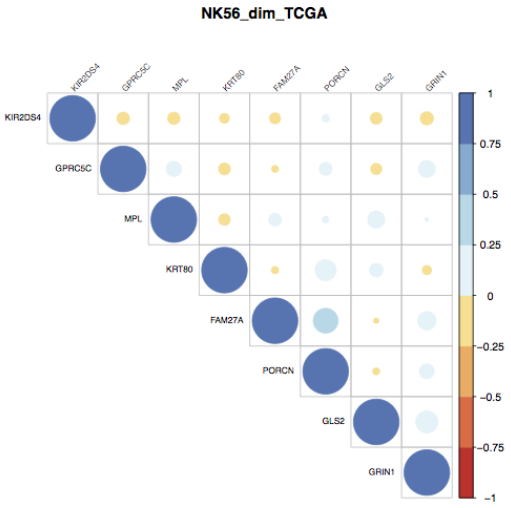
NKT_TCGA



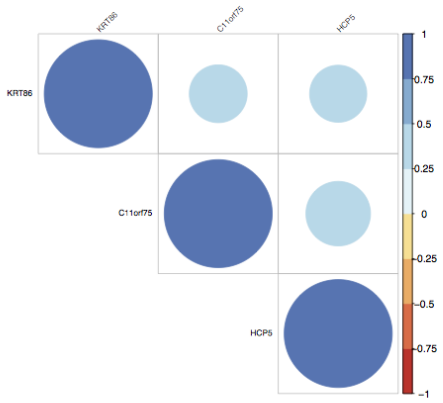

NK_cells
MAPK4
NOTCH3
SIGLEC9
SIGLEC7
LILRB5
LST1
PILRA
KLRG1
PIK3CG
CRTAM
FASLG
1
0.75
0.5
0.25
0
−0.25
−0.5
−0.75
−1


NK_cells_TCGA
LILRB5
KLRG1
PIK3CG
CRTAM
FASLG
LST1
SIGLEC7
PILRA
SIGLEC9
MAPK4
NOTCH3
1
0.75
0.5
0.25
0
−0.25
−0.5
−0.75
−1

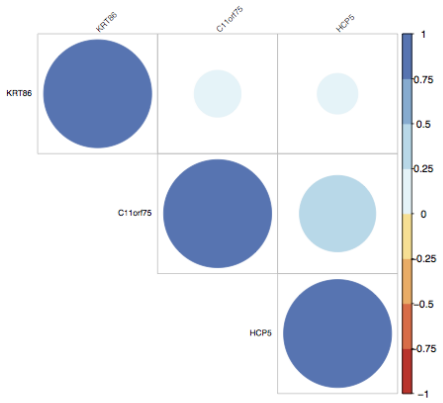

NK56_dim
GRIN1
KRT80
GPRC5C
MPL
GLS2
KIR2DS4
FAM27A
PORCN
1
0.75
0.5
0.25
0
−0.25
−0.5
−0.75
−1


NK56_dim_TCGA
KIR2DS4
GPRC5C
MPL
KRT80
FAM27A
PORCN
GLS2
GRIN1
1
0.75
0.5
0.25
0
−0.25
−0.5
−0.75
−1

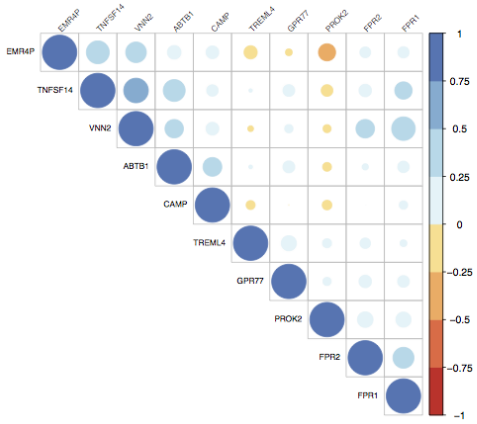
NK56_bright



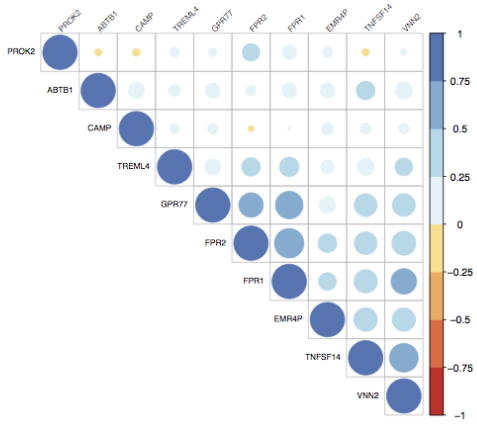
NK56_bright_TCGA

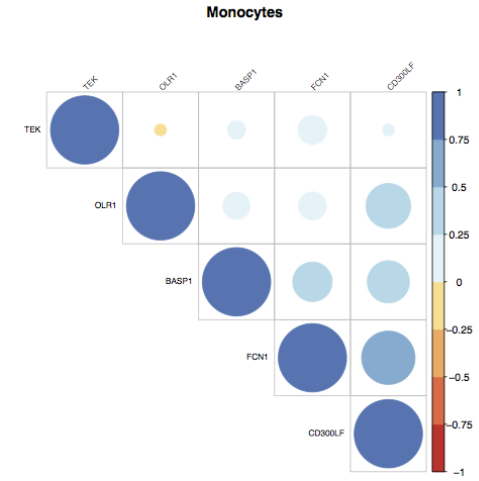


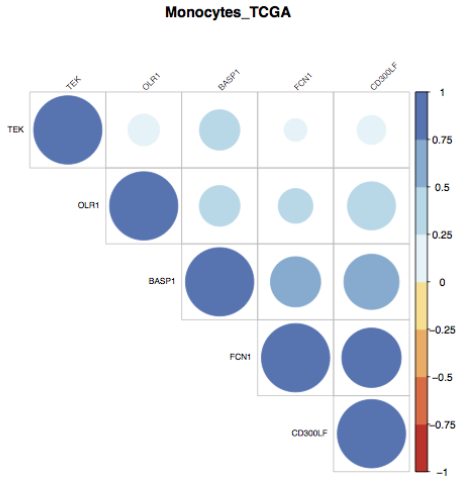
Neutrophils.pdf

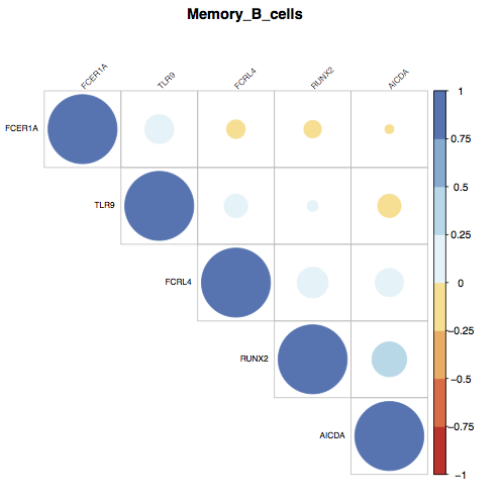


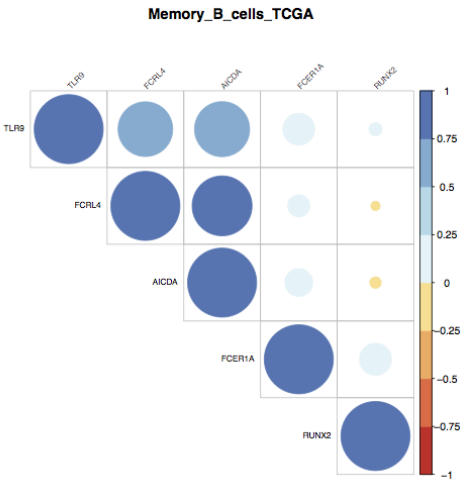
Neutrophils_TCGA.pdf



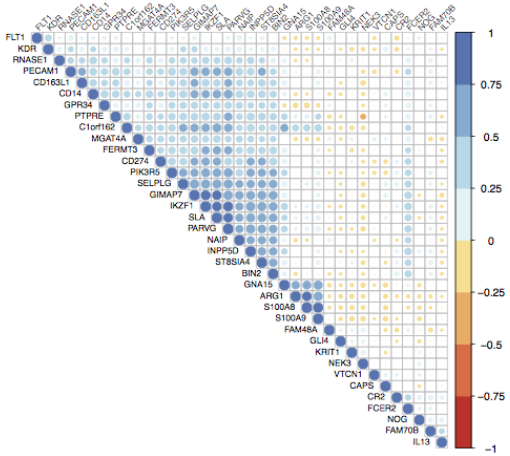

Monocytes
TEK
OLR1
BASP1
FCN1
CD300LF
1
0.75
0.5
0.25
0
−0.25
−0.5
−0.75
−1


Monocytes_TCGA
TEK
OLR1
BASP1
FCN1
CD300LF
1
0.75
0.5
0.25
0
−0.25
−0.5
−0.75
−1

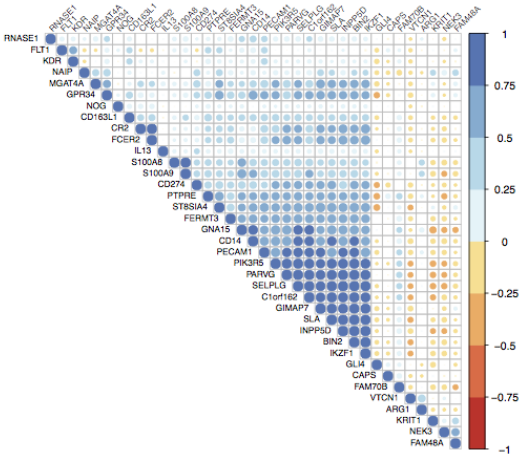

Memory_B_cells
FCER1A
TLR9
FCRL4
RUNX2
AICDA
1
0.75
0.5
0.25
0
−0.25
−0.5
−0.75
−1


Memory_B_cells_TCGA
TLR9
FCRL4
AICDA
FCER1A
RUNX2
1
0.75
0.5
0.25
0
−0.25
−0.5
−0.75
−1

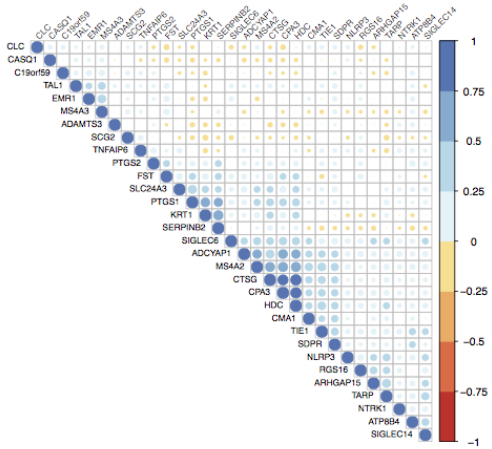
MDSC



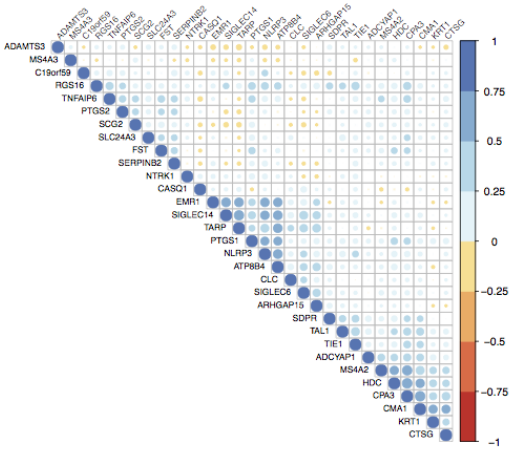
MDSC_TCGA



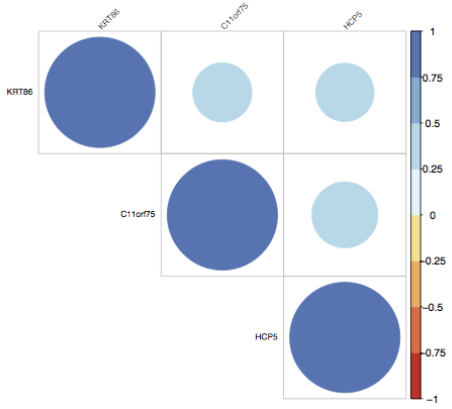
Mast_cells



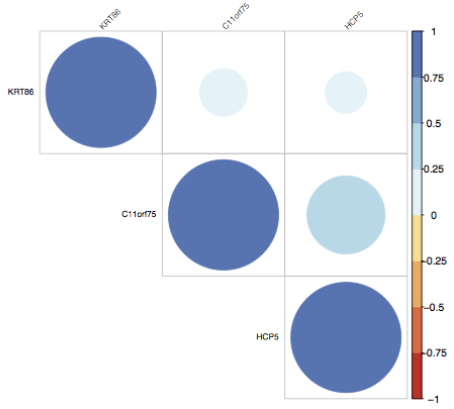
Mast_cells_TCGA



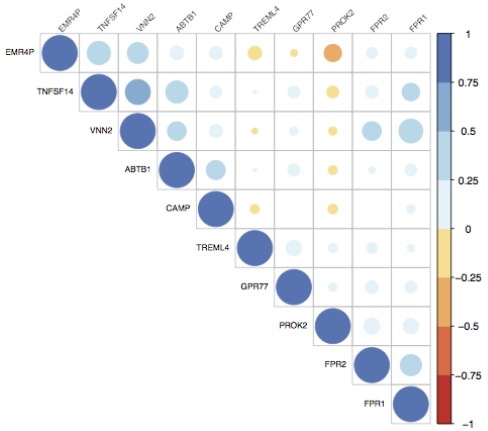
NK56_bright



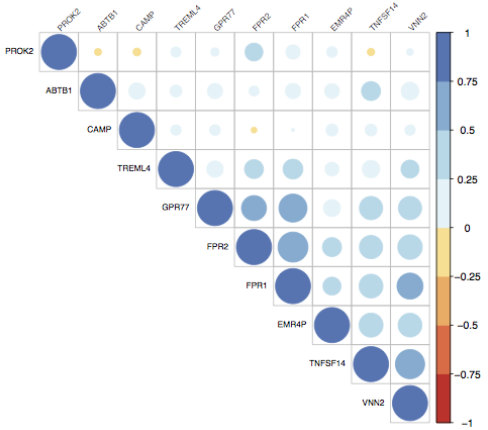
NK56_bright_TCGA



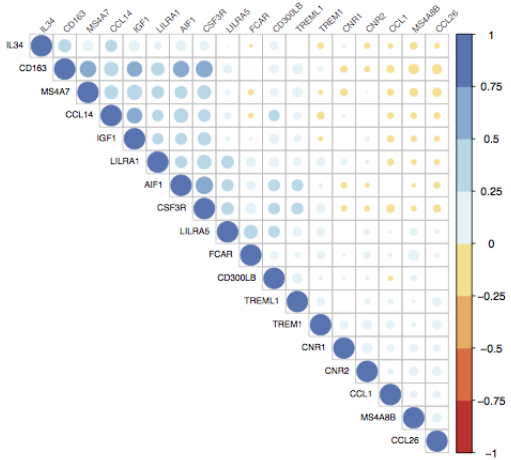
Neutrophils.pdf



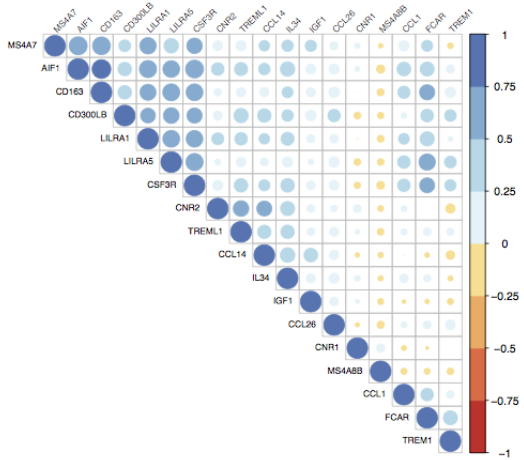
Neutrophils_TCGA.pdf



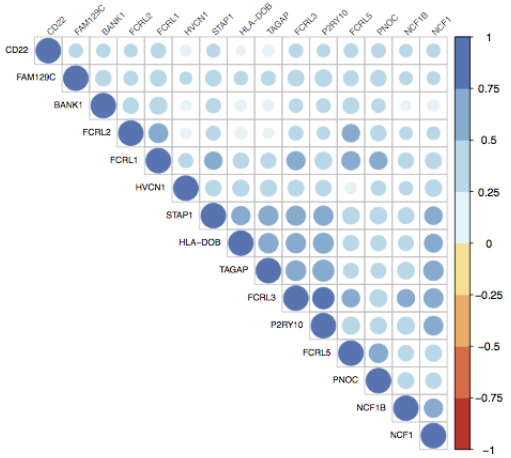
Macrophages



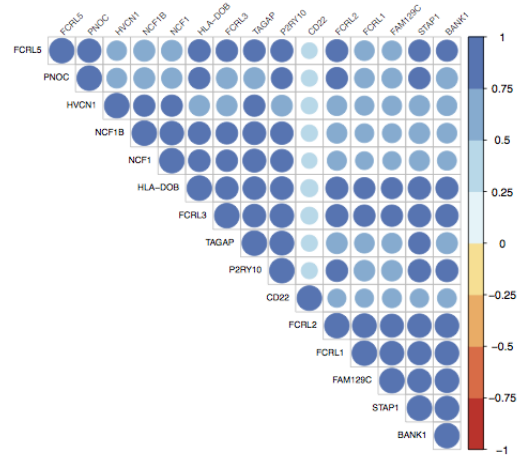
Macrophages_TCGA

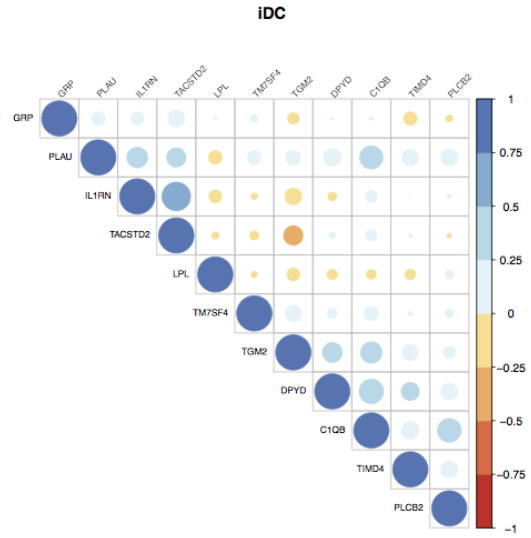


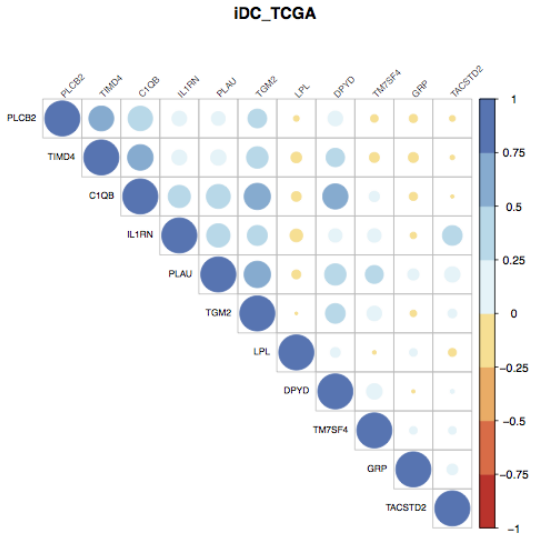
Immature_B_cells

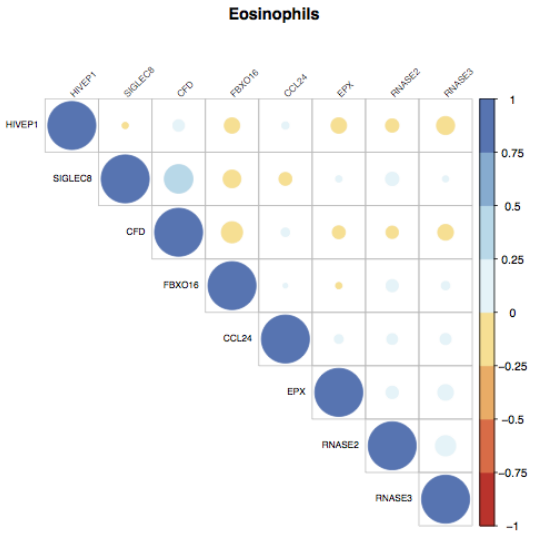


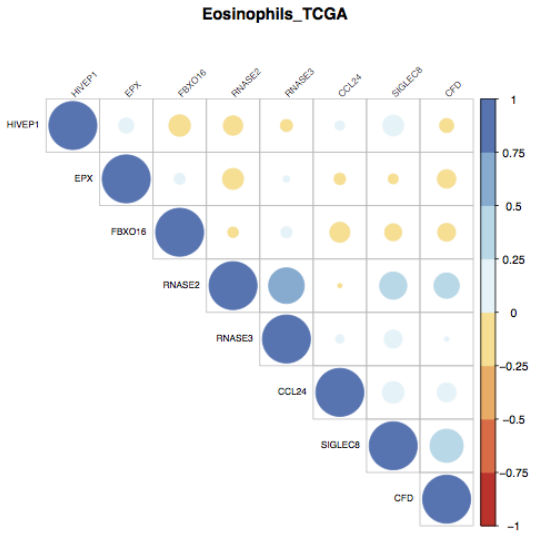
Immature_B_cells_TCGA



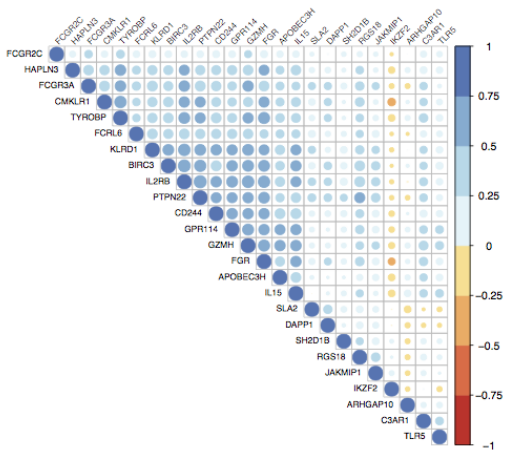

iDC
GRP
PLAU
IL1RN
TACSTD2
LPL
TM7SF4
TGM2
DPYD
C1QB
TIMD4
PLCB2
1
0.75
0.5
0.25
0
−0.25
−0.5
−0.75
−1


iDC_TCGA
PLCB2
TIMD4
C1QB
IL1RN
PLAU
TGM2
LPL
DPYD
TM7SF4
GRP
TACSTD2
1
0.75
0.5
0.25
0
−0.25
−0.5
−0.75
−1

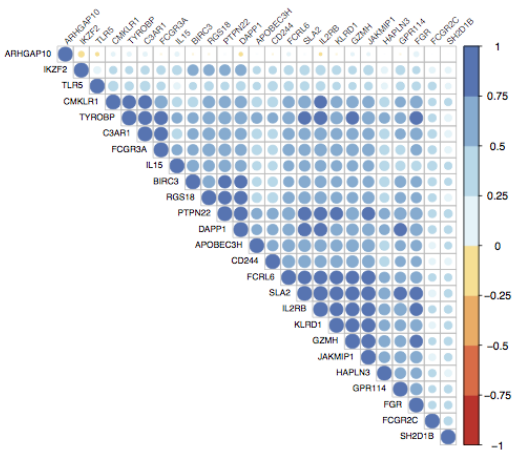

Eosinophils
HIVEP1
SIGLEC8
CFD
FBXO16
CCL24
EPX
RNASE2
RNASE3
1
0.75
0.5
0.25
0
−0.25
−0.5
−0.75
−1


Eosinophils_TCGA
HIVEP1
EPX
FBXO16
RNASE2
RNASE3
CCL24
SIGLEC8
CFD
1
0.75
0.5
0.25
0
−0.25
−0.5
−0.75
−1

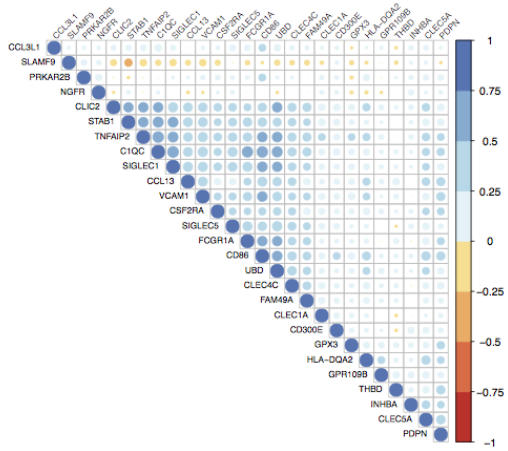
Effector_memory_CD8



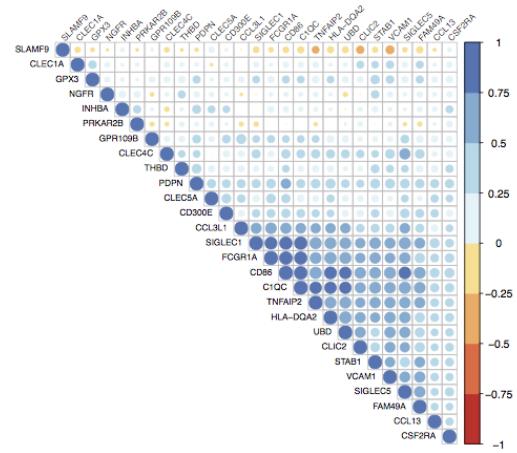
Effector_memory_CD8_TCGA



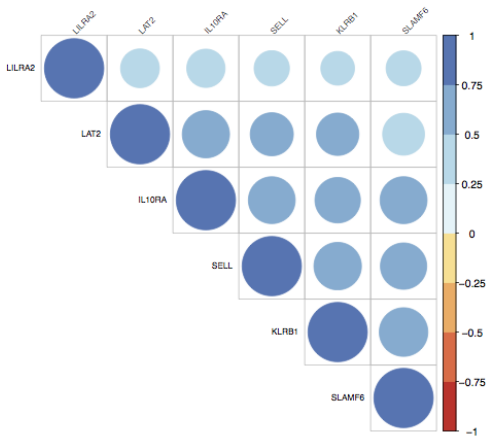
DC



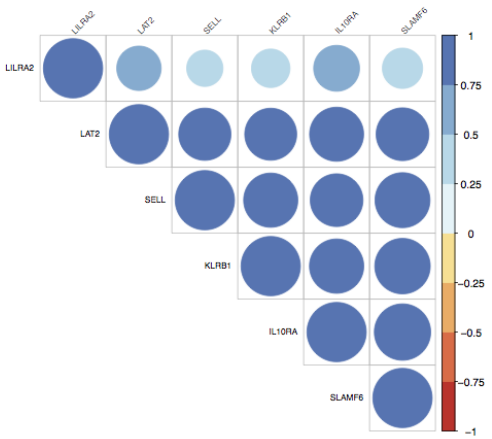
DC_TCGA



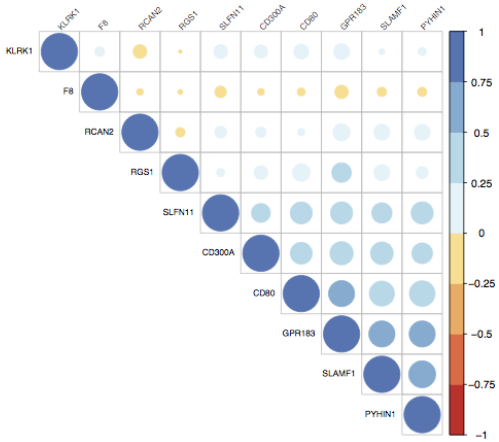
Cytotoxic_cells



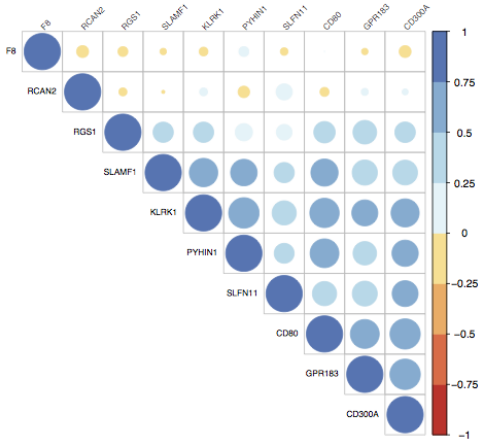
Cytotoxic_cells_TCGA

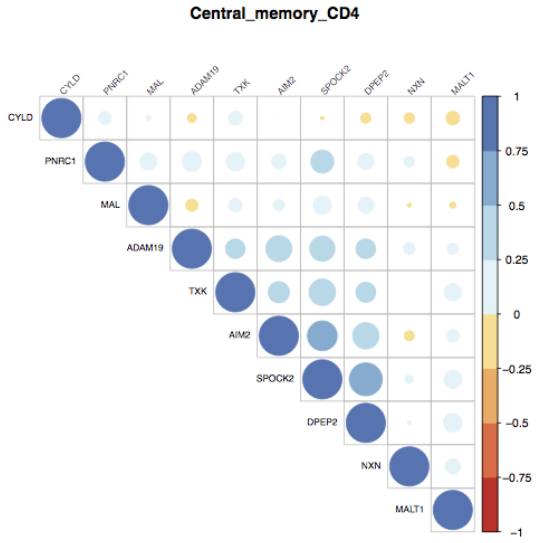


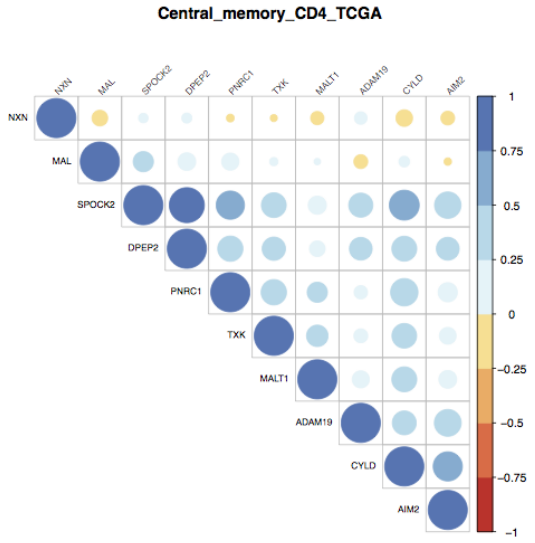
Central_memory_CD8



Central_memory_CD8_TCGA




Central_memory_CD4
CYLD
PNRC1
MAL
ADAM19
TXK
AIM2
SPOCK2
DPEP2
NXN
MALT1
1
0.75
0.5
0.25
0
−0.25
−0.5
−0.75
−1


Central_memory_CD4_TCGA
NXN
MAL
SPOCK2
DPEP2
PNRC1
TXK
MALT1
ADAM19
CYLD
AIM2
1
0.75
0.5
0.25
0
−0.25
−0.5
−0.75
−1

A.1.3 Final list of genes per immune cell scores after filtration process

Probe	*Gene*	*Cell_type*
ilmn_1777998	*ARHGAP25*	Activated_B_cells
ilmn_2068274	*CLEC9A*	Activated_B_cells
ilmn_1770673	*AKNA*	Activated_B_cells
ilmn_1688959	*CD27*	Activated_B_cells
ilmn_1780397	*TRAF3IP3*	Activated_B_cells
ilmn_1786176	*CD37*	Activated_B_cells
ilmn_2233783	*CD38*	Activated_B_cells
ilmn_1695025	*CD2*	Activated_B_cells
ilmn_1748473	*GIMAP4*	Activated_B_cells
ilmn_1687052	*PAX5*	Activated_B_cells
ilmn_1665943	*MAP4K1*	Activated_B_cells
ilmn_1677505	*CCL21*	Activated_B_cells
ilmn_1726597	*FAM65B*	Activated_B_cells
ilmn_1665647	*CD180*	Activated_B_cells
ilmn_1768016	*TNFRSF17*	Activated_B_cells
ilmn_1782729	*CLECL1*	Activated_B_cells
ilmn_2327860	*MAL*	Central_memory_CD4
ilmn_1689160	*DPEP2*	Central_memory_CD4
ilmn_1656287	*SPOCK2*	Central_memory_CD4
ilmn_1681301	*AIM2*	Central_memory_CD4
ilmn_1741143	*TXK*	Central_memory_CD4
ilmn_1791226	*NXN*	Central_memory_CD4
ilmn_1788481	*ADAM19*	Central_memory_CD4
ilmn_1730986	*MALT1*	Central_memory_CD4
ilmn_1789955	*PNRC1*	Central_memory_CD4
ilmn_1752520	*SLFN11*	Central_memory_CD8
ilmn_1716736	*CD80*	Central_memory_CD8
ilmn_1770768	*SLAMF1*	Central_memory_CD8
ilmn_2120210	*RCAN2*	Central_memory_CD8
ilmn_2395981	*PYHIN1*	Central_memory_CD8
ilmn_3235514	*GPR183*	Central_memory_CD8
ilmn_1693552	*CD300A*	Central_memory_CD8
ilmn_1656011	*RGS1*	Central_memory_CD8
ilmn_2222443	*KLRK1*	Central_memory_CD8
ilmn_2079655	*KLRB1*	Cytotoxic_cells
ilmn_1716983	*LILRA2*	Cytotoxic_cells
ilmn_1652825	*IL10RA*	Cytotoxic_cells
ilmn_2196078	*SLAMF6*	Cytotoxic_cells
ilmn_2326953	*LAT2*	Cytotoxic_cells
ilmn_1724422	*SELL*	Cytotoxic_cells
ilmn_1780465	*CLEC5A*	DC

Probe	*Gene*	*Cell_type*
ilmn_2176063	*FCGR1A*	DC
ilmn_1655987	*STAB1*	DC
ilmn_1751095	*CD300E*	DC
ilmn_1714602	*CD86*	DC
ilmn_1652381	*SIGLEC5*	DC
ilmn_1670490	*PDPN*	DC
ilmn_2307903	*VCAM1*	DC
ilmn_1783593	*CCL13*	DC
ilmn_1726666	*GPX3*	DC
ilmn_1752658	*NGFR*	DC
ilmn_1691339	*CLEC1A*	DC
ilmn_1759787	*THBD*	DC
ilmn_1794038	*FAM49A*	DC
ilmn_1719547	*INHBA*	DC
ilmn_1785902	*C1QC*	DC
ilmn_1727689	*TNFAIP2*	DC
ilmn_1795754	*CLIC2*	DC
ilmn_1682259	*CLEC4C*	DC
ilmn_1677693	*GPR109B*	DC
ilmn_1678841	*UBD*	DC
ilmn_1725320	*SIGLEC1*	DC
ilmn_1680144	*HLA-DQA2*	DC
ilmn_2376458	*CSF2RA*	DC
ilmn_2218856	*CCL3L1*	DC
ilmn_1684850	*PRKAR2B*	DC
ilmn_1787529	*C3AR1*	Effector_memory_CD8
ilmn_1722981	*TLR5*	Effector_memory_CD8
ilmn_2101278	*RGS18*	Effector_memory_CD8
ilmn_1715885	*PTPN22*	Effector_memory_CD8
ilmn_2409720	*SLA2*	Effector_memory_CD8
ilmn_1673030	*CMKLR1*	Effector_memory_CD8
ilmn_2097410	*DAPP1*	Effector_memory_CD8
ilmn_1664828	*APOBEC3H*	Effector_memory_CD8
ilmn_1684349	*IL2RB*	Effector_memory_CD8
ilmn_1724181	*IL15*	Effector_memory_CD8
ilmn_2368318	*FGR*	Effector_memory_CD8
ilmn_2405684	*BIRC3*	Effector_memory_CD8
ilmn_2074762	*FCRL6*	Effector_memory_CD8
ilmn_1702534	*CD244*	Effector_memory_CD8
ilmn_1701237	*SH2D1B*	Effector_memory_CD8
ilmn_1784141	*JAKMIP1*	Effector_memory_CD8
ilmn_1666902	*GPR114*	Effector_memory_CD8
ilmn_1654319	*HAPLN3*	Effector_memory_CD8

Probe	*Gene*	*Cell_type*
ilmn_1731233	*GZMH*	Effector_memory_CD8
ilmn_1778977	*TYROBP*	Effector_memory_CD8
ilmn_2112580	*FCGR3A*	Effector_memory_CD8
ilmn_2331121	*FCGR2C*	Effector_memory_CD8
ilmn_2386453	*KLRD1*	Effector_memory_CD8
ilmn_1730628	*RNASE2*	Eosinophil
ilmn_2113126	*RNASE3*	Eosinophil
ilmn_1653766	*CCL24*	Eosinophil
ilmn_1715823	*FBXO16*	Eosinophil
ilmn_1772631	*EPX*	Eosinophil
ilmn_1700428	*HLA-DOB*	Immature_B_cells
ilmn_1697309	*NCF1*	Immature_B_cells
ilmn_1792075	*CD22*	Immature_B_cells
ilmn_1815168	*HVCN1*	Immature_B_cells
ilmn_1664063	*FAM129C*	Immature_B_cells
ilmn_2333774	*TAGAP*	Immature_B_cells
ilmn_2106725	*NCF1B*	Immature_B_cells
ilmn_1699599	*FCRL3*	Immature_B_cells
ilmn_2105223	*FCRL5*	Immature_B_cells
ilmn_1661646	*BANK1*	Immature_B_cells
ilmn_2226183	*FCRL1*	Immature_B_cells
ilmn_1784774	*P2RY10*	Immature_B_cells
ilmn_3247998	*STAP1*	Immature_B_cells
ilmn_1676003	*PNOC*	Immature_B_cells
ilmn_1791329	*FCRL2*	Immature_B_cells
ilmn_1773963	*GNA15*	MDSC
ilmn_1681067	*PIK3R5*	MDSC
ilmn_1701947	*GPR34*	MDSC
ilmn_2394161	*ST8SIA4*	MDSC
ilmn_1701914	*CD274*	MDSC
ilmn_1744212	*INPP5D*	MDSC
ilmn_1662451	*FCER2*	MDSC
ilmn_2396444	*CD14*	MDSC
ilmn_1795183	*RNASE1*	MDSC
ilmn_1689518	*PECAM1*	MDSC
ilmn_2133316	*GIMAP7*	MDSC
ilmn_1686405	*KDR*	MDSC
ilmn_1760189	*NAIP*	MDSC
ilmn_1754894	*C1orf162*	MDSC
ilmn_3242540	*CD163L1*	MDSC
ilmn_2345898	*SLA*	MDSC
ilmn_2383611	*PTPRE*	MDSC
ilmn_1726342	*BIN2*	MDSC

Probe	*Gene*	*Cell_type*
ilmn_1752307	*FLT1*	MDSC
ilmn_1755643	*MGAT4A*	MDSC
ilmn_2366330	*FERMT3*	MDSC
ilmn_1734855	*SELPLG*	MDSC
ilmn_1676575	*IKZF1*	MDSC
ilmn_1695851	*PARVG*	MDSC
ilmn_1688231	*TREM1*	Macrophages
ilmn_1690783	*TREML1*	Macrophages
ilmn_2365091	*FCAR*	Macrophages
ilmn_2357419	*LILRA5*	Macrophages
ilmn_2379599	*CD163*	Macrophages
ilmn_1713686	*IL34*	Macrophages
ilmn_1734748	*LILRA1*	Macrophages
ilmn_3192001	*CCL14*	Macrophages
ilmn_1709613	*IGF1*	Macrophages
ilmn_2371280	*CSF3R*	Macrophages
ilmn_1792473	*AIF1*	Macrophages
ilmn_2331087	*MS4A7*	Macrophages
ilmn_1782741	*CD300LB*	Macrophages
ilmn_1715991	*SDPR*	Mast_cells
ilmn_2054297	*PTGS2*	Mast_cells
ilmn_1785732	*TNFAIP6*	Mast_cells
ilmn_1763344	*ADCYAP1*	Mast_cells
ilmn_2339835	*PTGS1*	Mast_cells
ilmn_2208413	*ARHGAP15*	Mast_cells
ilmn_1663519	*SLC24A3*	Mast_cells
ilmn_1808226	*RGS16*	Mast_cells
ilmn_2299095	*SIGLEC6*	Mast_cells
ilmn_1723944	*TARP*	Mast_cells
ilmn_3243061	*SIGLEC14*	Mast_cells
ilmn_1680424	*CTSG*	Mast_cells
ilmn_1783956	*ATP8B4*	Mast_cells
ilmn_1770772	*CMA1*	Mast_cells
ilmn_1700081	*FST*	Mast_cells
ilmn_1815690	*TIE1*	Mast_cells
ilmn_1806721	*MS4A2*	Mast_cells
ilmn_1766551	*CPA3*	Mast_cells
ilmn_1792323	*HDC*	Mast_cells
ilmn_1735712	*KRT1*	Mast_cells
ilmn_1731777	*NTRK1*	Mast_cells
ilmn_2310896	*NLRP3*	Mast_cells
ilmn_2377746	*RUNX2*	Memory_B_cells
ilmn_2164164	*AICDA*	Memory_B_cells

Probe	*Gene*	*Cell_type*
ilmn_1738517	*FCRL4*	Memory_B_cells
ilmn_1679798	*TLR9*	Memory_B_cells
ilmn_2066151	*TEK*	Monocytes
ilmn_1723035	*OLR1*	Monocytes
ilmn_1668063	*FCN1*	Monocytes
ilmn_2112357	*CD300LF*	Monocytes
ilmn_1651826	*BASP1*	Monocytes
ilmn_1658926	*NOTCH3*	NK
ilmn_2339294	*LILRB5*	NK
ilmn_1781824	*FASLG*	NK
ilmn_2409384	*SIGLEC7*	NK
ilmn_1770433	*PIK3CG*	NK
ilmn_1729915	*PILRA*	NK
ilmn_1795236	*SIGLEC9*	NK
ilmn_1658399	*KLRG1*	NK
ilmn_1796247	*CRTAM*	NK
ilmn_1688373	*LST1*	NK
ilmn_1803945	*HCP5*	NK56_bright
ilmn_1798270	*C11orf75*	NK56_bright
ilmn_2178226	*KRT86*	NK56_bright
ilmn_2352090	*GPRC5C*	NK56_dim
ilmn_1776640	*MPL*	NK56_dim
ilmn_1805404	*GRIN1*	NK56_dim
ilmn_1705814	*KRT80*	NK56_dim
ilmn_1698952	*TNFRSF11A*	NKT
ilmn_1667232	*KIR2DL3*	NKT
ilmn_2190842	*KIR3DL2*	NKT
ilmn_1786810	*KIR2DL1*	NKT
ilmn_1752647	*KLRC1*	NKT
ilmn_1750761	*NCR1*	NKT
ilmn_1790692	*GNLY*	NKT
ilmn_2131828	*KIR3DL1*	NKT
ilmn_1802096	*ABTB1*	Neutrophils
ilmn_2363392	*TNFSF14*	Neutrophils
ilmn_1678939	*VNN2*	Neutrophils
ilmn_1669317	*GPR77*	Neutrophils
ilmn_2092118	*FPR1*	Neutrophils
ilmn_1740875	*FPR2*	Neutrophils
ilmn_1688580	*CAMP*	Neutrophils
ilmn_3243190	*EMR4P*	Neutrophils
ilmn_1698367	*CD84*	TFH
ilmn_1715131	*CCR7*	TFH
ilmn_1780368	*GPR18*	TFH

Probe	*Gene*	*Cell_type*
ilmn_1806725	*PDCD1*	TFH
ilmn_2399363	*CLEC4A*	TGD
ilmn_1701248	*TREM2*	TGD
ilmn_1711332	*TFEC*	TGD
ilmn_2415303	*CLEC10A*	TGD
ilmn_2203271	*FPR3*	TGD
ilmn_1693009	*FGL2*	TGD
ilmn_1731503	*MARCO*	TGD
ilmn_2359800	*MS4A6A*	TGD
ilmn_2203926	*MRC1*	TGD
ilmn_1701195	*PLA2G7*	TGD
ilmn_1738992	*MNDA*	TGD
ilmn_1761778	*TNFSF8*	TGD
ilmn_1676372	*CD209*	TGD
ilmn_1686623	*CSF1R*	TGD
ilmn_1791771	*HCK*	TGD
ilmn_1782070	*NPL*	TGD
ilmn_2085862	*SLC15A3*	TGD
ilmn_1731240	*FGF7*	TGD
ilmn_1795762	*PLEK*	TGD
ilmn_1810275	*SLC7A7*	TGD
ilmn_1750961	*TM6SF1*	TGD
ilmn_2123743	*FCER1G*	TGD
ilmn_1670302	*HK3*	TGD
ilmn_1702231	*C1orf54*	TGD
ilmn_1747622	*CD33*	TGD
ilmn_1780533	*RNASE6*	TGD
ilmn_1667224	*SLAMF8*	TGD
ilmn_2382403	*FCGR2B*	TGD
ilmn_2342835	*P2RY14*	T_cells
ilmn_1790350	*TPRG1*	T_cells
ilmn_2370336	*MS4A4A*	T_cells
ilmn_1717197	*CD3G*	T_cells
ilmn_2086143	*CCR4*	T_cells
ilmn_2355953	*LILRB4*	T_cells
ilmn_1719433	*CD1D*	T_cells
ilmn_1668822	*BATF*	T_cells
ilmn_1772674	*IL21*	T_cells
ilmn_1721762	*IL18RAP*	T_cells
ilmn_1740633	*PRF1*	T_cells
ilmn_1674640	*CXCR6*	T_cells
ilmn_1749362	*CD28*	T_cells
ilmn_1710186	*CCL17*	T_cells

Probe	*Gene*	*Cell_type*
ilmn_1696463	*SPI1*	T_cells
ilmn_1781700	*IL18R1*	T_cells
ilmn_2389211	*LAIR1*	T_cells
ilmn_1723520	*CD1A*	T_cells
ilmn_1719756	*ZAP70*	T_cells
ilmn_1769782	*LAX1*	T_cells
ilmn_1728106	*TNF*	T_cells
ilmn_1736311	*POU2F2*	T_cells
ilmn_1739667	*JAK3*	T_cells
ilmn_2207291	*IFNG*	T_cells
ilmn_1749591	*ITGAL*	T_cells
ilmn_2377669	*CD247*	T_cells
ilmn_2109416	*NAPSB*	T_cells
ilmn_1651316	*CD69*	T_cells
ilmn_1722411	*HRH2*	T_cells
ilmn_1775501	*IL1B*	T_cells
ilmn_1798204	*IL21R*	T_cells
ilmn_1739421	*CCR8*	T_cells
ilmn_1726230	*CD1B*	T_cells
ilmn_1683774	*IL2RA*	T_cells
ilmn_2059744	*IL7*	T_cells
ilmn_2160476	*CCL22*	T_cells
ilmn_1785202	*STAT4*	T_cells
ilmn_1699908	*IL12RB1*	T_cells
ilmn_1725519	*CCL11*	T_cells
ilmn_2098126	*CCL5*	T_cells
ilmn_2384188	*NFATC1*	T_cells
ilmn_1666594	*IRF8*	T_cells
ilmn_2337928	*CXCR5*	T_cells
ilmn_2312340	*LILRB2*	T_cells
ilmn_1671509	*CCL3*	T_cells
ilmn_1681132	*IL12B*	T_cells
ilmn_1766363	*FLT3*	T_cells
ilmn_2406132	*LILRB3*	T_cells
ilmn_1699160	*ITK*	T_cells
ilmn_2377109	*LCK*	T_cells
ilmn_1778143	*GRAP2*	T_cells
ilmn_1683456	*CCL7*	T_cells
ilmn_2099528	*BTLA*	T_cells
ilmn_2335754	*CD1E*	T_cells
ilmn_1786303	*LILRA3*	T_cells
ilmn_2353732	*CD8A*	T_cells
ilmn_2320888	*CXCR4*	T_cells

Probe	*Gene*	*Cell_type*
ilmn_1654210	*CD1C*	T_cells
ilmn_2073307	*IL10*	T_cells
ilmn_2316974	*LILRB1*	T_cells
ilmn_1739794	*CD3E*	T_cells
ilmn_1683178	*JAK2*	T_cells
ilmn_2325837	*CD3D*	T_cells
ilmn_1702385	*LILRA4*	T_cells
ilmn_2061043	*CD48*	Th1
ilmn_2175912	*ITGB2*	Th1
ilmn_1792538	*CD7*	Th1
ilmn_1746565	*CD6*	Th1
ilmn_1677793	*P2RX5*	Th1
ilmn_1653498	*IGSF6*	Th1
ilmn_2359907	*CD68*	Th1
ilmn_2376205	*LTB*	Th1
ilmn_1671353	*IL12A*	Th1
ilmn_1777519	*ITGB7*	Th1
ilmn_2342066	*METRNL*	Th1
ilmn_2340217	*PTPRC*	Th1
ilmn_1802653	*EBI3*	Th1
ilmn_1693826	*HAVCR2*	Th1
ilmn_1813379	*TNFRSF9*	Th1
ilmn_1673363	*CD97*	Th1
ilmn_2415786	*CD96*	Th1
ilmn_1730176	*ITGAX*	Th1
ilmn_2145033	*CCR5*	Th1
ilmn_1708348	*ADAM8*	Th1
ilmn_1795464	*LTA*	Th1
ilmn_1692714	*TBX21*	Th1
ilmn_1662843	*CD53*	Th1
ilmn_1685009	*ITGAM*	Th1
ilmn_2342579	*IL7R*	Th1
ilmn_2334210	*ITGB4*	Th1
ilmn_1684943	*TRAT1*	Th1
ilmn_2208903	*CD52*	Th1
ilmn_1743570	*CEACAM3*	Th17
ilmn_1715603	*IL23A*	Th17
ilmn_1769925	*C2CD4A*	Th17
ilmn_1749744	*SH2D6*	Th17
ilmn_1774983	*IL17A*	Th17
ilmn_2227195	*CCDC65*	Th17
ilmn_2043079	*ILDR1*	Th17
ilmn_2188247	*IL17F*	Th17

Probe	*Gene*	*Cell_type*
ilmn_1788109	*IL17C*	Th17
ilmn_1761766	*GPR25*	Th17
ilmn_1655549	*SIGLEC10*	Th2
ilmn_2338348	*UBASH3A*	Th2
ilmn_2170813	*LAMP3*	Th2
ilmn_1675677	*TMPRSS3*	Th2
ilmn_2406656	*GATA3*	Th2
ilmn_1751400	*SKAP1*	Th2
ilmn_2376431	*CCR2*	Th2
ilmn_2125017	*TIGIT*	Treg
ilmn_1769129	*CCL19*	Treg
ilmn_1788531	*SIT1*	Treg
ilmn_1739393	*SELE*	Treg
ilmn_1698218	*TRAF1*	Treg
ilmn_1677827	*TLR7*	Treg
ilmn_1796642	*NCF2*	Treg
ilmn_2313672	*IL1RL1*	Treg
ilmn_1723004	*CD72*	Treg
ilmn_1715417	*SELP*	Treg
ilmn_2414762	*TLR10*	Treg
ilmn_1749006	*RCSD1*	Treg
ilmn_2341229	*CD34*	Treg
ilmn_2129668	*TGFB1*	Treg
ilmn_1763487	*CTLA4*	Treg
ilmn_1772387	*TLR2*	Treg
ilmn_1705047	*TLR8*	Treg
ilmn_1793730	*TM7SF4*	iDC
ilmn_1750678	*TIMD4*	iDC
ilmn_1797236	*TGM2*	iDC
ilmn_1795715	*DPYD*	iDC
ilmn_1796409	*C1QB*	iDC
ilmn_1724066	*PLCB2*	iDC
ilmn_1656057	*PLAU*	iDC
ilmn_1747344	*IL3RA*	pDC

A.2 Enriched pathways

A.2.1 Top enriched pathways in the Low Immune Subgroup

"GeneSet" is the pathway name, "Ratio of protein in Geneset" indicates ratios of numbers of genes contained in pathways to total genes in the Reactome FI network; "Number of protein in GeneSet" is the number of proteins within a given pathway; and "Protein from network" indicates number of genes from the input gene list per pathway. "Nodes" are proteins from the network.

GeneSet	*Ratio Of Protein In GeneSet*	*Number Of Protein In GeneSet*	*Protein From Network*	*P-value*	*FDR*	*Nodes*
The citric acid (TCA) cycle and respiratory electron transport(R)	0.0148	161	66	1.11E-16	5.27E-14	ATP5C1, UQCRC1, UQCRC2, ECSIT, DLD, ATP5A1, SLC16A1, UQCRB, UQCRH, UQCRQ, NDUFC1, NDUFB9, NDUFB7, NDUFB3, NDUFB2, NDUFA8, NDUFA7, NDUFA4, NDUFA1, SUCLA2, BSG, ATP5S, ATP5J, ATP5H, ATP5O, ATP5E, ATP5D, NDUFAF2, NDUFAF1, COX7B, COX7C, COX8A, COX5B, COX5A, COX6C, NDUFAB1, TACO1, COX6A1, SDHC, COX6B1, NDUFB11, COX11, CYC1, PDHA1, CYCS, UQCRFS1, NDUFS5, NDUFS4, NDUFS2, NDUFS1, LRPPRC, IDH3A, PDK2, PDHX, COX7A2L, SUCLG1, TRAP1, DLAT, PDP2, NDUFA13, NDUFA11, NDUFA10, ATP5G3, ATP5G2, ATP5G1, ATP5F1
Mitochondrial translation(R)	0.0085	92	58	1.11E-16	5.27E-14	MRPL19, MRPL16, MRPL14, MRPL15, MRPL12, MRPL13, MRPL10, MRPL11, MRPL27, MRPL28, MRPL23, MRPL24, MRPL21, MRPL30, MRPL36, MRPL34, MRPL35, MRPL32, MRPL33, MRPL42, MRPL40, MRPL48, MRPL45, MRPL46, MRPL43, MRPL53, MRPL51, MRPL4, MRPL3, MRPL2, MRPL1,

GeneSet	*Ratio Of Protein In GeneSet*	*Number Of Protein In GeneSet*	*Protein From Network*	*P-value*	*FDR*	*Nodes*
						MRPL9, GADD45GIP1, MRPS17, MRPS15, MRPS28, MRPS26, MRPS27, MRPS24, MRPS23, MRPS2, MRPS21, MRPS7, MRPS5, MRPS9, MRPS35, MRPS36, MRPS33, MRPS34, MRPS30, MRPS18B, MRPS18A, GFM2, MRRF, TUFM, DAP3, TSFM, MTIF2
Ribosome(K)	0.0142	154	59	1.32E-14	4.17E-12	MRPL19, MRPL16, MRPL14, MRPL15, MRPL12, MRPL13, MRPL10, MRPL11, MRPL27, MRPL28, MRPL23, MRPL24, MRPL21, MRPL30, RPLP1, RPLP0, MRPL36, MRPL34, MRPL35, MRPL32, MRPL33, MRPL4, MRPL3, MRPL2, MRPL1, MRPL9, RPL23A, MRPS17, MRPS15, MRPS2, MRPS21, MRPS7, MRPS5, MRPS9, RPL26L1, RPL4, RPL32, RPL34, RPL8, RPL6, RPL7, RPL36, RPL39, RPL22, RPL29, RPL12, RPS2, RPL10A, RPS7, RPS5, RPL7A, RPL37A, RPL36A, RPL35A, RPS15, RPS13, MRPS18A, RPS21, RPS3A
Generic Transcription Pathway(R)	0.0455	494	121	9.03E-14	2.14E-11	CSNK2A1, CSNK2A2, NR2F6, ZNF45, MED20, PMS2, ATAD2, CENPJ, STK11, CDC25C, MSH2, SMYD2, MNAT1, EXO1, RABGGTB, BIRC5, E2F1, E2F7, NDUFA4, TRIM28, CSNK2B, KCTD15, SUPT4H1, WWTR1, COX7B, BRCA1, COX7C, HNF4A, COX8A, KIT, YAP1, PCNA, COX5B, COX5A, TCEB2, TCEB1, NPM1, COX6C, RPA3, TACO1, COX6A1, COX6B1, CCNB1, COX11, TFAP2A, SUPT5H, CYCS, YWHAE, CHD4, TRIAP1, YWHAQ, YWHAG, TGIF1, TAF9, AURKB, AURKA, TGS1, SSRP1, ERCC3, ERCC2, TAF4, TAF2, GLS, MYC, PIP4K2B, MED1, NEDD4L, PRDX5, PRELID1, PLK2, GPI, PHF20, ING2, TEAD1, TEAD4, GTF2H2, GTF2H4, LRPPRC, YEATS4, USP7, SUPT16H, GTF2F1, GTF2F2, FANCI, CDK8, CDK5, RHEB, CDK2, RFC5, RFC3, RFC4, TFDP2, POLR2C, POLR2D, RBBP8,

GeneSet	Ratio Of Protein In GeneSet	Number Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
						POLR2E, POLR2F, RBBP7, POLR2I, TP53RK, KAT5, CHEK1, COX7A2L, RRAGB, RRAGD, DNA2, TPX2, RPTOR, PRMT5, PRMT1, SGK1, NR6A1, CDK5R1, KCTD1, TBL1XR1, SMAD2, KDM5B, THRA, HSPD1, ERBB2, BAX
rRNA processing(R)	0.0165	179	62	2.15E-13	4.08E-11	RPLP1, RPLP0, TRMT112, NSUN4, BYSL, TSR1, ISG20L2, HSD17B10, NOP56, NOP58, RPL23A, NOL11, UTP14A, EXOSC7, EXOSC5, EXOSC4, PES1, EXOSC8, BOP1, DKC1, RPP21, RPP25, DDX49, UTP15, UTP18, RPL4, RPL32, RPL34, RPL8, RPL6, RRP9, RPL7, RPL36, RPL39, RPL22, RPL29, GAR1, PNO1, RIOK2, RPL12, RPS2, WDR18, WDR12, RPL10A, RPS7, RPS5, NHP2, WDR3, RPL7A, NOB1, RPL37A, RPL36A, RPL35A, PWP2, RPS15, RPS13, RPS21, NOL9, EMG1, RPS3A, SENP3, SKIV2L2
Parkinson's disease(K)	0.0131	142	53	7.80E-13	1.23E-10	ATP5C1, UQCRC1, UQCRC2, GPR37, ATP5A1, UQCRB, UQCRH, UQCRQ, NDUFC1, SLC25A5, SLC25A4, NDUFB9, NDUFB7, NDUFB3, NDUFB2, NDUFA8, NDUFA7, NDUFA4, NDUFA1, ATP5J, ATP5H, ATP5O, ATP5E, ATP5D, COX7B, COX7C, COX8A, COX5B, COX5A, COX6C, NDUFAB1, COX6A1, SDHC, COX6B1, NDUFB11, CYC1, CYCS, UQCRFS1, NDUFS5, NDUFS4, NDUFS2, NDUFS1, COX7A2L, VDAC2, VDAC1, NDUFA13, NDUFA11, NDUFA10, ATP5G3, ATP5G2, ATP5G1, ATP5F1, SNCA
Mitotic Prometaphase(R)	0.0091	99	43	1.00E-12	1.36E-10	CSNK2A1, CSNK2A2, CDCA5, NCAPG, CDCA8, CENPA, CENPF, APITD1, CENPI, CENPL, CENPM, CENPN, CENPO, CENPQ, NDC80, BIRC5, CSNK2B, KIF2C, MAD2L1, ERCC6L, BUB1B, CCNB2, CCNB1, NUP133, AURKB, SMC4, SEH1L, SKA1, RAD21, PLK1, NUF2,

GeneSet	*Ratio Of Protein In GeneSet*	*Number Of Protein In GeneSet*	*Protein From Network*	*P-value*	*FDR*	*Nodes*
						HDAC8, ZW10, ZWINT, AHCTF1, XPO1, RANBP2, KIF18A, CKAP5, ITGB3BP, BUB1, CLASP1, CLASP2
Mitotic Metaphase and Anaphase(R)	0.0151	164	57	1.76E-12	2.08E-10	PSMD8, PSMD4, PSMD2, PSMD3, PSME3, CDCA5, CDCA8, CENPA, PSMB7, PSMB4, PSMB5, CENPF, APITD1, CENPI, CENPL, CENPM, CENPN, CENPO, CENPQ, ANAPC5, NDC80, BIRC5, CDC16, KIF2C, MAD2L1, ERCC6L, BUB1B, BANF1, PSMD10, PSMD12, PSMD11, PSMD14, PSMD13, UBE2C, NUP133, AURKB, SEH1L, SKA1, ANAPC11, RAD21, PLK1, NUF2, ESPL1, HDAC8, PPP2R1A, ZW10, ZWINT, AHCTF1, XPO1, RANBP2, PTTG1, KIF18A, CKAP5, ITGB3BP, BUB1, CLASP1, CLASP2
Oxidative phosphorylation(K)	0.0122	133	50	2.69E-12	2.78E-10	ATP5C1, UQCRC1, UQCRC2, ATP5A1, UQCRB, UQCRH, UQCRQ, NDUFC1, NDUFB9, NDUFB7, NDUFB3, NDUFB2, NDUFA8, NDUFA7, NDUFA4, NDUFA1, ATP6V1E2, ATP5J, ATP5H, ATP5O, ATP5E, ATP5D, ATP6V1C1, COX7B, COX7C, COX8A, COX5B, COX5A, COX6C, NDUFAB1, COX6A1, SDHC, COX6B1, NDUFB11, COX11, CYC1, UQCRFS1, NDUFS5, NDUFS4, NDUFS2, NDUFS1, COX7A2L, NDUFA13, NDUFA11, NDUFA10, ATP6V1D, ATP5G3, ATP5G2, ATP5G1, ATP5F1
Alzheimer's disease(K)	0.0157	171	58	2.92E-12	2.78E-10	ATP5C1, UQCRC1, UQCRC2, EIF2AK3, GRIN2C, ATP5A1, UQCRB, UQCRH, NAE1, UQCRQ, NDUFC1, NDUFB9, NDUFB7, NDUFB3, NDUFB2, HSD17B10, NDUFA8, NDUFA7, NDUFA4, NDUFA1, ATP5J, ATP5H, ATP5O, APH1B, ATP5E, ATP5D, COX7B, COX7C, COX8A, COX5B, COX5A, COX6C, NDUFAB1, COX6A1, SDHC, COX6B1, NDUFB11, CYC1, CYCS, GSK3B, UQCRFS1, NDUFS5, NDUFS4, NDUFS2, NDUFS1, PPP3R1, BACE2, CDK5, COX7A2L,

GeneSet	*Ratio Of Protein In GeneSet*	*Number Of Protein In GeneSet*	*Protein From Network*	*P-value*	*FDR*	*Nodes*
						CDK5R1, NDUFA13, NDUFA11, NDUFA10, ATP5G3, ATP5G2, ATP5G1, ATP5F1, SNCA
Huntington's disease(K)	0.0178	193	62	4.81E-12	4.14E-10	ATP5C1, UQCRC1, UQCRC2, PPARGC1A, ATP5A1, UQCRB, UQCRH, UQCRQ, NDUFC1, SLC25A5, SLC25A4, NDUFB9, NDUFB7, NDUFB3, NDUFB2, NDUFA8, NDUFA7, NDUFA4, NDUFA1, ATP5J, ATP5H, ATP5O, ATP5E, ATP5D, RCOR1, COX7B, COX7C, COX8A, COX5B, COX5A, COX6C, NDUFAB1, COX6A1, SDHC, COX6B1, NDUFB11, CYC1, CYCS, UQCRFS1, NDUFS5, NDUFS4, NDUFS2, NDUFS1, TAF4, CREB3L4, CLTC, POLR2C, POLR2D, POLR2E, POLR2F, POLR2I, COX7A2L, VDAC2, VDAC1, NDUFA13, NDUFA11, NDUFA10, ATP5G3, ATP5G2, ATP5G1, ATP5F1, BAX
Cell Cycle Checkpoints(R)	0.0151	164	52	3.95E-10	3.12E-08	PSMD8, PSMD4, PSMD2, PSMD3, GTSE1, PSME3, PSMB7, PSMB4, PSMB5, CDC25C, CDC25A, ANAPC5, PKMYT1, MCM10, EXO1, HERC2, BRCC3, CDC16, BRCA1, RPA3, MAD2L1, BUB1B, UBE2V2, PSMD10, PSMD12, PSMD11, PSMD14, PSMD13, CCNB2, CCNB1, UBE2C, YWHAE, YWHAQ, YWHAG, RNF8, ANAPC11, CDC7, CDC6, CDK2, RFC5, RFC3, RFC4, MCM3, MCM4, RBBP8, KAT5, CHEK1, DNA2, TP53BP1, CLSPN, H2AFX, WEE1
Eukaryotic Translation Initiation(R)	0.0103	112	41	5.25E-10	3.83E-08	RPLP1, RPLP0, EIF1AX, EIF4EBP1, EIF5B, EIF4B, RPL23A, EIF3M, EIF3K, EIF3I, EIF3H, EIF3F, EIF3D, PABPC1, RPL4, RPL32, RPL34, RPL8, RPL6, RPL7, RPL36, RPL39, RPL22, RPL29, RPL12, RPS2, RPL10A, RPS7, RPS5, EIF2B3, EIF2B2, RPL7A, RPL37A, RPL36A, RPL35A, RPS15, RPS13, RPS21, EIF2S2, EIF2S3, RPS3A

GeneSet	Ratio Of Protein In GeneSet	Number Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
Non-alcoholic fatty liver disease (NAFLD)(K)	0.0139	151	48	1.71E-09	1.15E-07	AKT2, AKT3, UQCRC1, UQCRC2, EIF2AK3, PKLR, UQCRB, UQCRH, UQCRQ, PIK3CB, NDUFC1, NDUFB9, NDUFB7, NDUFB3, NDUFB2, NDUFA8, NDUFA7, NDUFA4, NDUFA1, PIK3R2, COX7B, COX7C, COX8A, COX5B, COX5A, COX6C, NDUFAB1, COX6A1, SDHC, COX6B1, NDUFB11, CYC1, CYCS, GSK3B, UQCRFS1, NDUFS5, NDUFS4, NDUFS2, NDUFS1, IRS2, MLXIP, COX7A2L, NDUFA13, NDUFA11, NDUFA10, ADIPOR1, ADIPOR2, BAX
S Phase(R)	0.0113	123	42	2.40E-09	1.51E-07	PSMD8, PSMD4, PSMD2, PSMD3, PSME3, CDCA5, PSMB7, PSMB4, PSMB5, CDC25A, MNAT1, POLA1, PCNA, RPA3, PSMD10, PSMD12, PSMD11, PSMD14, PSMD13, POLD2, POLE4, POLE2, MYC, RAD21, CDC6, CDK4, CDK2, RFC5, RFC3, RFC4, MCM3, MCM4, SKP2, ESCO1, DNA2, FEN1, GINS1, GINS2, GINS3, LIG1, CKS1B, WEE1
Mitotic G2-G2/M phases(R)	0.016	174	52	2.81E-09	1.66E-07	PSMD8, PSMD4, PSMD2, PSMD3, GTSE1, PSME3, DYNLL1, PSMB7, PSMB4, PSMB5, CENPF, CENPJ, NEK2, CDC25C, CDC25A, PKMYT1, TUBGCP5, TUBGCP4, MNAT1, HSP90AB1, FOXM1, E2F1, CETN2, DYNC1I2, DYNC1H1, PSMD10, PSMD12, PSMD11, PSMD14, PSMD13, CCNB2, CCNB1, YWHAE, YWHAG, CCNA2, AURKA, CEP70, CEP76, PLK4, PLK1, PHLDA1, FKBPL, HMMR, CDK2, CEP152, PPP2R1A, TPX2, XPO1, TUBG1, CKAP5, CLASP1, WEE1
Cell cycle(K)	0.0114	124	40	2.59E-08	1.43E-06	CDC25C, CDC25A, ANAPC5, PKMYT1, E2F1, CDC16, PCNA, MAD2L1, BUB1B, CCNE2, CCNE1, CCNB3, CCNB2, CCNB1, YWHAE, GSK3B, YWHAQ, YWHAG, CCNA2, MYC, ANAPC13, PRKDC, ANAPC11, RAD21, PLK1, CDC7, CDC6, CDK4, CDK2,

GeneSet	*Ratio Of Protein In GeneSet*	*Number Of Protein In GeneSet*	*Protein From Network*	*P-value*	*FDR*	*Nodes*
						TFDP2, ESPL1, MCM3, MCM4, TTK, CHEK1, SKP2, PTTG1, SMAD2, BUB1, WEE1
Mitotic G1-G1/S phases(R)	0.012	130	40	8.81E-08	4.58E-06	PSMD8, PSMD4, PSMD2, PSMD3, PSME3, DHFR, PSMB7, PSMB4, PSMB5, CDC25A, MNAT1, MCM10, E2F1, POLA1, PCNA, RPA3, CCNE2, CCNE1, PSMD10, PSMD12, PSMD11, PSMD14, PSMD13, CCNB1, RRM2, POLE4, POLE2, MYC, CDC7, CDC6, CDK4, CDK2, TFDP2, MCM3, MCM4, SKP2, TOP2A, TYMS, CKS1B, WEE1
Nucleotide Excision Repair(R)	0.0094	102	34	1.34E-07	6.71E-06	ACTL6A, DDB1, RUVBL1, INO80C, MNAT1, CETN2, PCNA, RPA3, UBE2V2, PPIE, POLD2, ERCC3, ERCC1, ERCC2, ERCC8, GTF2H2, GTF2H4, USP7, CHD1L, COPS4, COPS5, COPS2, COPS8, RFC5, RFC3, RFC4, POLR2C, POLR2D, POLR2E, POLR2F, POLR2I, PRPF19, LIG1, LIG3

A.2.2 Top enriched pathways in the *CTNNB1* dominant group in the whole dataset

"GeneSet" is the pathway name, "Ratio of protein in Geneset" indicates ratios of numbers of genes contained in pathways to total genes in the Reactome FI network; "Number of protein in GeneSet" is the number of proteins within a given pathway; and "Protein from network" indicates number of genes from the input gene list per pathway. "Nodes" are proteins from the network.

GeneSet	*Ratio Of Protein In GeneSet*	*Number Of Protein In GeneSet*	*Protein From Network*	*P-value*	*FDR*	*Nodes*
Melanogenesis(K)	0.0093	101	3	3.36E-04	0.0221	GNAI2, CTNNB1, WNT2
role of brca1 brca2 and atr in cancer susceptibility(B)	0.0018	20	2	3.50E-04	0.0221	RAD1, ATR
Generic Transcription Pathway(R)	0.0455	494	5	3.97E-04	0.0221	PPP2R5C, POLR2C, RAD1, TAF9B, ATR
p53 pathway feedback loops 2(P)	0.0022	24	2	5.02E-04	0.0221	CTNNB1, ATR
Tight junction(K)	0.0126	137	3	8.14E-04	0.026	GNAI2, PARD6A, CTNNB1
Transcriptional regulation of pluripotent stem cells(R)	0.0033	36	2	1.12E-03	0.026	FOXP1, POLR2C
Hippo signaling pathway(K)	0.0142	154	3	1.14E-03	0.026	PARD6A, CTNNB1, WNT2
ATR signaling pathway(N)	0.0034	37	2	1.18E-03	0.026	RAD1, ATR
Fanconi anemia pathway(N)	0.0041	45	2	1.74E-03	0.0322	RAD1, ATR
TGF-beta receptor signaling(N)	0.0043	47	2	1.89E-03	0.0322	PARD6A, CTNNB1
Basal cell carcinoma(K)	0.0051	55	2	2.57E-03	0.0397	CTNNB1, WNT2
Rap1 signaling pathway(K)	0.0195	212	3	2.83E-03	0.0397	GNAI2, PARD6A, CTNNB1
CDC42 signaling events(N)	0.0064	70	2	4.12E-03	0.0492	PARD6A, CTNNB1
Gastric acid secretion(K)	0.0069	75	2	4.71E-03	0.0492	SSTR2, GNAI2

GeneSet	*Ratio Of Protein In GeneSet*	*Number Of Protein In GeneSet*	*Protein From Network*	*P-value*	*FDR*	*Nodes*
HTLV-I infection(K)	0.0237	258	3	4.92E-03	0.0492	CTNNB1, WNT2, ATR
Regulation of nuclear beta catenin signaling and target gene transcription(N)	0.0074	80	2	5.34E-03	0.0492	MDFIC, CTNNB1
Signaling events mediated by Hepatocyte Growth Factor Receptor (c-Met)(N)	0.0074	80	2	5.34E-03	0.0492	PARD6A, CTNNB1
Wnt signaling pathway(P)	0.0247	268	3	5.46E-03	0.0492	PPP2R5C, CTNNB1, WNT2

A.2.3 Top enriched pathways in the *MYC* dominant group in the whole dataset

"GeneSet" is the pathway name, "Ratio of protein in Geneset" indicates ratios of numbers of genes contained in pathways to total genes in the Reactome FI network; "Number of protein in GeneSet" is the number of proteins within a given pathway; and "Protein from network" indicates number of genes from the input gene list per pathway. "Nodes" are proteins from the network.

GeneSet	*Ratio Of Protein In GeneSet*	*Number Of Protein In GeneSet*	*Protein From Network*	*P-value*	*FDR*	*Nodes*
Keratinization(R)	0.0099	108	18	1.11E-16	2.26E-14	SPRR2E, KRT80, KLK5, KLK8, EVPL, CASP14, SPRR2A, DSP, JUP, KRT1, SPINK5, KRT10, KRT5, KRT15, KRT14, PKP3, DSC1, DSC3
Validated transcriptional targets of deltaNp63 isoforms(N)	0.0041	45	4	1.09E-05	1.11E-03	SFN, TP63, KRT5, KRT14
a6b1 and a6b4 Integrin signaling(N)	0.0032	35	3	1.70E-04	0.0115	COL17A1, EGFR, SFN
Bladder cancer(K)	0.0038	41	3	2.69E-04	0.0131	EGFR, MYC, FGFR3
p53 pathway(P)	0.004	44	3	3.31E-04	0.0131	SFN, TP63, SERPINB5
Posttranslational regulation of adherens junction stability and dissassembly(N)	0.0044	48	3	4.26E-04	0.0131	EGFR, DSP, JUP
Validated transcriptional targets of TAp63 isoforms(N)	0.0045	49	3	4.52E-04	0.0131	EVPL, TP63, SERPINB5
Direct p53 effectors(N)	0.0121	132	4	6.72E-04	0.0168	EGFR, SFN, TP63, SERPINB5
Central carbon metabolism in cancer(K)	0.0062	67	3	1.11E-03	0.0245	EGFR, MYC, FGFR3

GeneSet	*Ratio Of Protein In GeneSet*	*Number Of Protein In GeneSet*	*Protein From Network*	*P-value*	*FDR*	*Nodes*
p73 transcription factor network(N)	0.0069	75	3	1.54E-03	0.0275	MYC, SFN, TP63
Glucocorticoid receptor regulatory network(N)	0.0072	78	3	1.72E-03	0.0275	SFN, KRT5, KRT14
Regulation of nuclear beta catenin signaling and target gene transcription(N)	0.0074	80	3	1.85E-03	0.0275	MYC, SFN, KRT1
E-cadherin signaling in keratinocytes(N)	0.0019	21	2	1.89E-03	0.0275	EGFR, JUP
MicroRNAs in cancer(K)	0.0275	299	5	1.97E-03	0.0275	EGFR, MYC, TP63, SERPINB5, FGFR3
p53 pathway feedback loops 2(P)	0.0022	24	2	2.46E-03	0.032	MYC, TP63
Dorso-ventral axis formation(K)	0.0026	28	2	3.32E-03	0.0399	EGFR, ETS2

A.2.4 Top enriched pathways in the High Immune Subgroup

"GeneSet" is the pathway name, "Ratio of protein in Geneset" indicates ratios of numbers of genes contained in pathways to total genes in the Reactome FI network; "Number of protein in GeneSet" is the number of proteins within a given pathway; and "Protein from network" indicates number of genes from the input gene list per pathway. "Nodes" are proteins from the network.

	Ratio Of Protein In GeneSet	***Number Of Protein In GeneSet***	***Protein From Network***	***P-value***	***FDR***	***Nodes***
Interferon alpha/beta signaling(R)	0.0063	68	42	1.11E-16	1.99E-14	JAK1, RNASEL, STAT1, STAT2, PSMB8, SOCS3, SOCS1, PTPN6, IFI35, RSAD2, IFI27, BST2, ISG15, ISG20, ADAR, IRF3, IRF1, IRF2, IRF7, IRF5, IRF9, SAMHD1, MX1, IFNAR2, IFITM3, IFITM1, IFITM2, IFIT1, IFIT3, IFIT2, HLA-B, HLA-C, HLA-A, HLA-F, HLA-E, TYK2, IFI6, OASL, OAS1, OAS2, OAS3, GBP2
Antigen processing and presentation(K)	0.0071	77	39	1.11E-16	1.99E-14	RFX5, PSME1, PSME2, HLA-DRB5, HLA-DRB4, HLA-DRB3, HLA-DRB1, HLA-DPA1, HLA-DMA, HLA-DMB, KIR3DL3, HLA-DPB1, HLA-DOA, KIR2DL5A, HLA-DRA, CIITA, IFI30, CTSS, CTSB, CD74, CD8B, KIR2DL4, TAPBP, CD4, HLA-DQA1, KIR2DS3, KIR2DS5, HLA-DQB1, TAP2, TAP1, B2M, HLA-B, HLA-C, HLA-A, HLA-F, HLA-E, KLRC2, KLRC3, KLRC4

	Ratio Of Protein In GeneSet	*Number Of Protein In GeneSet*	*Protein From Network*	*P-value*	*FDR*	*Nodes*
Influenza A(K)	0.0161	175	56	1.11E-16	1.99E-14	HLA-DRB5, HLA-DRB4, EIF2AK2, EIF2AK4, HLA-DRB3, HLA-DRB1, IKBKB, IKBKE, JAK1, HLA-DPA1, RNASEL, HLA-DMA, HLA-DMB, STAT1, STAT2, HLA-DPB1, HLA-DOA, SOCS3, TNFRSF10A, NFKB1, TNFRSF10D, CXCL10, HLA-DRA, MYD88, CIITA, PIK3CD, CCL2, ICAM1, TNFSF10, RSAD2, PIK3R1, IL18, IL33, IL6, ADAR, DDX58, IRF3, IRF7, IRF9, MX1, HLA-DQA1, IFIH1, IFNGR1, HLA-DQB1, IFNAR2, MAP2K3, PML, TYK2, TLR4, TLR3, FAS, NFKBIA, CASP1, OAS1, OAS2, OAS3
Interferon gamma signaling(R)	0.0067	73	47	1.11E-16	1.99E-14	HLA-DRB5, HLA-DRB4, HLA-DRB3, HLA-DRB1, JAK1, HLA-DPA1, PTAFR, STAT1, HLA-DPB1, PIAS1, SOCS3, SOCS1, HLA-DRA, CIITA, PTPN6, IFI30, ICAM1, TRIM26, TRIM21, TRIM22, IRF3, IRF1, IRF2, IRF7, IRF5, IRF9, HLA-DQA1, TRIM8, TRIM5, IFNGR1, HLA-DQB2, HLA-DQB1, PML, B2M, HLA-B, HLA-C, HLA-A, HLA-F, HLA-E, SP100, OASL, OAS1, OAS2, OAS3, GBP2, GBP1, CAMK2G
Signaling by Interleukins(R)	0.0423	460	99	1.11E-16	1.99E-14	PSME1, PSME2, PDGFB, IL22RA2, IL18BP, IKBKB, IKBKG, JAK1, DUSP5, RAPGEF1, PTAFR, STAT1, STAT3, PSMB8, PSMB9, F13A1, RORA, PIM1, MAP3K8, SOCS3, SOCS1, NFKB1, NFKB2, CXCL10, MYD88, RASGRF2, PIK3CD, KSR1, TNFRSF1B, IL2RG, CCL2, FYN, PTPN6, IL27RA, ICAM1, PIK3R1, CCR1, STAT5B, IL24, SYK, IL18, OSMR, IL32, IL33, KITLG, PDGFRB, PDGFRA, IL4R, RASGRP1, RASGRP4, RASGRP3, IRAK2, IL6, MMP2, IRAK3, MMP9, IL9R, PSMB10, PTK2B, CD4, JUNB, HIF1A, HMOX1, TIMP1, ZEB1, ALOX5, CSF2, CSF1, FGF1, IL13RA1, RASAL3, MCL1, BCL6, LAT, SOS1, CEBPD, NOD1, NOD2, IL15RA, IL1R1,

	Ratio Of Protein In GeneSet	*Number Of Protein In GeneSet*	*Protein From Network*	*P-value*	*FDR*	*Nodes*
						IL1R2, TWIST1, LIF, GFRA1, TYK2, GFRA2, VAV1, RHOU, IL11RA, HGF, LIFR, CSF2RB, CXCL2, CASP1, BLNK, RASA4, MUC1, S1PR1, CAMK2G
Herpes simplex infection(K)	0.017	185	57	2.22E-16	3.31E-14	C3, HLA-DRB5, HLA-DRB4, EIF2AK2, EIF2AK4, HLA-DRB3, HLA-DRB1, IKBKB, IKBKG, IKBKE, JAK1, HLA-DPA1, RNASEL, HLA-DMA, HLA-DMB, STAT1, STAT2, HLA-DPB1, HLA-DOA, SOCS3, NFKB1, HLA-DRA, MYD88, CCL2, CD74, IL6, DDX58, IRF3, IRF7, IRF9, CFP, HLA-DQA1, IFIH1, IFNGR1, HLA-DQB1, IFNAR2, TAP2, TAP1, TRAF3, TRAF5, PML, IFIT1, HLA-B, HLA-C, HLA-A, HLA-F, HLA-E, TYK2, TLR3, SP100, ARNTL, FAS, NFKBIA, CASP8, OAS1, OAS2, OAS3

	Ratio Of Protein In GeneSet	*Number Of Protein In GeneSet*	*Protein From Network*	*P-value*	*FDR*	*Nodes*
Cytokine-cytokine receptor interaction(K)	0.0244	265	67	7.88E-15	1.01E-12	PDGFB, IL22RA2, PDGFC, CRLF2, CXCL13, CXCL14, CXCL16, LEPR, TNFRSF8, TNFRSF4, TNFRSF11B, TNFRSF10A, TNFRSF10D, CXCL10, CXCL11, CXCL12, FLT4, TNFSF13B, TNFRSF1B, IL2RG, CCL8, CCL2, TNFRSF6B, TNFSF10, TNFSF11, CCR1, TNFSF13, CD40, TNFRSF13B, IL24, IL18, OSMR, CD70, KITLG, PDGFRB, PDGFRA, IL4R, IL6, IL9R, TNFSF4, IFNGR1, CCL18, CCL16, CCL25, CCL23, CSF2, CSF1, IL13RA1, IFNAR2, CX3CR1, IL15RA, IL1R1, IL1R2, LIF, CX3CL1, CXCR3, CD40LG, FAS, CCL4L1, IL11RA, HGF, LIFR, CSF2RB, TGFB3, CXCL9, CXCL2, TGFBR2
Th17 cell differentiation(K)	0.0098	107	39	9.58E-14	1.07E-11	HLA-DRB5, HLA-DRB4, HLA-DRB3, HLA-DRB1, IKBKB, IKBKG, JAK1, HLA-DPA1, HLA-DMA, HLA-DMB, STAT1, STAT3, HLA-DPB1, RORA, HLA-DOA, NFKB1, HLA-DRA, IL2RG, IL27RA, STAT5B, PPP3CC, IL4R, IL6, CD4, PRKCQ, HIF1A, HLA-DQA1, NFATC2, IFNGR1, RARA, HLA-DQB1, AHR, LAT, IL1R1, RUNX1, TYK2, NFKBIA, NFKBIE, TGFBR2
Chemokine signaling pathway(K)	0.0172	187	52	2.71E-13	2.68E-11	PREX1, LYN, IKBKB, IKBKG, STAT1, STAT2, STAT3, CXCL13, CXCL14, CXCL16, RAP1A, NFKB1, CXCL10, CXCL11, CXCL12, PIK3CD, ARRB2, ROCK1, ROCK2, CCL8, CCL2, PIK3R1, CCR1, STAT5B, GNGT2, RAC2, GNAI2, PAK1, RASGRP2, GRK5, GRK6, PTK2B, ADCY4, ADCY3, ADCY7, CCL18, CCL16, CCL25, CCL23, SOS1, CX3CR1, CX3CL1, VAV3, VAV1, WAS, CXCR3, CCL4L1, ELMO1, NFKBIA, CXCL9, CXCL2, DOCK2

	Ratio Of Protein In GeneSet	*Number Of Protein In GeneSet*	*Protein From Network*	*P-value*	*FDR*	*Nodes*
NOD-like receptor signaling pathway(K)	0.0156	170	49	3.75E-13	3.34E-11	TXNIP, RBCK1, IKBKB, IKBKG, IKBKE, JAK1, RIPK3, RNASEL, STAT1, STAT2, NFKB1, MYD88, CCL2, CTSB, IL18, CYBA, CARD16, IL6, TNFAIP3, IRF3, IRF7, IRF9, ATG12, MEFV, GSDMD, TMEM173, CARD9, IFNAR2, ITPR1, PSTPIP1, TRAF3, TRAF5, RNF31, NOD1, NLRC4, NOD2, TYK2, TLR4, NLRP1, NFKBIA, CXCL2, CASP8, CASP5, CASP1, OAS1, OAS2, OAS3, GBP2, GBP1
TNF signaling pathway(K)	0.0101	110	38	9.68E-13	7.84E-11	IKBKB, CASP10, IKBKG, RIPK3, MAP3K8, MAP3K5, SOCS3, NFKB1, CXCL10, PIK3CD, TNFRSF1B, CCL2, ICAM1, PIK3R1, IL6, TNFAIP3, MMP9, CREB3L1, JUNB, CSF2, CSF1, MAP2K3, BCL3, MAP3K14, TRADD, TRAF3, TRAF5, CEBPB, NOD2, CFLAR, LIF, CX3CL1, FAS, MLKL, NFKBIA, CXCL2, CASP7, CASP8
NF-kappa B signaling pathway(K)	0.0087	95	35	1.32E-12	9.77E-11	LYN, IKBKB, IKBKG, TICAM2, LY96, NFKB1, NFKB2, CXCL12, MYD88, TNFSF13B, ICAM1, TNFSF11, BTK, CD40, SYK, RELB, CARD11, TNFAIP3, DDX58, PRKCQ, PLCG2, MAP3K14, LAT, TRADD, TRAF3, TRAF5, CFLAR, IL1R1, GADD45B, TLR4, CD40LG, CCL4L1, NFKBIA, CXCL2, BLNK
Measles(K)	0.0125	136	42	2.17E-12	1.48E-10	EIF2AK2, EIF2AK4, IKBKE, JAK1, STAT1, STAT2, STAT3, TNFRSF10A, NFKB1, TNFRSF10D, MYD88, PIK3CD, IL2RG, FYN, TNFSF10, PIK3R1, STAT5B, CCND3, CCND2, IL6, TNFAIP3, ADAR, DDX58, IRF3, IRF7, IRF9, MX1, PRKCQ, CDKN1B, IFIH1, IFNGR1, TACR1, IFNAR2, CBLB, SH2D1A, TYK2, TLR4, FAS, NFKBIA, OAS1, OAS2, OAS3

	Ratio Of Protein In GeneSet	*Number Of Protein In GeneSet*	*Protein From Network*	*P-value*	*FDR*	*Nodes*
Osteoclast differentiation(K)	0.0121	132	41	3.34E-12	2.14E-10	IKBKB, IKBKG, JAK1, STAT1, STAT2, NCF4, TNFRSF11B, SOCS3, SOCS1, NFKB1, NFKB2, PIK3CD, LCP2, FYN, FCGR2A, TNFSF11, BTK, PIK3R1, PPP3CC, SYK, RELB, CYBA, OSCAR, SIRPB1, IRF9, JUNB, PLCG2, NFATC2, FOSL2, CYLD, IFNGR1, CSF1, IFNAR2, MAP3K14, IL1R1, TYK2, CAMK4, SIRPG, NFKBIA, BLNK, TGFBR2
Viral myocarditis(K)	0.0054	59	27	4.24E-12	2.50E-10	HLA-DRB5, HLA-DRB4, HLA-DRB3, HLA-DRB1, HLA-DPA1, HLA-DMA, HLA-DMB, BID, HLA-DPB1, HLA-DOA, HLA-DRA, FYN, ICAM1, CD40, RAC2, LAMA2, HLA-DQA1, HLA-DQB1, SGCA, HLA-B, HLA-C, HLA-A, HLA-F, HLA-E, CD40LG, CASP8, MYH6
Natural killer cell mediated cytotoxicity(K)	0.0124	135	41	6.61E-12	3.70E-10	HCST, BID, KIR2DL5A, TNFRSF10A, TNFRSF10D, PIK3CD, LCP2, SH3BP2, FYN, GZMB, PTPN6, ICAM2, ICAM1, TNFSF10, PIK3R1, PPP3CC, RAC2, SYK, PAK1, MICB, KIR2DL4, PTK2B, PLCG2, KIR2DS3, KIR2DS5, NFATC2, IFNGR1, CSF2, IFNAR2, LAT, SOS1, HLA-B, HLA-C, HLA-A, HLA-E, SH2D1A, VAV3, VAV1, FAS, KLRC2, KLRC3
Staphylococcus aureus infection(K)	0.0052	56	26	7.71E-12	3.82E-10	C2, C3, HLA-DRB5, HLA-DRB4, HLA-DRB3, HLA-DRB1, HLA-DPA1, PTAFR, HLA-DMA, HLA-DMB, HLA-DPB1, HLA-DOA, HLA-DRA, C1S, C1R, FCGR2A, ICAM1, MASP1, CFB, CFD, CFH, CFI, HLA-DQA1, HLA-DQB1, C1QA, C5AR1

	Ratio Of Protein In GeneSet	*Number Of Protein In GeneSet*	*Protein From Network*	*P-value*	*FDR*	*Nodes*
Tuberculosis(K)	0.0165	179	48	7.80E-12	3.82E-10	RFX5, C3, HLA-DRB5, HLA-DRB4, HLA-DRB3, HLA-DRB1, CASP10, JAK1, HLA-DPA1, HLA-DMA, HLA-DMB, BID, STAT1, HLA-DPB1, HLA-DOA, CORO1A, NFKB1, HLA-DRA, MYD88, CIITA, APAF1, KSR1, FCGR2A, CTSS, PPP3CC, SYK, IL18, CD74, IRAK2, VDR, IL6, HLA-DQA1, MRC2, IFNGR1, HLA-DQB1, CARD9, CLEC4E, CLEC7A, TRADD, LSP1, CR1, CEBPB, NOD2, TLR1, TLR4, TGFB3, CASP8, CAMK2G
Th1 and Th2 cell differentiation(K)	0.0085	92	33	1.15E-11	5.41E-10	RBPJ, HLA-DRB5, HLA-DRB4, HLA-DRB3, HLA-DRB1, IKBKB, IKBKG, JAK1, HLA-DPA1, HLA-DMA, HLA-DMB, STAT1, HLA-DPB1, HLA-DOA, MAF, NFKB1, HLA-DRA, IL2RG, STAT5B, PPP3CC, IL4R, CD4, MAML1, PRKCQ, HLA-DQA1, NFATC2, IFNGR1, HLA-DQB1, NOTCH1, LAT, TYK2, NFKBIA, NFKBIE

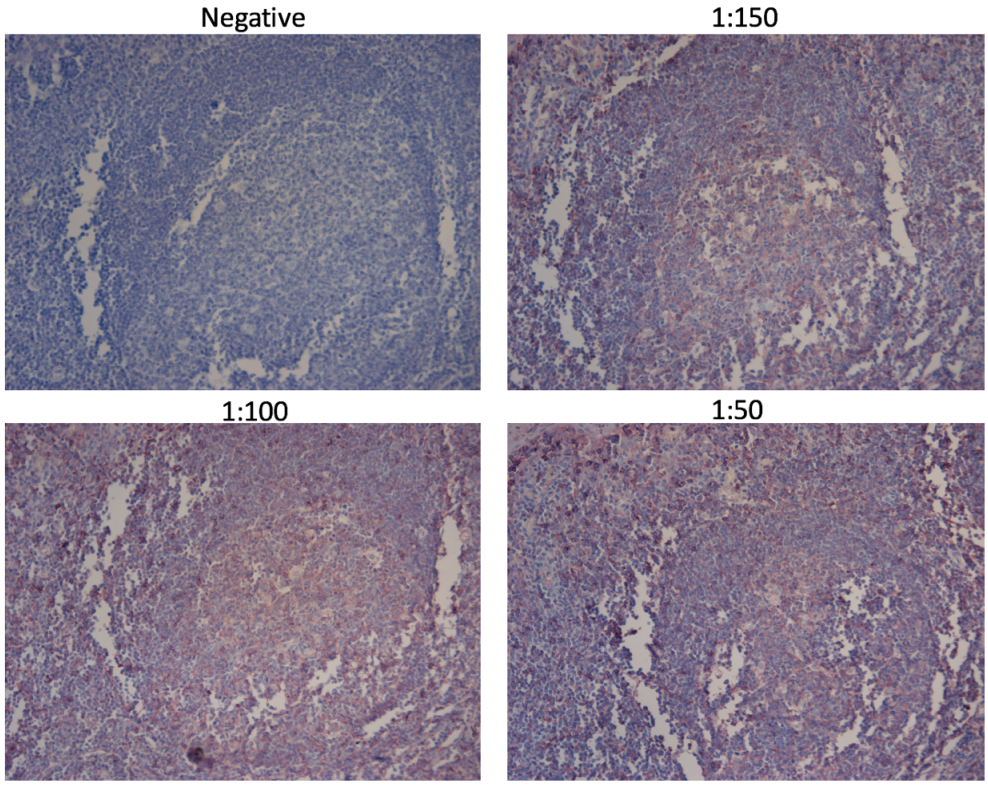
A.2.5 Genes differentially expressed between *CNNTB1* dominant and *MYC* dominant group in the Low Immune Subgroup

probe	*gene*	*detected05*	*z*	*p_value*
ilmn_1663158	*ZNF174*	0.8375	-5.9197	3.23E-09
ilmn_2084391	*RAD18*	0.7729167	-5.819325	5.91E-09
ilmn_2335198	*NCOA1*	0.7020833	-5.648805	1.62E-08
ilmn_2248863	*ZBTB38*	0.8708333	-5.635275	1.75E-08
ilmn_1654112	*PARD6A*	0.8333333	-5.595262	2.20E-08
ilmn_1688041	*TMEM53*	0.9166667	-5.36422	8.13E-08
ilmn_1764297	*ARF6*	0.9083334	-5.350971	8.75E-08
ilmn_1652309	*TTC8*	0.81875	-5.337494	9.42E-08
ilmn_1794560	*TMEM93*	0.8416666	-5.2772	1.31E-07
ilmn_3248941	*C6orf225*	0.8270833	-5.170131	2.34E-07
ilmn_1699362	*IK*	0.8270833	-5.156789	2.51E-07
ilmn_1809259	*HRASLS2*	0.4958333	-5.089804	3.58E-07
ilmn_2356786	*ADD1*	0.875	-5.063047	4.13E-07
ilmn_1765880	*C16orf57*	0.9458333	-5.063033	4.13E-07
ilmn_1700419	*HSPC171*	0.8104166	-5.02957	4.92E-07
ilmn_1660223	*CREBL2*	0.90625	-5.026237	5.00E-07
ilmn_2155516	*QTRTD1*	0.8083333	-4.962656	6.95E-07
ilmn_3237779	*TMEM184C*	0.8604167	-4.956006	7.20E-07
ilmn_1741462	*FAHD1*	0.7916667	-4.9225	8.55E-07
ilmn_1718265	*ATG5*	0.5541667	-4.922486	8.55E-07
ilmn_1798053	*SEPT14*	0.6229166	-4.902463	9.46E-07
ilmn_1807807	*SKA2*	0.9041666	-4.88233	1.05E-06
ilmn_1715901	*FBXL17*	0.7791666	-4.878997	1.07E-06
ilmn_1659082	*ZCRB1*	0.7958333	-4.868945	1.12E-06
ilmn_2371470	*C1orf124*	0.75	-4.842653	1.28E-06
ilmn_3234775	*SLAIN2*	0.8104166	-4.828843	1.37E-06
ilmn_1746426	*TOMM70A*	0.8979167	-4.825456	1.40E-06
ilmn_1812254	*KIF5A*	0.7854167	-4.822095	1.42E-06
ilmn_1667201	*WDR51B*	0.6875	-4.822095	1.42E-06
ilmn_1708580	*PDZK1IP1*	0.4229167	4.808724	1.52E-06
ilmn_1665761	*BCL11B*	0.7333333	4.838964	1.30E-06
ilmn_1722718	*BMP2*	0.4604167	4.848866	1.24E-06
ilmn_1730223	*RNF39*	0.55	4.855559	1.20E-06
ilmn_1770922	*TMEM45A*	0.86875	4.862252	1.16E-06
ilmn_1795359	*SPRR2A*	0.4354167	4.865612	1.14E-06
ilmn_1674386	*PITX1*	0.6604167	4.875638	1.09E-06
ilmn_1765668	*IL20RB*	0.6666667	4.889078	1.01E-06

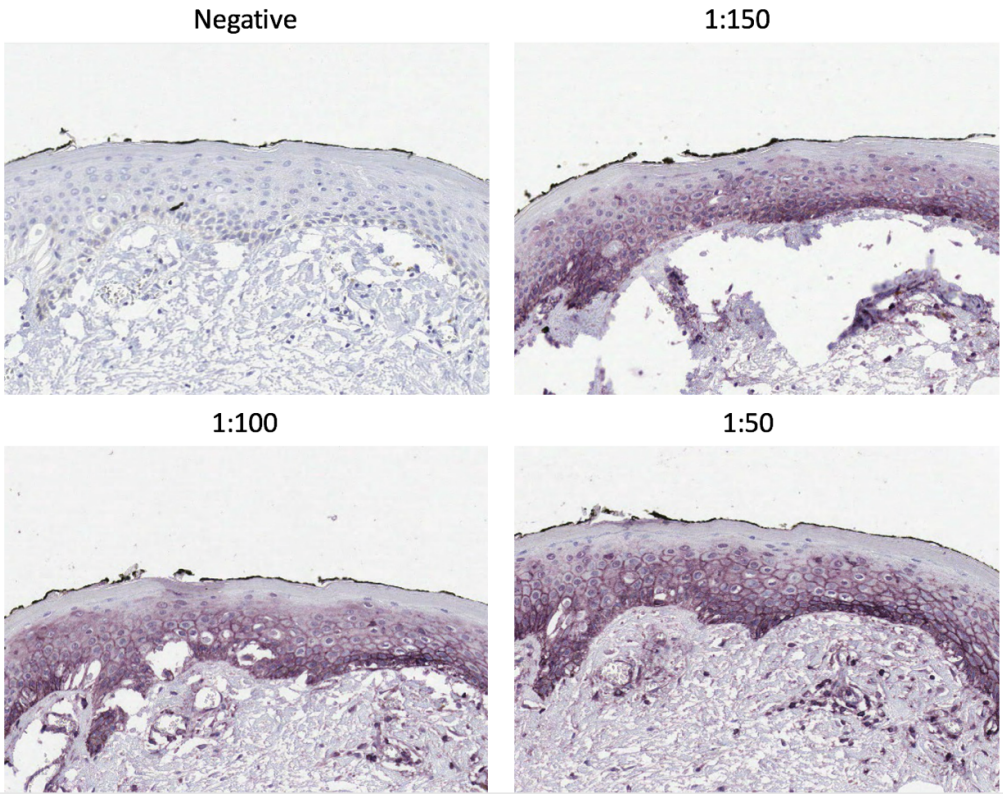
probe	*gene*	*detected05*	*z*	*p_value*
ilmn_2094952	*NUAK2*	0.7354167	4.909115	9.15E-07
ilmn_1752813	*UGT1A6*	0.3604167	4.919279	8.69E-07
ilmn_2067656	*CCND2*	0.8729166	4.982777	6.27E-07
ilmn_2110271	*OLFM2*	0.825	5.006159	5.55E-07
ilmn_1798577	*SLC6A11*	0.55	5.009548	5.46E-07
ilmn_1763491	*CKMT1B*	0.6145833	5.009987	5.44E-07
ilmn_1750785	*SYTL1*	0.4	5.196902	2.03E-07
ilmn_1768772	*DEGS2*	0.4583333	5.23041	1.69E-07
ilmn_1712283	*WNT10B*	0.38125	5.310663	1.09E-07

A.3 Representative images of antibody optimisation

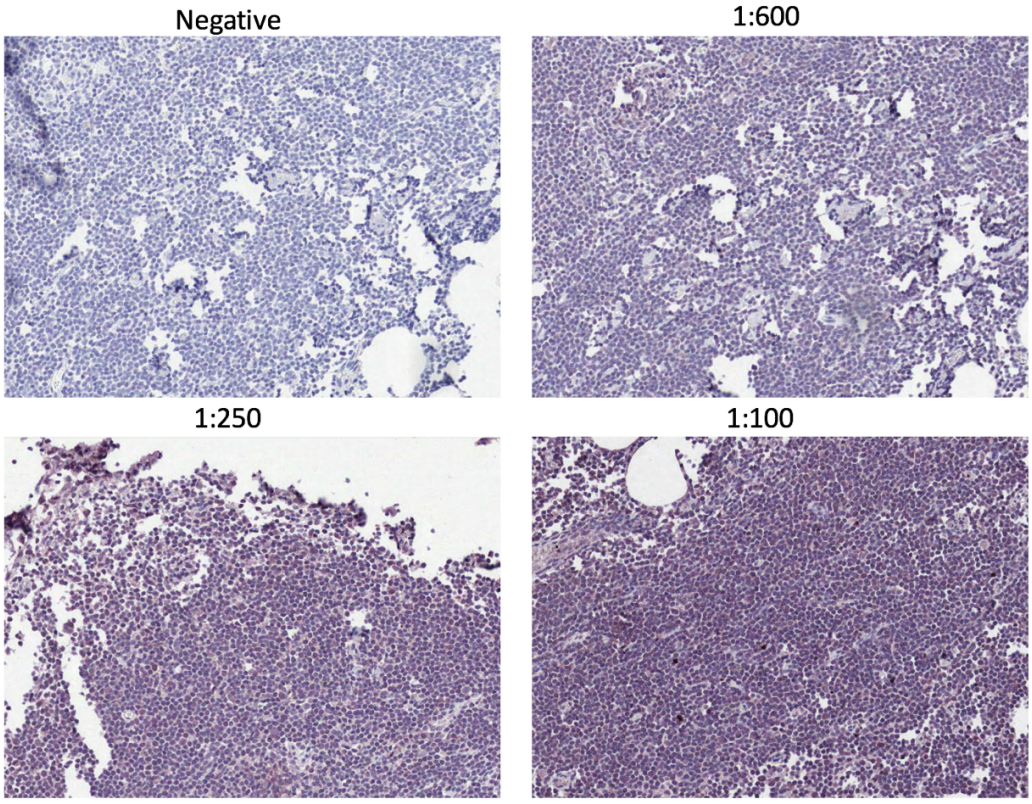
A.3.1 Representation of antibody optimisation for MYC. Magnification 20X



A.3.2 Representation of antibody optimasation for HLA-B. Magnification 20X



A.3.3 Represation of antibody optimasation for NF-κB p105. Maginifcation 20X



A.4 Top genes differentially expressed by smoking status (never/ever) in the whole dataset and in the High Immune Subgroup

A.4.1 The top 20 genes differentially expressed by smoking status (never/ever) in the whole dataset. The negative z indicates that the gene is higher in ever smokers

probe	*gene*	*z*	*P-value*	*FDR*
ilmn_1772163	*PRKY*	-4.16	0.00003	0.14
ilmn_1693338	*CYP1B1*	-4.09	0.00004	0.14
ilmn_1732039	*DDX3Y*	-4.26	0.00002	0.15
ilmn_1756506	*CYorf15B*	-4.16	0.00003	0.15
ilmn_1721218	*KRT13*	4.09	0.00004	0.16
ilmn_2052433	*CYorf14*	-4.20	0.00003	0.16
ilmn_1730670	*FSTL3*	-4.01	0.00006	0.17
ilmn_1787831	*ODF3L2*	4.02	0.00006	0.17
ilmn_1783142	*RPS4Y1*	-4.28	0.00002	0.18
ilmn_1776195	*TMSB4Y*	-3.94	0.00008	0.20
ilmn_2090059	*ZFY*	-3.92	0.00009	0.20
ilmn_1734205	*RASSF1*	-4.49	0.00001	0.21
ilmn_1678425	*WFDC8*	-3.85	0.00012	0.23
ilmn_1685690	*JARID1D*	-3.81	0.00014	0.24
ilmn_1770266	*ZNF354B*	-3.86	0.00011	0.24
ilmn_1667750	*SLC4A5*	3.83	0.00013	0.24
ilmn_2056795	*USP9Y*	-3.80	0.00015	0.24
ilmn_1741674	*PPP1R9A*	4.30	0.00002	0.25
ilmn_1670821	*CYorf15A*	-3.69	0.00022	0.28
ilmn_2179083	*LOXL4*	3.71	0.00020	0.29

A.4.2 The top 20 genes differentially expressed by smoking status (never/ever) in the High Immune Subgroup. The negative z indicates that the gene is higher in ever smokers

probe	*gene*	*z*	*P-value*	*FDR*
ilmn_2098126	*CCL5*	1.49	0.14	0.96
ilmn_1722825	*FLJ36701*	1.49	0.14	0.96
ilmn_2053679	*ACADM*	1.49	0.14	0.96
ilmn_1707799	*DGKK*	-1.51	0.13	0.96
ilmn_1704870	*PGLYRP1*	-1.48	0.14	0.96

probe	*gene*	*z*	*P-value*	*FDR*
ilmn_2062524	*RBBP4*	1.49	0.14	0.96
ilmn_1740505	*SGMS1*	-1.48	0.14	0.96
ilmn_1683475	*TOMM40*	1.51	0.13	0.96
ilmn_1813344	*C20orf7*	1.48	0.14	0.96
ilmn_1726786	*TNRC6B*	1.48	0.14	0.96
ilmn_1722059	*SAFB*	1.51	0.13	0.96
ilmn_1735959	*CNOT4*	-1.49	0.14	0.96
ilmn_1789653	*PBLD*	-1.51	0.13	0.96
ilmn_2149815	*CYP2F1*	1.48	0.14	0.96
ilmn_1748889	*PCDHGA7*	-1.49	0.14	0.96
ilmn_1713301	*DGCR2*	-1.48	0.14	0.96
ilmn_1656134	*CNOT7*	-1.51	0.13	0.96
ilmn_3249351	*SNORA35*	-1.48	0.14	0.96
ilmn_2173651	*TBC1D3*	-1.48	0.14	0.96
ilmn_1768035	*MMP12*	-1.48	0.14	0.96

References

[1] Angelova M, Charoentong P, Hackl H, Fischer ML, Snajder R, Krogsdam AM, et al. Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy. Genome Biol 2015;16:64. doi:10.1186/s13059-015-0620-6.

[2] Newton-Bishop JA, Chang Y-M, Elliott F, Chan M, Leake S, Karpavicius B, et al. Relationship between sun exposure and melanoma risk for tumours in different body sites in a large case-control study in a temperate climate. Eur J Cancer 2011;47:732–41. doi:10.1016/j.ejca.2010.10.008.

[3] Armstrong BK, Cust AE. Sun exposure and skin cancer, and the puzzle of cutaneous melanoma. Cancer Epidemiol 2017;48:147–56. doi:10.1016/j.canep.2017.04.004.

[4] Cancer Research UK 2018. http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/skin-cancer/incidence (accessed June 26, 2018).

[5] Balch CM, Gershenwald JE, Soong SJ, Thompson JF, Atkins MB, Byrd DR, et al. Final version of 2009 AJCC melanoma staging and classification. J Clin Oncol 2009;27:6199–206. doi:10.1200/JCO.2009.23.4799.

[6] Dickson P V., Gershenwald JE. Staging and Prognosis of Cutaneous Melanoma. Surg Oncol Clin N Am 2011;20:1–17. doi:10.1016/j.soc.2010.09.007.

[7] Balch CM, Buzaid AC, Soong S-J, Atkins MB, Cascinelli N, Coit DG, et al. Final Version of the American Joint Committee on Cancer Staging System for Cutaneous Melanoma. J Clin Oncol 2001;19:3635–48. doi:10.1200/JCO.2001.19.16.3635.

[8] Spatz A, Cook M., Elder D., Piepkorn M, Ruiter D., Barnhill R. Interobserver reproducibility of ulceration assessment in primary cutaneous melanomas. Eur J Cancer 2003;39:1861–5. doi:10.1016/S0959-8049(03)00325-3.

[9] Bønnelykke-Behrndtz ML, Schmidt H, Christensen IJ, Damsgaard TE, Møller HJ, Bastholt L, et al. Prognostic stratification of ulcerated melanoma: not only the extent matters. Am J Clin Pathol 2014;142:845–56. doi:10.1309/AJCPW56PHGLFTKZC.

[10] Mandalà M, Galli F, Cattaneo L, Merelli B, Rulli E, Ribero S, et al. Mitotic rate correlates with sentinel lymph node status and outcome in cutaneous melanoma greater than 1 millimeter in thickness: A multi-institutional study of 1524 cases. J Am Acad Dermatol 2017;76:264–273.e2. doi:10.1016/j.jaad.2016.08.066.

[11] Tas F, Erturk K. Histological lymphovascular invasion is associated with nodal involvement, recurrence, and survival in patients with cutaneous malignant melanoma. Int J Dermatol 2017;56:166–70. doi:10.1111/ijd.13405.

[12] Schmidt CR, Panageas KS, Coit DG, Patel A, Brady MS. An increased number of sentinel lymph nodes is associated with advanced breslow depth and lymphovascular invasion in patients with primary melanoma. Ann Surg Oncol 2009;16:948–52. doi:10.1245/s10434-009-0331-x.

[13] Callender GG, Egger ME, Burton AL, Scoggins CR, Ross MI, Stromberg AJ, et al. Prognostic implications of anatomic location of primary cutaneous melanoma of 1 mm or thicker. Am J Surg 2011;202:659–65. doi:10.1016/j.amjsurg.2011.06.048.

[14] Duarte CA, Flórez JP, López HG, Meneses MX, de Vries E. Survival of acral lentiginous melanoma in the National Cancer Institute of Colombia. J Eur Acad Dermatology Venereol 2017;31:438–42. doi:10.1111/jdv.13913.

[15] Sanlorenzo M, Osella-Abate S, Ribero S, Marenco F, Nardò T, Fierro MT, et al. Melanoma of the lower extremities: foot site is an independent risk factor for clinical outcome. Int J Dermatol 2015;54:1023–9. doi:10.1111/ijd.12730.

[16] Asgari MM, Shen L, Sokil MM, Yeh I, Jorgenson E. Prognostic factors and survival in acral lentiginous melanoma. Br J Dermatol 2017;177:428–35. doi:10.1111/bjd.15600.

[17] Sanchez A, Rodríguez D, Allard CB, Bechis SK, Sullivan RJ, Boeke CE, et al. Primary genitourinary melanoma: Epidemiology and disease-specific survival in a large population-based cohort. Urol Oncol Semin Orig Investig 2016;34:166.e7-166.e14. doi:10.1016/j.urolonc.2015.11.009.

[18] Tchelebi L, Guirguis A, Ashamalla H. Rectal melanoma: epidemiology, prognosis, and role of adjuvant radiation therapy. J Cancer Res Clin Oncol 2016;142:2569–75. doi:10.1007/s00432-016-2245-x.

[19] Hayward NK, Wilmott JS, Waddell N, Johansson PA, Field MA, Nones K, et al. Whole-genome landscapes of major melanoma subtypes. Nature 2017;545:175–80. doi:10.1038/nature22071.

[20] Joosse A, Collette S, Suciu S, Nijsten T, Patel PM, Keilholz U, et al. Sex Is an Independent Prognostic Indicator for Survival and Relapse/Progression-Free Survival in Metastasized Stage III to IV Melanoma: A Pooled Analysis of Five European Organisation for Research and Treatment of Cancer Randomized Controlled Trials. J Clin Oncol 2013;31:2337–46. doi:10.1200/JCO.2012.44.5031.

[21] Lasithiotakis K, Leiter U, Meier F, Eigentler T, Metzler G, Moehrle M, et al. Age and gender are significant independent predictors of survival in primary cutaneous melanoma. Cancer 2008;112:1795–804. doi:10.1002/cncr.23359.

[22] Weiss SA, Han J, Darvishian F, Tchack J, Han SW, Malecek K, et al. Impact of aging on host immune response and survival in melanoma: An analysis of 3 patient cohorts. J Transl Med 2016;14:1–11. doi:10.1186/s12967-016-1026-2.

[23] Derhovanessian E, Solana R, Larbi A, Pawelec G. Immunity & Ageing Immunity , ageing and cancer 2008;16:1–16. doi:10.1186/1742-4933-5-11.

[24] Kaur A, Ecker BL, Douglass SM, Kugel CH, Webster MR, Almeida F V, et al. Remodeling of the collagen matrix in aging skin promotes melanoma metastasis and affects immune cell motility. Cancer Discov 2018:CD-18-0193. doi:10.1158/2159-8290.CD-18-0193.

[25] Trolle L, Henrik-Nielsen R, Gniadecki R. Ability to self-detect malignant melanoma decreases with age. Clin Exp Dermatol 2011;36:499–501. doi:10.1111/j.1365-2230.2011.04023.x.

[26] Virchow R. Die Krankhaften Geschwulste: Dressig Vorlesungen, gehalten waehrend des Wintersemesters 1862–1863 an der Universitaet zu Berlin: Vorlesungen ueber Pathologie: Verlag von August Hirschwald 1863.

[27] Nairn R. Detection and Significance of Immunity to Skin Cancer. Melanoma Ski. Cancer Proc. Int. Cancer Conf., Sydney, Australia: V.C.N. Blight: 1972, p. 257–71.

[28] Clark WH, From L, Bernardino E a, Clark H, Bernardino E, Mihm MC. The Histogenesis and Biologic Behavior of Primary Human Malignant Melanomas of the Skin The Histogenesis Malignant Behavior of Primary Melanomas of the Skin1. Cancer Res 1969:705–27.

[29] Thomas NE, Busam KJ, From L, Kricker A, Armstrong BK, Anton-Culver H, et al. Tumor-infiltrating lymphocyte grade in primary melanomas is independently associated with melanoma-specific survival in the population-based genes, environment and melanoma study. J Clin Oncol 2013;31:4252–9. doi:10.1200/JCO.2013.51.3002.

[30] Van Houdt IS, Sluijter BJR, Moesbergen LM, Vos WM, De Gruijl TD, Molenkamp BG, et al. Favorable outcome in clinically stage II melanoma patients is associated with the presence of activated tumor infiltrating T-lymphocytes and preserved MHC class I antigen expression. Int J Cancer 2008;123:609–15. doi:10.1002/ijc.23543.

[31] Taylor RC, Patel A, Panageas KS, Busam KJ, Brady MS. Tumor-infiltrating lymphocytes predict sentinel lymph node positivity in patients with cutaneous melanoma. J Clin Oncol 2007;25:869–75. doi:10.1200/JCO.2006.08.9755.

[32] Eriksson H, Frohm-Nilsson M, Järås J, Kanter-Lewensohn L, Kjellman P, Månsson-Brahme E, et al. Prognostic factors in localized invasive primary cutaneous malignant melanoma: Results of a large population-based study. Br J Dermatol 2015;172:175–86. doi:10.1111/bjd.13171.

[33] Monshizadeh L, Hanikeri M, Beer TW, Heenan PJ. A critical review of melanoma pathology reports for patients referred to the Western Australian Melanoma Advisory Service. Pathology 2012;44:441–7. doi:10.1097/PAT.0b013e328355767e.

[34] Urso C. Interobserver reproducibility of histological features in cutaneous malignant melanoma. J Clin Pathol 2005;58:1194–8. doi:10.1136/jcp.2005.026765.

[35] Busam KJ, Antonescu CR, Marghoob AA, Nehal KS, Sachs DL, Shia J, et al. Histologic Classification of Tumor-Infiltrating Lymphocytes in Primary Cutaneous Malignant Melanoma. Am J Clin Pathol 2001;115:856–60. doi:10.1309/G6EK-Y6EH-0LGY-6D6P.

[36] Azimi F, Scolyer RA, Rumcheva P, Moncrieff M, Murali R, McCarthy SW, et al. Tumor-infiltrating lymphocyte grade is an independent predictor of sentinel lymph node status and survival in patients with cutaneous melanoma. J Clin Oncol 2012;30:2678–83. doi:10.1200/JCO.2011.37.8539.

[37] Dhillon AS, Hagan S, Rath O, Kolch W. MAP kinase signalling pathways in cancer. Oncogene 2007;26:3279–90. doi:10.1038/sj.onc.1210421.

[38] Paluncic J, Kovacevic Z, Jansson PJ, Kalinowski D, Merlot AM, Huang MLH, et al. Roads to melanoma: Key pathways and emerging players in melanoma progression and oncogenic signaling. Biochim Biophys Acta - Mol Cell Res 2016;1863:770–84. doi:10.1016/j.bbamcr.2016.01.025.

[39] Bradish JR, Cheng L. Molecular pathology of malignant melanoma: changing the clinical practice paradigm toward a personalized approach. Hum Pathol 2014;45:1315–26. doi:10.1016/j.humpath.2014.04.001.

[40] Smalley KSM, Sondak VK. Inhibition of BRAF and MEK in BRAF-mutant melanoma. Lancet 2015;386:410–2. doi:10.1016/S0140-6736(15)60972-2.

[41] Fedorenko I V, Gibney GT, Smalley KSM. NRAS mutant melanoma: biological behavior and future strategies for therapeutic management. Oncogene 2013;32:3009–18. doi:10.1038/onc.2012.453.

[42] Mandalà M, Merelli B, Massi D. Nras in melanoma: Targeting the undruggable target. Crit Rev Oncol Hematol 2014;92:107–22. doi:10.1016/j.critrevonc.2014.05.005.

[43] Liu P, Cheng H, Roberts TM, Zhao JJ. Targeting the phosphoinositide 3-kinase pathway in cancer. Nat Rev Drug Discov 2009;8:627–44. doi:10.1038/nrd2926.

[44] Leevers SJ, Vanhaesebroeck B, Waterfield MD. Signalling through phosphoinositide 3-kinases: the lipids take centre stage. Curr Opin Cell Biol 1999;11:219–25.

[45] Okkenhaug K. Signaling by the phosphoinositide 3-kinase family in immune cells. Annu Rev Immunol 2013;31:675–704. doi:10.1146/annurev-immunol-032712-095946.

[46] Akbani R, Akdemir KC, Aksoy BA, Albert M, Ally A, Amin SB, et al. Genomic Classification of Cutaneous Melanoma. Cell 2015;161:1681–96. doi:10.1016/j.cell.2015.05.044.

[47] Dankort D, Curley DP, Cartlidge RA, Nelson B, Karnezis AN, Damsky Jr WE, et al. BrafV600E cooperates with Pten loss to induce metastatic melanoma. Nat Genet 2009;41:544–52. doi:10.1038/ng.356.

[48] Hodis E, Watson IR, Kryukov G V., Arold ST, Imielinski M, Theurillat JP, et al. A

landscape of driver mutations in melanoma. Cell 2012;150:251–63. doi:10.1016/j.cell.2012.06.024.

[49] Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin A V., et al. Signatures of mutational processes in human cancer. Nature 2013;500:415–21. doi:10.1038/nature12477.

[50] Guan J, Gupta R, Filipp F V. Cancer systems biology of TCGA SKCM: efficient detection of genomic drivers in melanoma. Sci Rep 2015;5:7857. doi:10.1038/srep07857.

[51] Woo SR, Fuertes MB, Corrales L, Spranger S, Furdyna MJ, Leung MYK, et al. STING-dependent cytosolic DNA sensing mediates innate immune recognition of immunogenic tumors. Immunity 2014;41:830–42. doi:10.1016/j.immuni.2014.10.017.

[52] Corrales L, McWhirter SM, Dubensky TW, Gajewski TF. The host STING pathway at the interface of cancer and immunity. J Clin Invest 2016;126:2404–11. doi:10.1172/JCI86892.

[53] Murphy K, Weaver C. Janeway's Immunobiology, 9th edition. CRC Press; 2016.

[54] Engel P, Boumsell L, Balderas R, Bensussan A, Gattei V, Horejsi V, et al. CD Nomenclature 2015: Human Leukocyte Differentiation Antigen Workshops as a Driving Force in Immunology: Table I. J Immunol 2015;195:4555–63. doi:10.4049/jimmunol.1502033.

[55] Smyth MJ, Dunn GP, Schreiber RD. Cancer Immunosurveillance and Immunoediting: The Roles of Immunity in Suppressing Tumor Development and Shaping Tumor Immunogenicity, 2006, p. 1–50. doi:10.1016/S0065-2776(06)90001-7.

[56] Gajewski TF, Schreiber H, Fu Y-X. Innate and adaptive immune cells in the tumor microenvironment. Nat Immunol 2013;14:1014–22. doi:10.1038/ni.2703.

[57] Ehrlich PU. Ueber den jetzigen stand der Karzinomforschung. NedTijdschr Geneeskd 1909:273–290.

[58] Burnet M. Cancer--A Biological Approach: III. Viruses Associated with Neoplastic Conditions. IV. Practical Applications. Bmj 1957;1:841–7. doi:10.1136/bmj.1.5023.841.

[59] Khosravi H, Akabane AL, Alloo A, Nazarian RM, Boland GM. Metastatic

melanoma with spontaneous complete regression of a thick primary lesion. JAAD Case Reports 2016;2:439–41. doi:10.1016/j.jdcr.2016.09.011.

[60] Zorn E, Hercend T. A natural cytotoxic T cell response in a spontaneously regressing human melanoma targets a neoantigen resulting from a somatic point mutation. Eur J Immunol 1999;29:592–601. doi:10.1002/(SICI)1521-4141(199902)29:02<592::AID-IMMU592>3.0.CO;2-2.

[61] E. Z, T. H. Thin regressing malignant melanoma: significance of concurrent regional lymph node metastases. Histopathology 1989;3:257–65.

[62] Quaglino P, Marenco F, Osella-Abate S, Cappello N, Ortoncelli M, Salomone B, et al. Vitiligo is an independent favourable prognostic factor in stage III and IV metastatic melanoma patients: results from a single-institution hospital-based observational cohort study. Ann Oncol 2010;21:409–14. doi:10.1093/annonc/mdp325.

[63] Byrne KT, Turk MJ. New Perspectives on the Role of Vitiligo in Immune Responses to Melanoma. Oncotarget 2011;2. doi:10.18632/oncotarget.323.

[64] Byrne KT, Côté AL, Zhang P, Steinberg SM, Guo Y, Allie R, et al. Autoimmune melanocyte destruction is required for robust CD8+ memory T cell responses to mouse melanoma. J Clin Invest 2011;121:1797–809. doi:10.1172/JCI44849.

[65] Kubica AW, Brewer JD. Melanoma in Immunosuppressed Patients. Mayo Clin Proc 2012;87:991–1003. doi:10.1016/j.mayocp.2012.04.018.

[66] MacKie RM, Reid R, Junor B. Fatal Melanoma Transferred in a Donated Kidney 16 Years after Melanoma Surgery. N Engl J Med 2003;348:567–8. doi:10.1056/NEJM200302063480620.

[67] Durgeau A, Virk Y, Corgnac S, Mami-Chouaib F. Recent advances in targeting CD8 T-cell immunity for more effective cancer immunotherapy. Front Immunol 2018;9. doi:10.3389/fimmu.2018.00014.

[68] Wirth TC, Kühnel F. Neoantigen Targeting-Dawn of a New Era in Cancer Immunotherapy? Front Immunol 2017;8:1848. doi:10.3389/fimmu.2017.01848.

[69] O'Donnell JS, Teng MWL, Smyth MJ. Cancer immunoediting and resistance to T cell-based immunotherapy. Nat Rev Clin Oncol 2018. doi:10.1038/s41571-018-0142-8.

[70] Tarazona R, Duran E, Solana R. Natural Killer Cell Recognition of Melanoma: New

Clues for a More Effective Immunotherapy. Front Immunol 2015;6:649. doi:10.3389/fimmu.2015.00649.

[71] Spranger S, Dai D, Horton B, Gajewski TF. Tumor-Residing Batf3 Dendritic Cells Are Required for Effector T Cell Trafficking and Adoptive T Cell Therapy. Cancer Cell 2017;31:711–723.e4. doi:10.1016/j.ccell.2017.04.003.

[72] Vanbervliet B, Bendriss-Vermare N, Massacrier C, Homey B, de Bouteiller O, Brière F, et al. The Inducible CXCR3 Ligands Control Plasmacytoid Dendritic Cell Responsiveness to the Constitutive Chemokine Stromal Cell–derived Factor 1 (SDF-1)/CXCL12. J Exp Med 2003;198:823–30. doi:10.1084/jem.20020437.

[73] Liu Y-J. IPC: professional type 1 interferon-producing cells and plasmacytoid dendritic cell precursors. Annu Rev Immunol 2005;23:275–306. doi:10.1146/annurev.immunol.23.021704.115633.

[74] Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. Cell 2011;144:646–74. doi:10.1016/j.cell.2011.02.013.

[75] Spranger S, Bao R, Gajewski TF. Melanoma-intrinsic β-catenin signalling prevents anti-tumour immunity. Nature 2015;523:231–5. doi:10.1038/nature14404.

[76] Nsengimana J, Laye J, Filia A, O'Shea S, Muralidhar S, Poźniak J, et al. β-Catenin–mediated immune evasion pathway frequently operates in primary cutaneous melanomas. J Clin Invest 2018;128:2048–63. doi:10.1172/JCI95351.

[77] Spranger S, Spaapen RM, Zha Y, Williams J, Meng Y, Ha TT, et al. Up-Regulation of PD-L1, IDO, and Tregs in the Melanoma Tumor Microenvironment Is Driven by CD8+ T Cells. Sci Transl Med 2013;5:200ra116-200ra116. doi:10.1126/scitranslmed.3006504.

[78] Mantovani A. The growing diversity and spectrum of action of myeloid-derived suppressor cells. Eur J Immunol 2010;40:3317–20. doi:10.1002/eji.201041170.

[79] Gros A, Turcotte S, Wunderlich JR, Ahmadzadeh M, Dudley ME, Rosenberg SA. Myeloid Cells Obtained from the Blood but Not from the Tumor Can Suppress T-cell Proliferation in Patients with Melanoma. Clin Cancer Res 2012;18:5212–23. doi:10.1158/1078-0432.CCR-12-1108.

[80] Veglia F, Perego M, Gabrilovich D. Myeloid-derived suppressor cells coming of age. Nat Immunol 2018;19:108–19. doi:10.1038/s41590-017-0022-x.

[81] Oleinika K, Nibbs RJ, Graham GJ, Fraser AR. Suppression, subversion and escape:

the role of regulatory T cells in cancer progression. Clin Exp Immunol 2013;171:36–45. doi:10.1111/j.1365-2249.2012.04657.x.

[82] Wang M, Zhao J, Zhang L, Wei F, Lian Y, Wu Y, et al. Role of tumor microenvironment in tumorigenesis. J Cancer 2017;8:761–73. doi:10.7150/jca.17648.

[83] Lund AW, Duraes F V, Hirosue S, Raghavan VR, Nembrini C, Thomas SN, et al. VEGF-C promotes immune tolerance in B16 melanomas and cross-presentation of tumor antigen by lymph node lymphatics. Cell Rep 2012;1:191–9. doi:10.1016/j.celrep.2012.01.005.

[84] Lund AW, Wagner M, Fankhauser M, Steinskog ES, Broggi MA, Spranger S, et al. Lymphatic vessels regulate immune microenvironments in human and murine melanoma. J Clin Invest 2016;126:3389–402. doi:10.1172/JCI79434.

[85] Ziani L, Safta-Saadoun T Ben, Gourbeix J, Cavalcanti A, Robert C, Favre G, et al. Melanoma-associated fibroblasts decrease tumor cell susceptibility to NK cell-mediated killing through matrix-metalloproteinases secretion. Oncotarget 2017;8:19780–94. doi:10.18632/oncotarget.15540.

[86] Xing F, Saidou J, Watabe K. Cancer associated fibroblasts (CAFs) in tumor microenvironment. Front Biosci (Landmark Ed 2010;15:166–79. doi:20036813.

[87] Chen DS, Mellman I. Elements of cancer immunity and the cancer-immune set point. Nature 2017;541:321–30. doi:10.1038/nature21349.

[88] Qiu F, Liang C-L, Liu H, Zeng Y-Q, Hou S, Huang S, et al. Impacts of cigarette smoking on immune responsiveness: Up and down or upside down? Oncotarget 2017;8:268–84. doi:10.18632/oncotarget.13613.

[89] Aranow C. Vitamin D and the immune system. J Investig Med 2011;59:881–6. doi:10.2310/JIM.0b013e31821b8755.

[90] Walker LSK, Sansom DM. The emerging role of CTLA4 as a cell-extrinsic regulator of T cell responses. Nat Rev Immunol 2011;11:852–63. doi:10.1038/nri3108.

[91] Wei SC, Duffy CR, Allison JP. Fundamental mechanisms of immune checkpoint blockade therapy. Cancer Discov 2018;8:1069–86. doi:10.1158/2159-8290.CD-18-0367.

[92] Hodi FS, O'Day SJ, McDermott DF, Weber RW, Sosman JA, Haanen JB, et al. Improved Survival with Ipilimumab in Patients with Metastatic Melanoma. N

Engl J Med 2010;363:711–23. doi:10.1056/NEJMoa1003466.

[93] Robert C, Schachter J, Long G V., Arance A, Grob JJ, Mortier L, et al. Pembrolizumab versus Ipilimumab in Advanced Melanoma. N Engl J Med 2015;372:2521–32. doi:10.1056/NEJMoa1503093.

[94] Weber JS. Nivolumab versus chemotherapy in patients with advanced melanoma who progress ... Página 1 de 2 Nivolumab versus chemotherapy in patients with advanced melanoma who progressed after anti-CTLA-4 treatment (CheckMate 037): a randomised , controlled , open. Lancet Oncol 2015;2045:1–2.

[95] Wolchok JD, Chiarion-Sileni V, Gonzalez R, Rutkowski P, Grob J-J, Cowey CL, et al. Overall Survival with Combined Nivolumab and Ipilimumab in Advanced Melanoma. N Engl J Med 2017;377:1345–56. doi:10.1056/NEJMoa1709684.

[96] Amaria RN, Reddy SM, Tawbi HA, Davies MA, Ross MI, Glitza IC, et al. Neoadjuvant immune checkpoint blockade in high-risk resectable melanoma. Nat Med 2018;24:1649–54. doi:10.1038/s41591-018-0197-1.

[97] Blank CU, Rozeman EA, Fanchi LF, Sikorska K, van de Wiel B, Kvistborg P, et al. Neoadjuvant versus adjuvant ipilimumab plus nivolumab in macroscopic stage III melanoma. Nat Med 2018;24:1655–61. doi:10.1038/s41591-018-0198-0.

[98] Rosenberg SA, Packard BS, Aebersold PM, Solomon D, Topalian SL, Toy ST, et al. Use of Tumor-Infiltrating Lymphocytes and Interleukin-2 in the Immunotherapy of Patients with Metastatic Melanoma. N Engl J Med 1988;319:1676–80. doi:10.1056/NEJM198812223192527.

[99] Perica K, Varela JC, Oelke M, Schneck J. Adoptive T Cell Immunotherapy For Cancer. Rambam Maimonides Med J 2015;6:e0004. doi:10.5041/RMMJ.10179.

[100] June CH. Adoptive T cell therapy for cancer in the clinic. J Clin Invest 2007;117:1466–76. doi:10.1172/JCI32446.

[101] June CH, O'Connor RS, Kawalekar OU, Ghassemi S, Milone MC. CAR T cell immunotherapy for human cancer. Science (80-) 2018;359:1361–5. doi:10.1126/science.aar6711.

[102] Larkin J, Chiarion-Sileni V, Gonzalez R, Grob JJ, Cowey CL, Lao CD, et al. Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma. N Engl J Med 2015;373:23–34. doi:10.1056/NEJMoa1504030.

[103] Svane IM, Verdegaal EM. Achievements and challenges of adoptive T cell therapy with tumor-infiltrating or blood-derived lymphocytes for metastatic melanoma: what is needed to achieve standard of care? Cancer Immunol Immunother 2014;63:1081–91. doi:10.1007/s00262-014-1580-5.

[104] Rosenberg SA, Yang JC, Sherry RM, Kammula US, Hughes MS, Phan GQ, et al. Durable Complete Responses in Heavily Pretreated Patients with Metastatic Melanoma Using T-Cell Transfer Immunotherapy. Clin Cancer Res 2011;17:4550–7. doi:10.1158/1078-0432.CCR-11-0116.

[105] Stack EC, Wang C, Roman KA, Hoyt CC. Multiplexed immunohistochemistry, imaging, and quantitation: A review, with an assessment of Tyramide signal amplification, multispectral imaging and multiplex analysis. Methods 2014;70:46–58. doi:10.1016/j.ymeth.2014.08.016.

[106] Angelo M, Bendall SC, Finck R, Hale MB, Hitzman C, Borowsky AD, et al. Multiplexed ion beam imaging of human breast tumors. Nat Med 2014;20:436–42. doi:10.1038/nm.3488.

[107] Adan A, Alizada G, Kiraz Y, Baran Y, Nalbant A. Flow cytometry: basic principles and applications. Crit Rev Biotechnol 2017;37:163–76. doi:10.3109/07388551.2015.1128876.

[108] Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, Obenauf AC, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. Immunity 2013;39:782–95. doi:10.1016/j.immuni.2013.10.003.

[109] Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol 2016;17:218. doi:10.1186/s13059-016-1070-5.

[110] Racle J, de Jonge K, Baumgaertner P, Speiser DE, Gfeller D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. Elife 2017;6. doi:10.7554/eLife.26476.

[111] Danaher P, Warren S, Dennis L, D'Amico L, White A, Disis ML, et al. Gene expression markers of Tumor Infiltrating Leukocytes. J Immunother Cancer 2017;5:1–15. doi:10.1186/s40425-017-0215-8.

[112] Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol 2017;18:220. doi:10.1186/s13059-017-1349-1.

[113] Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 2015;12:453–7. doi:10.1038/nmeth.3337.

[114] Newton-Bishop JA, Beswick S, Randerson-Moor J, Chang YM, Affleck P, Elliott F, et al. Serum 25-hydroxyvitamin D3 levels are associated with Breslow thickness at presentation and survival from melanoma. J Clin Oncol 2009;27:5439–44. doi:10.1200/JCO.2009.22.1135.

[115] Du P, Kibbe WA, Lin SM. lumi: A pipeline for processing Illumina microarray. Bioinformatics 2008;24:1547–8. doi:10.1093/bioinformatics/btn224.

[116] Lauss M, Visne I, Kriegner A, Ringnér M, Jönsson G, Höglund M. Monitoring of Technical Variation in Quantitative High-Throughput Datasets. Cancer Inform 2013;12:CIN.S12862. doi:10.4137/CIN.S12862.

[117] Jönsson G, Busch C, Knappskog S, Geisler J, Miletic H, Ringnér M, et al. Gene expression profiling-based identification of molecular subtypes in stage IV melanomas with different clinical outcome. Clin Cancer Res 2010;16:3356–67. doi:10.1158/1078-0432.CCR-09-2509.

[118] Harbst K, Staaf J, Lauss M, Karlsson A, Masback A, Johansson I, et al. Molecular Profiling Reveals Low- and High-Grade Forms of Primary Melanoma. Clin Cancer Res 2012;18:4026–36. doi:10.1158/1078-0432.CCR-12-0343.

[119] Abbotts R, Jewell R, Nsengimana J, Maloney DJ, Simeonov A, Seedhouse C, et al. Targeting human apurinic/apyrimidinic endonuclease 1 (APE1) in phosphatase and tensin homolog (PTEN) deficient melanoma cells for personalized therapy. Oncotarget 2014;5. doi:10.18632/oncotarget.1926.

[120] Jewell R, Elliott F, Laye J, Nsengimana J, Davies J, Walker C, et al. The clinicopathological and gene expression patterns associated with ulceration of primary melanoma. Pigment Cell Melanoma Res 2015;28:94–104. doi:10.1111/pcmr.12315.

[121] Ronaghi M. DNA SEQUENCING:A Sequencing Method Based on Real-Time Pyrophosphate. Science (80-) 1998;281:363–5. doi:10.1126/science.281.5375.363.

[122] Irizarry RA. Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res 2003;31:15e–15. doi:10.1093/nar/gng015.

[123] Brettschneider J, Collin F, Bolstad BM, Speed TP. Quality Assessment for Short Oligonucleotide Microarray Data. Technometrics 2008;50:241–64. doi:10.1198/004017008000000334.

[124] Raaijmakers MIG, Widmer DS, Maudrich M, Koch T, Langer A, Flace A, et al. A new live-cell biobank workflow efficiently recovers heterogeneous melanoma cells from native biopsies. Exp Dermatol 2015;24:377–80. doi:10.1111/exd.12683.

[125] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15–21. doi:10.1093/bioinformatics/bts635.

[126] Murdoch DJ, Chow ED. A Graphical Display of Large Correlation Matrices. Am Stat 1996;50:178. doi:10.2307/2684435.

[127] Friendly M. Corrgrams. Am Stat 2002;56:316–24. doi:10.1198/000313002533.

[128] D'Haeseleer P. How does gene expression clustering work? Nat Biotechnol 2005;23:1499–501. doi:10.1038/nbt1205-1499.

[129] Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 1998;95:14863–8.

[130] Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM. Systematic determination of genetic network architecture. Nat Genet 1999;22:281–5. doi:10.1038/10343.

[131] Gibbons FD. Judging the Quality of Gene Expression-Based Clustering Methods Using Gene Annotation. Genome Res 2002;12:1574–81. doi:10.1101/gr.397002.

[132] Costa IG, Carvalho F de AT de, Souto MCP de. Comparative analysis of clustering methods for gene expression time course data. Genet Mol Biol 2004;27:623–31. doi:10.1590/S1415-47572004000400025.

[133] Datta S, Datta S. Comparisons and validation of statistical clustering techniques for microarray gene expression data. Bioinformatics 2003;19:459–66. doi:10.1093/bioinformatics/btg025.

[134] Monti S, Tamayo P, Mesirov J, Golub T. Consensus clustering: A resampling based method for class discovery and visualization of gene expression

microarray data. Mach Learn 2003;52:91–118.

[135] Wilkerson MD, Hayes DN. ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. Bioinformatics 2010;26:1572–3. doi:10.1093/bioinformatics/btq170.

[136] Cirenajwis H, Ekedahl H, Lauss M, Harbst K, Carneiro A, Enoksson J, et al. Molecular stratification of metastatic melanoma using gene expression profiling: Prediction of survival outcome and benefit from molecular targeted therapy. Oncotarget 2015;6:12297–309. doi:10.18632/oncotarget.3655.

[137] Nsengimana J, Laye J, Filia A, Walker C, Jewell R, Van den Oord JJ, et al. Independent replication of a melanoma subtype gene signature and evaluation of its prognostic value and biological correlates in a population cohort. Oncotarget 2015;6:11683–93. doi:10.18632/oncotarget.3549.

[138] Gentles AJ, Newman AM, Liu CL, Bratman S V, Feng W, Kim D, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat Med 2015;21:938–45. doi:10.1038/nm.3909.

[139] Masten BJ, Olson GK, Tarleton CA, Rund C, Schuyler M, Mehran R, et al. Characterization of Myeloid and Plasmacytoid Dendritic Cells in Human Lung. J Immunol 2006;177:7784–93. doi:10.4049/jimmunol.177.11.7784.

[140] Fening K, Parekh V, McKay K. CD123 immunohistochemistry for plasmacytoid dendritic cells is useful in the diagnosis of scarring alopecia. J Cutan Pathol 2016;43:643–8. doi:10.1111/cup.12725.

[141] Hori S. Control of Regulatory T Cell Development by the Transcription Factor Foxp3. Science (80-) 2003;299:1057–61. doi:10.1126/science.1079490.

[142] Zhou M, Ouyang W. The Function Role of GATA-3 in Th1 and Th2 Differentiation. Immunol Res 2003;28:25–38. doi:10.1385/IR:28:1:25.

[143] Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. Science (80-) 2015;347:1260419–1260419. doi:10.1126/science.1260419.

[144] Aras S, Raza Zaidi M. TAMeless traitors: Macrophages in cancer progression and metastasis. Br J Cancer 2017;117:1583–91. doi:10.1038/bjc.2017.356.

[145] Riaz N, Havel JJ, Makarov V, Desrichard A, Urba WJ, Sims JS, et al. Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab. Cell

2017;171:934–949.e15. doi:10.1016/j.cell.2017.09.028.

[146] Lauss M, Donia M, Harbst K, Andersen R, Mitra S, Rosengren F, et al. Mutational and putative neoantigen load predict clinical benefit of adoptive T cell therapy in melanoma. Nat Commun 2017;8:1–10. doi:10.1038/s41467-017-01460-0.

[147] Van Allen EM, Miao D, Schilling B, Shukla SA, Blank C, Zimmer L, et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. Science (80-) 2015;350:207–11. doi:10.1126/science.aad0095.

[148] Trucco LD, Mundra PA, Hogan K, Garcia-Martinez P, Viros A, Mandal AK, et al. Ultraviolet radiation-induced DNA damage is prognostic for outcome in melanoma. Nat Med 2018. doi:10.1038/s41591-018-0265-6.

[149] Spranger S, Gajewski TF, Kline J. MYC — a thorn in the side of cancer immunity. Cell Res 2016;26:639–40. doi:10.1038/cr.2016.50.

[150] Łuksza M, Riaz N, Makarov V, Balachandran VP, Hellmann MD, Solovyov A, et al. A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy. Nature 2017;551:517–20. doi:10.1038/nature24473.

[151] Balachandran VP, Łuksza M, Zhao JN, Makarov V, Moral JA, Remark R, et al. Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. Nature 2017;551:512–6. doi:10.1038/nature24462.

[152] Yadav M, Jhunjhunwala S, Phung QT, Lupardus P, Tanguay J, Bumbaca S, et al. Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. Nature 2014;515:572–6. doi:10.1038/nature14001.

[153] Sharkey MS, Lizée G, Gonzales MI, Patel S, Topalian SL. CD4+ T-Cell Recognition of Mutated B-RAF in Melanoma Patients Harboring the V599E Mutation. Cancer Res 2004;64:1595–9. doi:10.1158/0008-5472.CAN-03-3231.

[154] Andersen MH, Fensterle J, Ugurel S, Reker S, Houben R, Guldberg P, et al. Immunogenicity of Constitutively Active BRaf. J Immunol 2004;2:5456–60.

[155] Somasundaram R, Swoboda R, Caputo L, Otvos L, Weber B, Volpe P, et al. Human leukocyte antigen-A2-restricted CTL responses to mutated BRAF peptides in melanoma patients. Cancer Res 2006;66:3287–93. doi:10.1158/0008-5472.CAN-05-1932.

[156] Thomas NE, Edmiston SN, Alexander A, Groben PA, Parrish E, Kricker A, et al. Association Between NRAS and BRAF Mutational Status and Melanoma-Specific

Survival Among Patients With Higher-Risk Primary Melanoma. JAMA Oncol 2015;1:359. doi:10.1001/jamaoncol.2015.0493.

[157] Seliger B, Harders C, Wollscheid U, Staege MS, Reske-Kunz AB, Huber C. Suppression of MHC class I antigens in oncogenic transformants: association with decreased recognition by cytotoxic T lymphocytes. Exp Hematol 1996;24:1275–9.

[158] Johnson DB, Lovly CM, Flavin M, Panageas KS, Ayers GD, Zhao Z, et al. Impact of NRAS Mutations for Patients with Advanced Melanoma Treated with Immune Therapies. Cancer Immunol Res 2015;3:288–95. doi:10.1158/2326-6066.CIR-14-0207.

[159] Tarca AL, Bhatti G, Romero R. A comparison of gene set analysis methods in terms of sensitivity, prioritization and specificity. PLoS One 2013;8. doi:10.1371/journal.pone.0079217.

[160] Wu G, Feng X, Stein L, Hanahan D, Weinberg R, Vogelstein B, et al. A human functional protein interaction network and its application to cancer data analysis. Genome Biol 2010;11:R53. doi:10.1186/gb-2010-11-5-r53.

[161] Scardoni G, Petterlini M, Laudanna C. Analyzing biological network parameters with CentiScaPe. Bioinformatics 2009;25:2857–9. doi:10.1093/bioinformatics/btp517.

[162] Shannon P. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res 2003;13:2498–504. doi:10.1101/gr.1239303.

[163] Kerrien S, Alam-Faruque Y, Aranda B, Bancarz I, Bridge A, Derow C, et al. IntAct--open source resource for molecular interaction data. Nucleic Acids Res 2007;35:D561–5. doi:10.1093/nar/gkl958.

[164] Peri S. Development of Human Protein Reference Database as an Initial Platform for Approaching Systems Biology in Humans. Genome Res 2003;13:2363–71. doi:10.1101/gr.1680803.

[165] Breitkreutz B-J, Stark C, Reguly T, Boucher L, Breitkreutz A, Livstone M, et al. The BioGRID Interaction Database: 2008 update. Nucleic Acids Res 2007;36:D637–40. doi:10.1093/nar/gkm1001.

[166] Flicek P, Aken BL, Ballester B, Beal K, Bragin E, Brent S, et al. Ensembl's 10th year.

Nucleic Acids Res 2010;38:D557–62. doi:10.1093/nar/gkp972.

[167] Lee HK. Coexpression Analysis of Human Genes Across Many Microarray Data Sets. Genome Res 2004;14:1085–94. doi:10.1101/gr.1910904.

[168] Prieto C, Risueño A, Fontanillo C, De Las Rivas J. Human Gene Coexpression Landscape: Confident Network Derived from Tissue Transcriptomic Profiles. PLoS One 2008;3:e3911. doi:10.1371/journal.pone.0003911.

[169] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nat Genet 2000;25:25–9. doi:10.1038/75556.

[170] Finn RD. Pfam: clans, web tools and services. Nucleic Acids Res 2006;34:D247–51. doi:10.1093/nar/gkj149.

[171] Rzhetsky A, Iossifov I, Koike T, Krauthammer M, Kra P, Morris M, et al. GeneWays: a system for extracting, analyzing, visualizing, and integrating molecular pathway data. J Biomed Inform 2004;37:43–53. doi:10.1016/j.jbi.2003.10.001.

[172] Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L, et al. Reactome: A database of reactions, pathways and biological processes. Nucleic Acids Res 2011;39:691–7. doi:10.1093/nar/gkq1018.

[173] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 1995;57:289–300. doi:0035-9246/95/57289.

[174] Bastian M, Heymann S, Jacomy M. Gephi: An Open Source Software for Exploring and Manipulating Networks. Third Int AAAI Conf Weblogs Soc Media 2009:361–2. doi:10.1136/qshc.2004.010033.

[175] Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4:44–57. doi:10.1038/nprot.2008.211.

[176] Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. New approach for understanding genome variations in KEGG. Nucleic Acids Res 2018. doi:10.1093/nar/gky962.

[177] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting

genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545–50. doi:10.1073/pnas.0506580102.

[178] Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, et al. The Reactome Pathway Knowledgebase. Nucleic Acids Res 2018;46:D649–55. doi:10.1093/nar/gkx1132.

[179] Charitou T, Bryan K, Lynn DJ. Using biological networks to integrate, visualize and analyze genomics data. Genet Sel Evol 2016;48:27. doi:10.1186/s12711-016-0205-1.

[180] Schaefer MH, Serrano L, Andrade-Navarro MA. Correcting for the study bias associated with protein-protein interaction measurements reveals differences between protein degree distributions from different cancer types. Front Genet 2015;6:260. doi:10.3389/fgene.2015.00260.

[181] Dang C V. c-Myc Target Genes Involved in Cell Growth, Apoptosis, and Metabolism. Mol Cell Biol 1999;19:1–11. doi:10.1128/MCB.19.1.1.

[182] Schlagbauer-Wadl H, Griffioen M, van Elsas A, Schrier PI, Pustelnik T, Eichler HG, et al. Influence of increased c-Myc expression on the growth characteristics of human melanoma. J Invest Dermatol 1999;112:332–6. doi:10.1046/j.1523-1747.1999.00506.x.

[183] Preston GC, Sinclair L V., Kaskar A, Hukelmann JL, Navarro MN, Ferrero I, et al. Single cell tuning of Myc expression by antigen receptor signal strength and interleukin-2 in T lymphocytes. EMBO J 2015;34:2008–24. doi:10.15252/embj.201490252.

[184] Kortlever RM, Sodir NM, Wilson CH, Burkhart DL, Pellegrinet L, Brown Swigart L, et al. Myc Cooperates with Ras by Programming Inflammation and Immune Suppression. Cell 2017;171:1301–1315.e14. doi:10.1016/j.cell.2017.11.013.

[185] Gnanaprakasam JNR, Sherman JW, Wang R. MYC and HIF in shaping immune response and immune metabolism. Cytokine Growth Factor Rev 2017;35:63–70. doi:10.1016/j.cytogfr.2017.03.004.

[186] Versteeg R, Noordermeer IA, Krüse-Wolters M, Ruiter DJ, Schrier PI. c-myc down-regulates class I HLA expression in human melanomas. EMBO J 1988;7:1023–9.

[187] Peltenburg LTC, Schrier PI. Transcriptional suppression of HLA-B expression by

c-Myc is mediated through the core promoter elements. Immunogenetics 1994;40:54–61. doi:10.1007/BF00163964.

[188] God JM, Cameron C, Figueroa J, Amria S, Hossain A, Kempkes B, et al. Elevation of c-MYC Disrupts HLA Class II–Mediated Immune Recognition of Human B Cell Tumors. J Immunol 2015;194:1434–45. doi:10.4049/jimmunol.1402382.

[189] Spranger S, Spaapen RM, Zha Y, Williams J, Meng Y, Ha TT, et al. Up-Regulation of PD-L1, IDO, and Tregs in the Melanoma Tumor Microenvironment Is Driven by CD8+ T Cells. Sci Transl Med 2013;5:200ra116-200ra116. doi:10.1126/scitranslmed.3006504.

[190] Zhang Q, Lenardo MJ, Baltimore D. 30 Years of NF-κB: A Blossoming of Relevance to Human Pathobiology. Cell 2017;168:37–57. doi:10.1016/j.cell.2016.12.012.

[191] Concetti J, Wilson CL. NFKB1 and Cancer: Friend or Foe? Cells 2018;7. doi:10.3390/cells7090133.

[192] Chen F, Castranova V. Nuclear factor-kappaB, an unappreciated tumor suppressor. Cancer Res 2007;67:11093–8. doi:10.1158/0008-5472.CAN-07-1576.

[193] Taniguchi K, Karin M. NF-κB, inflammation, immunity and cancer: coming of age. Nat Rev Immunol 2018. doi:10.1038/nri.2017.142.

[194] Tuijnenburg P, Allen L, Burns SO, Greene D, Chandra A, Kiani-alikhan S, et al. Loss-of-function nuclear factor k B subunit 1 (NFKB1) variants are the most common monogenic cause of common variable immunodeficiency in Europeans 2018;1:1–12.

[195] Muthuswamy R, Berk E, Junecko BF, Zeh HJ, Zureikat AH, Normolle D, et al. NF-B Hyperactivation in Tumor Tissues Allows Tumor-Selective Reprogramming of the Chemokine Microenvironment to Enhance the Recruitment of Cytolytic T Effector Cells. Cancer Res 2012;72:3735–43. doi:10.1158/0008-5472.CAN-11-4136.

[196] O'Reilly LA, Putoczki TL, Mielke LA, Low JT, Lin A, Preaudet A, et al. Loss of NF-κB1 Causes Gastric Cancer with Aberrant Inflammation and Expression of Immune Checkpoint Regulators in a STAT-1-Dependent Manner. Immunity 2018;48:570–583.e8. doi:10.1016/j.immuni.2018.03.003.

[197] Campbell PJ, Stephens PJ, Pleasance ED, O'Meara S, Li H, Santarius T, et al.

Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. Nat Genet 2008;40:722–9. doi:10.1038/ng.128.

[198] Bonnet E, Moutet M-L, Baulard C, Bacq-Daian D, Sandron F, Mesrob L, et al. Performance comparison of three DNA extraction kits on human whole-exome data from formalin-fixed paraffin-embedded normal and tumor samples. PLoS One 2018;13:e0195471. doi:10.1371/journal.pone.0195471.

[199] Eckhart L, Bach J, Ban J, Tschachler E. Melanin binds reversibly to thermostable DNA polymerase and inhibits its activity. Biochem Biophys Res Commun 2000;271:726–30. doi:10.1006/bbrc.2000.2716.

[200] Häcker H, Karin M. Regulation and function of IKK and IKK-related kinases. Sci STKE 2006;2006:re13. doi:10.1126/stke.3572006re13.

[201] Jain A, Kaczanowska S, Davila E. IL-1 Receptor-Associated Kinase Signaling and Its Role in Inflammation, Cancer Progression, and Therapy Resistance. Front Immunol 2014;5:553. doi:10.3389/fimmu.2014.00553.

[202] Thu YM, Richmond A. NF-κB inducing kinase: a key regulator in the immune system and in cancer. Cytokine Growth Factor Rev 2010;21:213–26. doi:10.1016/j.cytogfr.2010.06.002.

[203] Yamamoto T, Tsutsumi N, Tochio H, Ohnishi H, Kubota K, Kato Z, et al. Functional assessment of the mutational effects of human IRAK4 and MyD88 genes. Mol Immunol 2014;58:66–76. doi:10.1016/j.molimm.2013.11.008.

[204] Takaesu G, Surabhi RM, Park K-J, Ninomiya-Tsuji J, Matsumoto K, Gaynor RB. TAK1 is critical for IkappaB kinase-mediated activation of the NF-kappaB pathway. J Mol Biol 2003;326:105–15.

[205] Schmid JA, Birbach A. IkappaB kinase beta (IKKbeta/IKK2/IKBKB)--a key molecule in signaling to the transcription factor NF-kappaB. Cytokine Growth Factor Rev 2008;19:157–65. doi:10.1016/j.cytogfr.2008.01.006.

[206] Yamamoto Y, Gaynor RB. IkappaB kinases: key regulators of the NF-kappaB pathway. Trends Biochem Sci 2004;29:72–9. doi:10.1016/j.tibs.2003.12.003.

[207] Liu G, Park Y-J, Abraham E. Interleukin-1 receptor-associated kinase (IRAK) -1-mediated NF-kappaB activation requires cytosolic and nuclear activity. FASEB J 2008;22:2285–96. doi:10.1096/fj.07-101816.

[208] Rothwarf DM, Zandi E, Natoli G, Karin M. IKK-gamma is an essential regulatory subunit of the IkappaB kinase complex. Nature 1998;395:297–300. doi:10.1038/26261.

[209] Horvath CM. The Jak-STAT Pathway Stimulated by Interferon. Sci Signal 2004;2004:tr8-tr8. doi:10.1126/stke.2602004tr8.

[210] Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, et al. Pan-cancer patterns of somatic copy number alteration. Nat Genet 2013;45:1134–40. doi:10.1038/ng.2760.

[211] Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 2016;32:2847–9. doi:10.1093/bioinformatics/btw313.

[212] UniProt Consortium T. UniProt: the universal protein knowledgebase. Nucleic Acids Res 2018;46:2699. doi:10.1093/nar/gky092.

[213] Camp RL. X-Tile: A New Bio-Informatics Tool for Biomarker Assessment and Outcome-Based Cut-Point Optimization. Clin Cancer Res 2004;10:7252–9. doi:10.1158/1078-0432.CCR-04-0713.

[214] Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol 2011;12:R41. doi:10.1186/gb-2011-12-4-r41.

[215] Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, et al. Absolute quantification of somatic DNA alterations in human cancer. Nat Biotechnol 2012;30:413–21. doi:10.1038/nbt.2203.

[216] Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, et al. The landscape of somatic copy-number alteration across human cancers. Nature 2010;463:899–905. doi:10.1038/nature08822.

[217] Zaretsky JM, Garcia-Diaz A, Shin DS, Escuin-Ordinas H, Hugo W, Hu-Lieskovan S, et al. Mutations Associated with Acquired Resistance to PD-1 Blockade in Melanoma. N Engl J Med 2016;375:819–29. doi:10.1056/NEJMoa1604958.

[218] Shin DS, Zaretsky JM, Escuin-Ordinas H, Garcia-Diaz A, Hu-Lieskovan S, Kalbasi A, et al. Primary Resistance to PD-1 Blockade Mediated by JAK1/2 Mutations. Cancer Discov 2017;7:188–201. doi:10.1158/2159-8290.CD-16-1223.

[219] Dellgren C, Ekwelum VAC, Ormhøj M, Pallesen N, Knudsen J, Nehlin JO, et al. Low Constitutive Cell Surface Expression of HLA-B Is Caused by a Posttranslational Mechanism Involving Glu ^{180} and Arg ^{239}. J Immunol 2016;197:4807–16. doi:10.4049/jimmunol.1502546.

[220] Storr SJ, Safuan S, Mitra A, Elliott F, Walker C, Vasko MJ, et al. Objective assessment of blood and lymphatic vessel invasion and association with macrophage infiltration in cutaneous melanoma. Mod Pathol 2012;25:493–504. doi:10.1038/modpathol.2011.182.

[221] Newton-Bishop JA, Davies JR, Latheef F, Randerson-Moor J, Chan M, Gascoyne J, et al. 25-Hydroxyvitamin D2 /D3 levels and factors associated with systemic inflammation and melanoma survival in the Leeds Melanoma Cohort. Int J Cancer 2015;136:2890–9. doi:10.1002/ijc.29334.

[222] Timerman D, McEnery-Stonelake M, Joyce CJ, Nambudiri VE, Hodi FS, Claus EB, et al. Vitamin D deficiency is associated with a worse prognosis in metastatic melanoma. Oncotarget 2017;8:6873–82. doi:10.18632/oncotarget.14316.

[223] Fang S, Sui D, Wang Y, Liu H, Chiang Y-J, Ross MI, et al. Association of Vitamin D Levels With Outcome in Patients With Melanoma After Adjustment For C-Reactive Protein. J Clin Oncol 2016;34:1741–7. doi:10.1200/JCO.2015.64.1357.

[224] Gambichler T, Bindsteiner M, Höxtermann S, Kreuter A. Serum 25-hydroxyvitamin D serum levels in a large German cohort of patients with melanoma. Br J Dermatol 2013;168:625–8. doi:10.1111/j.1365-2133.2012.11212.x.

[225] Saiag P, Aegerter P, Vitoux D, Lebbé C, Wolkenstein P, Dupin N, et al. Prognostic Value of 25-hydroxyvitamin D3 Levels at Diagnosis and During Follow-up in Melanoma Patients. J Natl Cancer Inst 2015;107:djv264. doi:10.1093/jnci/djv264.

[226] Reichrath J, Rech M, Moeini M, Meese E, Tilgen W, Seifert M. In vitro comparison of the vitamin D endocrine system in 1,25(OH)2D3-responsive and -resistant melanoma cells. Cancer Biol Ther 2007;6:48–55.

[227] Amer M, Qayyum R. Relation between serum 25-hydroxyvitamin D and C-reactive protein in asymptomatic adults (from the continuous National Health and Nutrition Examination Survey 2001 to 2006). Am J Cardiol 2012;109:226–30.

doi:10.1016/j.amjcard.2011.08.032.

[228] O'Shea SJ, Davies JR, Newton-Bishop JA. Vitamin D, vitamin A, the primary melanoma transcriptome and survival. Br J Dermatol 2016;175 Suppl:30–4. doi:10.1111/bjd.14919.

[229] Jha P, Peto R. Global Effects of Smoking, of Quitting, and of Taxing Tobacco. N Engl J Med 2014;370:60–8. doi:10.1056/NEJMra1308383.

[230] Kessides MC, Wheless L, Hoffman-Bolton J, Clipp S, Alani RM, Alberg AJ. Cigarette smoking and malignant melanoma: A case-control study. J Am Acad Dermatol 2011;64:84–90. doi:10.1016/j.jaad.2010.01.041.

[231] Song F, Qureshi AA, Gao X, Li T, Han J. Smoking and risk of skin cancer: A prospective analysis and a meta-analysis. Int J Epidemiol 2012;41:1694–705. doi:10.1093/ije/dys146.

[232] Jones MS, Jones PC, Stern SL, Elashoff D, Hoon DSB, Thompson J, et al. The Impact of Smoking on Sentinel Node Metastasis of Primary Cutaneous Melanoma. Ann Surg Oncol 2017;24:2089–94. doi:10.1245/s10434-017-5775-9.

[233] Wyatt C, Lucas RM, Hurst C, Kimlin MG. Vitamin D deficiency at melanoma diagnosis is associated with higher Breslow thickness. PLoS One 2015;10:e0126394. doi:10.1371/journal.pone.0126394.

[234] Shiels MS, Katki HA, Freedman ND, Purdue MP, Wentzensen N, Trabert B, et al. Cigarette smoking and variations in systemic immune and inflammation markers. J Natl Cancer Inst 2014;106:1–8. doi:10.1093/jnci/dju294.

[235] Walser T, Cui X, Yanagawa J, Lee JM, Heinrich E, Lee G, et al. Smoking and lung cancer: the role of inflammation. Proc Am Thorac Soc 2008;5:811–5. doi:10.1513/pats.200809-100TH.

[236] Bauer M, Linsel G, Fink B, Offenberg K, Hahn AM, Sack U, et al. A varying T cell subtype explains apparent tobacco smoking induced single CpG hypomethylation in whole blood. Clin Epigenetics 2015;7:1–11. doi:10.1186/s13148-015-0113-1.

[237] Bauer M, Fink B, Thürmann L, Eszlinger M, Herberth G, Lehmann I. Tobacco smoking differently influences cell types of the innate and adaptive immune system—indications from CpG site methylation. Clin Epigenetics 2016;8:83. doi:10.1186/s13148-016-0249-7.

[238] Kõks S, Kõks G. Activation of GPR15 and its involvement in the biological effects of smoking. Exp Biol Med 2017;242:1207–12. doi:10.1177/1535370217703977.

[239] Schumacher TN, Schreiber RD. Neoantigens in cancer immunotherapy. Science (80-) 2015;348:69–74. doi:10.1126/science.aaa4971.

[240] Kleinschmidt I, Hills M, Elliott P. Smoking behaviour can be predicted by neighbourhood deprivation measures. J Epidemiol Community Health 1995;49 Suppl 2:S72-7.

[241] Passarelli MN, Newcomb PA. Survival Benefits of Smoking Cessation After Breast Cancer Diagnosis. JNCI Cancer Spectr 2017;1:19–20. doi:10.1093/jncics/pkx005.

[242] Osazuwa-Peters N, Adjei Boakye E, Chen BY, Tobo BB, Varvares MA. Association Between Head and Neck Squamous Cell Carcinoma Survival, Smoking at Diagnosis, and Marital Status. JAMA Otolaryngol Neck Surg 2017. doi:10.1001/jamaoto.2017.1880.

[243] Nagle CM, Bain CJ, Webb PM. Cigarette Smoking and Survival after Ovarian Cancer Diagnosis. Cancer Epidemiol Biomarkers Prev 2006;15:2557–60. doi:10.1158/1055-9965.EPI-06-0592.

[244] Foerster B, Pozo C, Abufaraj M, Mari A, Kimura S, D'Andrea D, et al. Association of Smoking Status With Recurrence, Metastasis, and Mortality Among Patients With Localized Prostate Cancer Undergoing Prostatectomy or Radiotherapy. JAMA Oncol 2018;4:953. doi:10.1001/jamaoncol.2018.1071.

[245] Ordóñez-Mena JM, Walter V, Schöttker B, Jenab M, O'Doherty MG, Kee F, et al. Impact of prediagnostic smoking and smoking cessation on colorectal cancer prognosis: A meta-analysis of individual patient data from cohorts within the CHANCES consortium. Ann Oncol 2018;29:472–83. doi:10.1093/annonc/mdx761.

[246] Tonstad S, Cowan JL. C-reactive protein as a predictor of disease in smokers and former smokers: a review. Int J Clin Pract 2009;63:1634–41. doi:10.1111/j.1742-1241.2009.02179.x.

[247] Lee J, Taneja V, Vassallo R. Cigarette Smoking and Inflammation. J Dent Res 2012;91:142–9. doi:10.1177/0022034511421200.

[248] Çolak Y, Afzal S, Lange P, Nordestgaard BG. Smoking, Systemic Inflammation, and Airflow Limitation: A Mendelian Randomization Analysis of 98 085 Individuals

from the General Population. Nicotine Tob Res 2018. doi:10.1093/ntr/nty077.

[249] Savin Z, Kivity S, Yonath H, Yehuda S. Smoking and the intestinal microbiome. Arch Microbiol 2018;200:677–84. doi:10.1007/s00203-018-1506-2.

[250] Gopalakrishnan V, Spencer CN, Nezi L, Reuben A, Andrews MC, Karpinets T V, et al. Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. Science 2018;359:97–103. doi:10.1126/science.aan4236.

[251] Wang H, Zhao J-X, Hu N, Ren J, Du M, Zhu M-J. Side-stream smoking reduces intestinal inflammation and increases expression of tight junction proteins. World J Gastroenterol 2012;18:2180–7. doi:10.3748/wjg.v18.i18.2180.

[252] Biedermann L, Zeitz J, Mwinyi J, Sutter-Minder E, Rehman A, Ott SJ, et al. Smoking cessation induces profound changes in the composition of the intestinal microbiota in humans. PLoS One 2013;8:e59260. doi:10.1371/journal.pone.0059260.

[253] Sopori M. Science and society: Effects of cigarette smoke on the immune system. Nat Rev Immunol 2002;2:372–7. doi:10.1038/nri803.

[254] Sopori ML, Kozak W, Savage SM, Geng Y, Soszynski D, Kluger MJ, et al. Effect of nicotine on the immune system: Possible regulation of immune responses by central and peripheral mechanisms. Psychoneuroendocrinology 1998;23:189–204. doi:10.1016/S0306-4530(97)00076-0.

[255] Kalra R, Singh SP, Pena-Philippides JC, Langley RJ, Razani-Boroujerdi S, Sopori ML. Immunosuppressive and anti-inflammatory effects of nicotine administered by patch in an animal model. Clin Diagn Lab Immunol 2004;11:563–8. doi:10.1128/CDLI.11.3.563-568.2004.

[256] Scott A, Lugg ST, Aldridge K, Lewis KE, Bowden A, Mahida RY, et al. Pro-inflammatory effects of e-cigarette vapour condensate on human alveolar macrophages. Thorax 2018;73:1161–9. doi:10.1136/thoraxjnl-2018-211663.

[257] Martin EM, Clapp PW, Rebuli ME, Pawlak EA, Glista-Baker E, Benowitz NL, et al. E-cigarette use results in suppression of immune and inflammatory-response genes in nasal epithelial cells similar to cigarette smoke. Am J Physiol Cell Mol Physiol 2016;311:L135–44. doi:10.1152/ajplung.00170.2016.

[258] Kõks G, Uudelepp ML, Limbach M, Peterson P, Reimann E, Kõks S. Smoking-induced expression of the GPR15 gene indicates its potential role in chronic

inflammatory pathologies. Am J Pathol 2015;185:2898–906. doi:10.1016/j.ajpath.2015.07.006.

[259] Gabay C. Interleukin-6 and chronic inflammation. Arthritis Res Ther 2006;8 Suppl 2:S3. doi:10.1186/ar1917.

[260] Fonseca JE, Santos MJ, Canhão H, Choy E. Interleukin-6 as a key player in systemic inflammation and joint destruction. Autoimmun Rev 2009;8:538–42. doi:10.1016/j.autrev.2009.01.012.

[261] Kumari N, Dwarakanath BS, Das A, Bhatt AN. Role of interleukin-6 in cancer progression and therapeutic resistance. Tumour Biol 2016;37:11553–72. doi:10.1007/s13277-016-5098-7.

[262] Hoejberg L, Bastholt L, Schmidt H. Interleukin-6 and melanoma. Melanoma Res 2012;22:327–33. doi:10.1097/CMR.0b013e3283543d72.

[263] Lyu G-Y, Yeh Y-H, Yeh Y-C, Wang Y-C. Mutation load estimation model as a predictor of the response to cancer immunotherapy. NPJ Genomic Med 2018;3:12. doi:10.1038/s41525-018-0051-x.

[264] Spranger S, Luke JJ, Bao R, Zha Y, Hernandez KM, Li Y, et al. Density of immunogenic antigens does not explain the presence or absence of the T-cell-inflamed tumor microenvironment in melanoma. Proc Natl Acad Sci U S A 2016;113:E7759–68. doi:10.1073/pnas.1609376113.

[265] Kokkat TJ, Patel MS, McGarvey D, LiVolsi VA, Baloch ZW. Archived formalin-fixed paraffin-embedded (FFPE) blocks: A valuable underexploited resource for extraction of DNA, RNA, and protein. Biopreserv Biobank 2013;11:101–6. doi:10.1089/bio.2012.0052.

[266] Evers DL, Fowler CB, Cunningham BR, Mason JT, O'Leary TJ. The effect of formaldehyde fixation on RNA: optimization of formaldehyde adduct removal. J Mol Diagn 2011;13:282–8. doi:10.1016/j.jmoldx.2011.01.010.

[267] Hoffman EA, Frey BL, Smith LM, Auble DT. Formaldehyde crosslinking: A tool for the study of chromatin complexes. J Biol Chem 2015;290:26404–11. doi:10.1074/jbc.R115.651679.

[268] Espina V, Heiby M, Pierobon M, Liotta LA. Laser capture microdissection technology. Expert Rev Mol Diagn 2007;7:647–57. doi:10.1586/14737159.7.5.647.

[269] Emmert-Buck MR, Bonner RF, Smith PD, Chuaqui RF, Zhuang Z, Goldstein SR, et al. Laser Capture Microdissection. Science (80-) 1996;274:998–1001. doi:10.1126/science.274.5289.998.

[270] Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun 2013;4:2612. doi:10.1038/ncomms3612.

[271] Zhang H, Chen J. Current status and future directions of cancer immunotherapy. J Cancer 2018;9:1773–81. doi:10.7150/jca.24577.

[272] Ribas A, Wolchok JD. Cancer immunotherapy using checkpoint blockade. Science (80-) 2018;359:1350–5. doi:10.1126/science.aar4060.

[273] Dunn GP, Bruce AT, Ikeda H, Old LJ, Schreiber RD. Cancer immunoediting: from immunosurveillance to tumor escape. Nat Immunol 2002;3:991–8. doi:10.1038/ni1102-991.

[274] Finotello F, Trajanoski Z. Quantifying tumor-infiltrating immune cells from transcriptomics data. Cancer Immunol Immunother 2018;67:1031–40. doi:10.1007/s00262-018-2150-z.

[275] Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The Immune Landscape of Cancer. Immunity 2018;48:812–830.e14. doi:10.1016/j.immuni.2018.03.023.

[276] Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. Cell 2015;160:48–61. doi:10.1016/j.cell.2014.12.033.

[277] Seliger B. Molecular mechanisms of HLA class I-mediated immune evasion of human tumors and their role in resistance to immunotherapies. Hla 2016;88:213–20. doi:10.1111/tan.12898.

[278] Jerby-Arnon L, Shah P, Cuoco MS, Rodman C, Su MJ, Melms JC, et al. A Cancer Cell Program Promotes T Cell Exclusion and Resistance to Checkpoint Blockade. Cell 2018;175:984–997.e24. doi:10.1016/j.cell.2018.09.006.

[279] Spranger S, Gajewski TF. Tumor-intrinsic oncogene pathways mediating immune avoidance. Oncoimmunology 2016;5:e1086862. doi:10.1080/2162402X.2015.1086862.

[280] Casey SC, Tong L, Li Y, Do R, Walz S, Fitzgerald KN, et al. MYC regulates the

antitumor immune response through CD47 and PD-L1. Science (80-) 2016;352:227–31. doi:10.1126/science.aac9935.

[281] Peng W, Chen JQ, Liu C, Malu S, Creasy C, Tetzlaff MT, et al. Loss of PTEN Promotes Resistance to T Cell-Mediated Immunotherapy. Cancer Discov 2016;6:202–16. doi:10.1158/2159-8290.CD-15-0283.

[282] Chen H, Liu H, Qing G. Targeting oncogenic Myc as a strategy for cancer treatment. Signal Transduct Target Ther 2018;3:5. doi:10.1038/s41392-018-0008-7.

[283] Balachandran S, Porosnicu M, Barber GN. Oncolytic activity of vesicular stomatitis virus is effective against tumors exhibiting aberrant p53, Ras, or myc function and involves the induction of apoptosis. J Virol 2001;75:3474–9. doi:10.1128/JVI.75.7.3474-3479.2001.

[284] Pastorino F, Brignole C, Marimpietri D, Di Paolo D, Zancolli M, Pagnan G, et al. Targeted delivery of oncogene-selective antisense oligonucleotides in neuroectodermal tumors: therapeutic implications. Ann N Y Acad Sci 2004;1028:90–103. doi:10.1196/annals.1322.010.

[285] Hopewell EL, Zhao W, Fulp WJ, Bronk CC, Lopez AS, Massengill M, et al. Lung tumor NF-kB signaling promotes T cell–mediated immune surveillance. J Clin Invest 2013;123:2509–22. doi:10.1172/JCI67250DS1.

[286] Grinberg-bleyer Y, Oh H, Desrichard A, Klein U, Hayden MS, Ghosh S, et al. NF- k B c-Rel Is Crucial for the Regulatory T Cell Immune Checkpoint in Cancer Article NF- k B c-Rel Is Crucial for the Regulatory T Cell Immune Checkpoint in Cancer. Cell n.d.;170:1096–1100.e13. doi:10.1016/j.cell.2017.08.004.

[287] Madonna G, Ullman CD, Gentilcore G, Palmieri G, Ascierto PA. NF-κB as potential target in the treatment of melanoma. J Transl Med 2012;10:53. doi:10.1186/1479-5876-10-53.

[288] Pai SG, Carneiro BA, Mota JM, Costa R, Leite CA, Barroso-Sousa R, et al. Wnt/beta-catenin pathway: Modulating anticancer immune response. J Hematol Oncol 2017;10:1–12. doi:10.1186/s13045-017-0471-6.

[289] Tanaka A, Sakaguchi S. Regulatory T cells in cancer immunotherapy. Cell Res 2017;27:109–18. doi:10.1038/cr.2016.151.

[290] Ha T-Y. The Role of Regulatory T Cells in Cancer. Immune Netw 2009;9:209.

doi:10.4110/in.2009.9.6.209.

[291] Umansky V, Blattner C, Gebhardt C, Utikal J. The Role of Myeloid-Derived Suppressor Cells (MDSC) in Cancer Progression. Vaccines 2016;4:36. doi:10.3390/vaccines4040036.

[292] Umansky V, Sevko A, Gebhardt C, Utikal J. Myeloid-derived suppressor cells in malignant melanoma. JDDG J Der Dtsch Dermatologischen Gesellschaft 2014;12:1021–7. doi:10.1111/ddg.12411.

[293] Obeid JM, Erdag G, Smolkin ME, Deacon DH, Patterson JW, Chen L, et al. PD-L1, PD-L2 and PD-1 expression in metastatic melanoma: Correlation with tumor-infiltrating immune cells and clinical outcome. Oncoimmunology 2016;5:1–12. doi:10.1080/2162402X.2016.1235107.

[294] Karachaliou N, Gonzalez-Cao M, Crespo G, Drozdowskyj A, Aldeguer E, Gimenez-Capitan A, et al. Interferon gamma, an important marker of response to immune checkpoint blockade in non-small cell lung cancer and melanoma patients. Ther Adv Med Oncol 2018;10:1758834017749748. doi:10.1177/1758834017749748.

[295] Gao J, Shi LZ, Zhao H, Chen J, Xiong L, He Q, et al. Loss of IFN-γ Pathway Genes in Tumor Cells as a Mechanism of Resistance to Anti-CTLA-4 Therapy. Cell 2016;167:397–404.e9. doi:10.1016/j.cell.2016.08.069.

[296] Bach EA, Aguet M, Schreiber RD. The IFN gamma receptor: a paradigm for cytokine receptor signaling. Annu Rev Immunol 1997;15:563–91. doi:10.1146/annurev.immunol.15.1.563.