Establishing the Role of Vitamin D signalling in Immunity and Melanoma Specific Survival

Sathya Muralidhar

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

The University of Leeds
Leeds Institute of Medical Research at St James's
School of Medicine

February 2019

The candidate confirms that the work submitted is 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. Further details of the jointly-authored publications and the contributions of the candidate and the other authors to the work should be included below this statement.

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 thank the following members from their respective institutions for their contributions to my PhD research. The specific contributions of each member to my research projects are listed at the beginning of the relevant Chapters in this thesis.

I extend my sincere gratitude to the participants and family members of the Leeds Melanoma Cohort, without whom my research would not have been possible.

I would like to thank all members of the department of Epidemiology and Biostatistics, who have been an integral part of my time as PhD student. I would especially like to mention:

- My primary supervisors Prof. Julia Newton-Bishop and Prof. Tim Bishop for providing the guidance, supervision and encouragement that made my PhD possible.
- Dr Jeremie Nsengimana for his statistical guidance and abundant patience while teaching me.
- Dr Juliette Randerson-Moor for her constant support, guidance, for bringing the Doctor to my notice and for her assistance with formatting this thesis.
- Dr Jonathan Laye for his support and extensive contributions to the immunohistochemistry work performed in this thesis.
- Dr Mark Harland for his guidance in the design of RT-PCR experiments
- Ms Tracey Mell for her assistance with Immunohistochemistry
- My fellow PhD students Mr. Joey Diaz, Mr. Rohit Thakur, Mr. Adam Trower and Ms. Joanna Pozniak for being a wonderful peer group.
- Particular mention to Ms. Joanna Pozniak for being my ‘PhD support system’ and a friend.
- Dr John Davies, Dr Mark Iles and Prof. Jenny Barrett for their insightful comments and suggestions during lab meetings and journal clubs, especially during the early days of my PhD.
- Ms. May Chan and Mrs. Joanne Gascoyne for helping with my queries and providing data access
- Mrs Katie Cairns for her assistance with paperwork, especially during the early days of my PhD

I would also like to thank the Human Resources team: Mrs. Sharon Collins, Mrs. Sharon Pinder and Mrs. Wendy Kennedy, who have been very supportive in providing assistance with regards to my visa and contracts.

I would like to thank the members of the MELGEN consortium, especially my fellow MELGEN PhD students, who have been an important support system.

I would like to extend my thanks to Dr David Adams and Dr Louise van der Weyden from the Experimental Cancer Genetic team at the Wellcome Sanger Institute for their support and patience in designing and executing the *in vitro* and *in vivo* experiments. I would like to thank the other members of the Experimental Cancer Genetic team for their help and for making my time at the Sanger an enjoyable one.

I would like to thank Eagle genomic (the industrial partners of the MELGEN consortium) who hosted me for a secondment, which in addition to being a useful experience, was instrumental in my consideration of a career in non-academic research.

I would like to thank my family, both old and new: the former for being the pillar of strength behind my journey so far and the latter for choosing to be part of my journey ahead.

Finally, I would like to specially thank Prof. Julia Newton-Bishop and Prof. Tim Bishop for their roles beyond my supervisors: for their trust, mentorship and for support with more things than can be listed.

Abstract

1α,25-dihydroxyvitamin D_3 signals via its canonical nuclear receptor: Vitamin D Receptor (VDR). While higher levels of serum vitamin D have been reported to be associated with thinner primary melanomas and better outcome, increased VDR expression has been associated with decreased tumour progression and improved prognosis in melanoma primaries. However, the genomic basis of this effect remains to be explored and a causal mechanism is yet to be established. To address this question, I have used microarray data from a cohort of 703 treatment-naïve primary melanomas from the Leeds Melanoma Cohort (LMC) and corresponding clinical data.

In the LMC primary melanomas, serum vitamin D was not significantly associated with melanoma survival. However, tumour *VDR* expression was significantly (and independently) protective for melanoma death in both the LMC and the TCGA metastatic melanoma datasets. Tumour *VDR* expression was found to be significantly positively correlated with genes enriched for ECM organization, TNF signalling, IFNγ signalling, IL12-mediated signalling and NFκB signalling, which are predominantly immune-related. Concordantly, *VDR* expression was lower in tumours graded by the pathologist as having no immune infiltrate, compared to tumours with brisk and non-brisk immune infiltrate. Additionally, *VDR* correlated positively with imputed immune cells scores. Conversely, the negatively correlated genes were enriched for Mitotic Prophase, Wnt signalling pathway, Mitochondrial translation, citric acid cycle and oxidative phosphorylation, which are predominantly proliferation-related. Of particular interest among the negatively correlated pathways was the Wnt/β-catenin signalling pathway. Functional validation using an *in vivo* tail-vein metastasis assay revealed that murine melanoma cells stably expressing VDR produced significantly fewer pulmonary metastases compared to control cells with null VDR expression. VDR-expressing cells also had significantly lower expression of Wnt/β-catenin signalling genes compared to control cells. These findings indicate that vitamin D-VDR signalling contributes to control of pro-proliferative and immunosuppressive Wnt/β-catenin signalling in melanoma and that this is associated with less proliferative, less metastatic disease and stronger host immune responses.

Table of Contents

List of Tables

List of Figures

Abbreviations

VDR	Vitamin D Receptor
UV	Ultra Violet
DALY	Disability Adjusted Life Year
IARC	International Agency for Cancer Research
WHO	World Health Organisation
ASR	Age Standardised Rate
AJCC	American Joint Committee on Cancer
TNM	Tumour Node Metastasis
GWAS	Genome Wide Association Studies
UVA	Ultraviolet-A
UVB	Ultraviolet-B
WES	Whole Exome Sequencing
TCGA	The Cancer Genome Atlas
DNS	Dysplastic Naevus Syndrome
ROS	Reactive Oxygen Species
RHC	Red Hair Colour
NER	Nucleotide Excision Repair
TME	Tumour Microenvironment
ECM	Extracellular Matrix
CAF	Cancer Associated Fibroblasts
TILs	Tumour Infiltrating Lymphocytes
MHC	Major Histocompatibility Complex
NK cells	Natural killer cells
MDSC	Myeloid Derived Suppressor Cell
MMP	Matrix Metalloproteinase
FDA	Food and Drug Administration
CTL	Cytotoxic T Lymphocyte
T-reg	T-regulatory cells
CYP	Cytochrome P450
VDRE	Vitamin D Response Element
7-DHC	7-dehydrocholesterol
RXR	Retinoid X Receptor
NHR	Nuclear Hormone Receptor
hVDR	Human Vitamin D Receptor

DBD	DNA Binding Domain
LBD	Ligand Binding Domain
HRE	Hormone Receptor Elements
nVDRE	negative Vitamin D Response Element
ChIP-Seq	Chromatin Immunoprecipitation-Sequencing
APC	Antigen Presenting Cells
TLR	Toll Like Receptor
CRP	C-Reactive Protein
NMSC	Non Melanoma Skin Cancer
MSS	Melanoma Specific Survival
LMC	Leeds Melanoma Cohort
ONS	Office of National Statistics
PIAG	Patient Advisory Group
MREC	Multicentre Research Ethics Committee
FFPE	Formalin Fixed Paraffin Embedded
TMA	Tissue Microarray
mRNA	messenger Ribonucleic acid
SD	Standard Deviation
mm (suffix)	Millimetre
H&E	Haematoxylin and Eosin
DASL	cDNA-mediated Annealing Selection Extension and Ligation
CNA	Copy Number Alteration
Reactome FI network	Reactome Functional interaction network
DIP	Database of Interacting Proteins
MINT	Molecular Interaction Database
GEO	Gene Expression Omnibus
KEGG	Kyoto Encyclopaedia of Genes and Genomes
PANTHER	Protein Analysis Through Evolutionary Relationship
FDR	False Discovery Rate
SLE	Systemic Lupus Erythematosus
LC-MS/MS	Liquid Chromatography-Mass Spectrometry
HR	Hazard Ratio
FLG2	Fillagrin-2
GREAT	Genome Regions Enrichment of Annotations Tool
NR1L	Nuclear Receptor-1L
SACN	Scientific Advisory Committee on Nutrition

IHC	Immunohistochemistry
nmol/L (suffix)	nano moles per litre
CIC	Consensus Immunome Cluster
NLR	Nuclear Ligand Receptor
cDNA	Complementary Deoxyribonucleic acid
qRT-PCR	quantitative Real Time-Polymerase Chain Reaction
BSA	Bovine Serum Albumin
IRES	Internal Ribosomal Entry Site
PB	*Piggybac* Transposon
SOC	Super Optimal Broth
LB	Lysogeny Broth
Pen-Strep-Glutamine	Penicillin-Streptomycin-Glutamin
Opti-MEM	Opti-Minimum Eagles Medium
DMEM	Dulbecco's Minimum Eagles Medium
FBS	Fetal Bovine Serum
PBS	Phosphate Buffered Saline
MeOH	Methanol
TBST	Tris-buffered saline-Tween
HRP	Horse Radish Peroxidase
ECL	Enhanced Chemiluminescence
SDS	Sodium Dodecyl Sulphate
MOPS	3-(N-morppholino)propanesulfanoic acid
μm^2 (suffix)	square micrometre
mm^2 (suffix)	square millimetre
QS5	Quantstudio 5
FC	Fold Change
ITR	Inverted Terminal Repeats
µg (suffix)	microgram
kDa (suffix)	kilo dalton
COE	Consumption of Epidermis
SCF	Sub-epidermal Cleft Formation
RFS	Relapse Free Survival
OS	Overall Survival
DMFS	Distant Metastasis Free Survival
IFN/PEG-IFN	Interferon/pegylated Interferon

Definition of terms

Leeds Group: Refers to the members (past and present) of the Section of Epidemiology and Biostatistics at the University of Leeds

Breslow thickness: the measurement of Breslow thickness is described as: *"Breslow thickness is measured from the top of the granular layer of the epidermis (or, if the surface is ulcerated, from the base of the ulcer) to the deepest invasive cell across the broad base of the tumour (dermal/subcutaneous) as described by Breslow."* (1, 2).

Mitotic rate: Mitotic rate is defined as the number of tumour cells which are mitotically active per square millimetre of tumour area, as estimated by the reviewing pathologist.

Ulceration: melanoma ulceration is defined as : "*full thickness absence of an intact epidermis above any portion of the primary tumour with an associated host reaction (characterized by a fibrinous and acute inflammatory exudate) above the primary tumour based on histopathological examination."* (2).

Age: Age at diagnosis, in years.

Sex: Sex at diagnosis, indicated as Male or Female (self-reported in questionnaire)

AJCC stage: Classification of melanomas according to the recommendations of the 7th edition of the American Joint Committee on Cancer

Vascular invasion: vascular invasion indicates the presence (or absence) of tumour cells which are fixed to the walls and within the lumens of lymphatic or blood vessels (3).

Melanoma survival/death: survival information for participants was obtained both directly (by annual re-contact) and indirectly from review of national cancer registries and the ONS. In the case of deceased participants, the cause of death was obtained from death certificates and medical records. This was reviewed by research nurses in the Leeds Melanoma Research group led Prof. Julia Newton-Bishop, to generate Melanoma Specific Survival (MSS).

…

Chapter 1
Introduction

Melanoma is a type of skin cancer, originating from the melanocytes, which are neural crest-derived cells residing in the epidermis. The transformation of melanocyte to a melanoma malignant phenotype involves a series of genomic and molecular events, which lead to aberrant signalling in pathways pertaining to cell cycle arrest, melanocyte development and differentiation, immune response and DNA-damage repair. Extensive research including studies of familial susceptibility, genome-wide association, mutation burden, *in vitro* and *in vivo* models have contributed to the understanding the hierarchy of molecular and genetic events involved in melanoma initiation and progression (4). These studies demonstrate that while melanoma initiation mandates genetic predisposition and/or somatic mutations, melanoma progression requires additional mutation and/or copy number variation events, which eventually culminate in disruption of the above mentioned cellular processes (5). Melanoma development and progression is also influenced by various host factors, which contribute to disease development and progression.

Chapter Aim

This chapter aims to provide an overview of the disease that is melanoma. This includes discussion of melanoma incidence and mortality, melanoma aetiology, the molecular pathogenesis of melanoma, role of host factors in melanoma development and melanoma therapy. This chapter also includes discussion of vitamin D-VDR signalling, focusing on the functional relevance of this signalling axis in human disease, in cancer and in melanoma itself.

1.1 Melanoma incidence

Melanoma of the skin is the 15th most commonly occurring cancer worldwide but is more common in the UK being the 5th most common (6). Among the different skin cancers, malignant melanoma accounts for the majority of the skin cancer deaths, despite representing less than 5% of all cutaneous malignancies (7, 8). There has been an increase in the worldwide incidence for melanoma in the past decades (7-

9). Incidence rates have been shown to vary with ethnicity, geographical location, age and sex, all of which are discussed below.

Melanoma is a cancer attributed to uncontrolled division of melanocytes: the pigment-producing cells in the skin, which is the human organ most exposed to sun and UV radiation. Thus, it is to be expected that melanoma incidence varies as a function of geography, skin-type and ethnicity. The role of sun/UV exposure, pigmentation and skin type in melanoma aetiology is discussed in detail in the sections below. In the case of ethnicity, melanoma incidence varies across different ethnicities more than do most cancers (8). Melanoma burden is highest in regions where the population at risk is pale-skinned: Australasia (Australia, New Zealand), North America and Europe (10), but living in a hot country or with access, usually on holiday to hot countries (as of 2012 estimates, represented in Figure 1.1). Even among people of the same ethnicity, melanoma incidence has been shown to vary by geographical region: melanoma incidence increased with decreasing latitude in North America and England (11) and Australia (12). Within Europe, melanoma incidence is however higher in Scandinavian countries compared to southern countries like Spain and Italy (8). These differences have been attributed to pale-skinned and olive-skinned populations in the north and south of Europe respectively (13).

Since skin type, ethnicity and geography are strongly linked to each other, studies using ethnically heterogeneous populations, but in the same geographical region, have offered better insight into delineating the effects of these factors. For instance, a pooled study reported that melanoma incidence increased with higher ultraviolet (UV) indices[i] and lower altitude only in non-Hispanic whites: not in black or Hispanic populations in the United States (15). Similarly, within countries whose population consists of heterogeneous ethnicities, melanoma rates are highest among the palest skinned, while incidence is lower among people of darker-skinned ethnicities (8, 16). Taken together, melanoma is predominantly incident in white skinned populations, but geographic location is also a significant factor.

[i] Ultraviolet (UV) index: The Global Solar UV Index is a measure of incident solar UV radiation levels on the earth's surface. Known commonly as the 'UV index', this measure includes values of zero and upward, with higher indices being associated with greater potential damage to the skin and eye (14. WHO. 2002.)

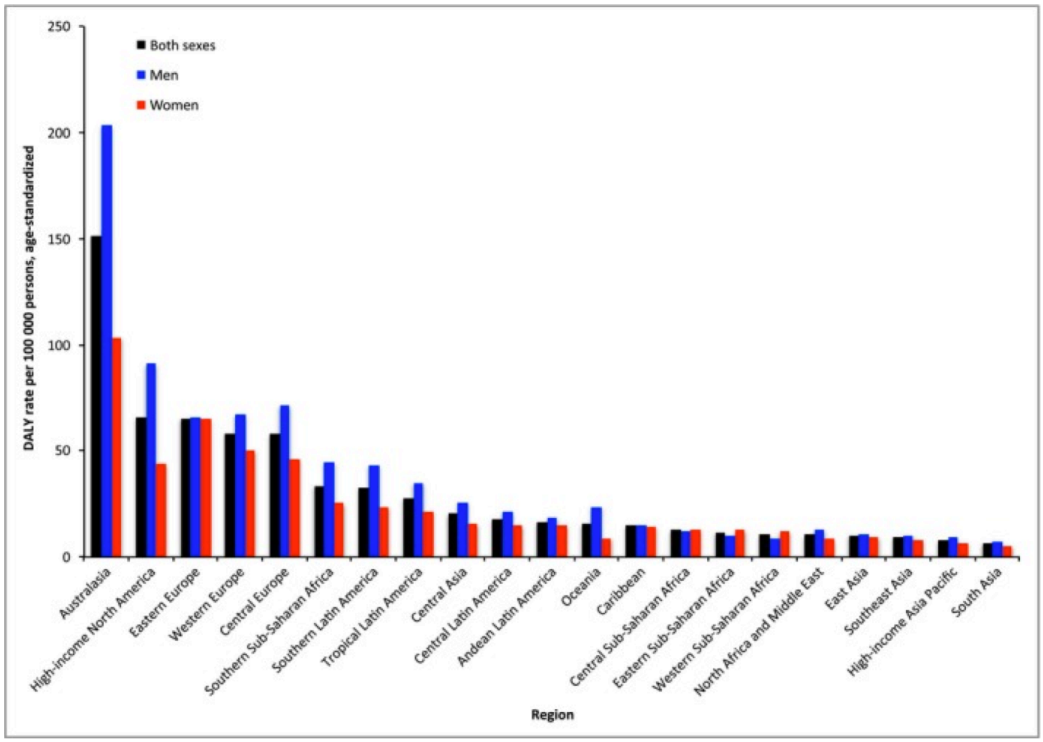


Figure 1.1 Worldwide incidence of melanoma sourced from GLOBOCAN 2012 estimates

Disability Adjusted Life Year (DALY) is a summary measure used to give an indication of the burden of disease. One DALY represents the loss of the equivalent of one year of full health. Source: international Agency for cancer Research-World Health Organisation (IARC-WHO) online Glossary of Terms: http://www-dep.iarc.fr/WHOdb/glossary.htm

Melanoma incidence also increases with age, with the trend being observed in high risk populations such as Australia, New Zealand and Northern Europe (8) (depicted in Figure 1.2). Even though melanoma incidence is lower in the younger population (<40 years) however, it is still the most commonly diagnosed cancer among young adults (17, 18).

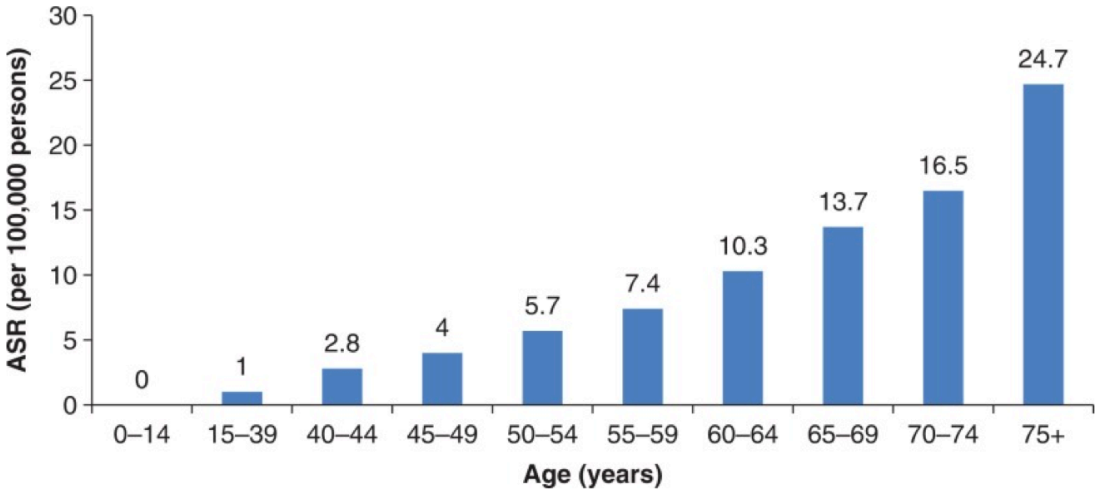


Figure 1.2 Worldwide incidence of melanoma by age, as per GLOBOCAN 2012 estimates

Age standardized Rate (ASR) is a summary measure of the rate that a population would have if it had a standard age structure. Source: international Agency for cancer Research-World Health Organisation (IARC-WHO) online Glossary of Terms: http://www-dep.iarc.fr/WHOdb/glossary.htm

Worldwide melanoma incidence varies between men and women (depicted in Figure 1.3). At age over 40 years, melanoma incidence is greater in men than in women, worldwide (8, 19) and in high incidence populations such as United States, Australia and New Zealand (8, 20). This pattern of higher incidence in men is consistent even across different ethnicities, for instance: in the United states, melanoma incidence is higher in males of non-Hispanic Caucasian, Asian/Pacific Islander and African American ethnicities, compared to their respective female populations (21). This suggests that the male-female disparity in incidence is not confounded by geography or ethnic background. It has been postulated that this differential incidence could be reflections of androgen-related effects (22, 23). However, the current excess of melanoma in males in the UK is a new phenomenon. As recently as 2006, melanoma in the UK was more common in women and this had been the case since the incidence started to rise at the beginning of the 20th century (24). The change in sex incidence over time suggests that it is more likely behavioural change is responsible for these sex differences than hormonal factors.

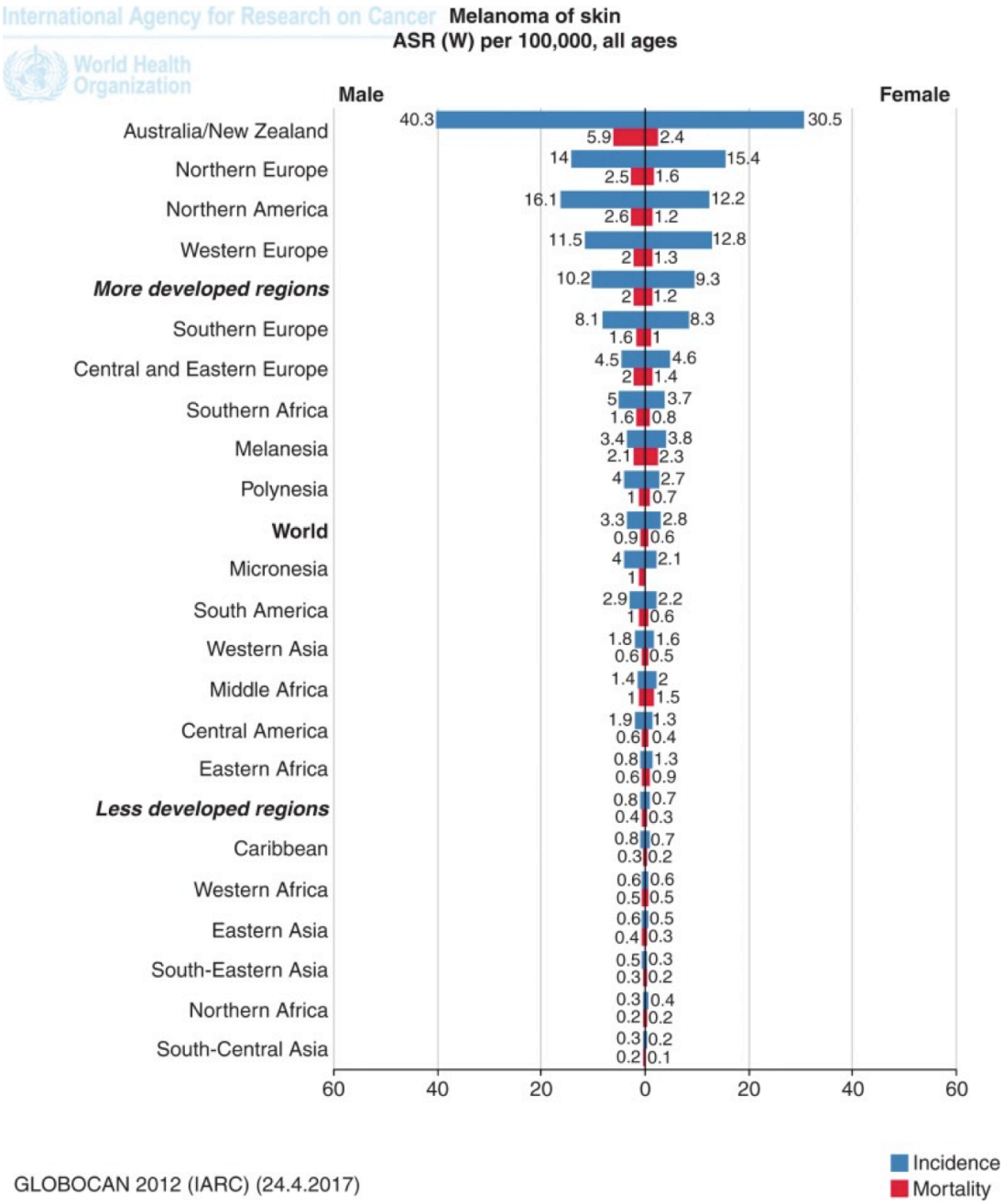


Figure 1.3 Worldwide incidence of melanoma by sex, as per GLOBOCAN 2012 estimates

Age standardized Rate (ASR) is a summary measure of the rate that a population would have if it had a standard age structure. Source: international Agency for cancer Research-World Health Organisation (IARC-WHO) online Glossary of Terms: http://www-dep.iarc.fr/WHOdb/glossary.htm

1.2 **Melanoma mortality**

Much like melanoma incidence, trends in melanoma mortality vary by geography, ethnicity, age and sex. The worldwide melanoma mortality is depicted in Figure 1.4, as per 2012 estimates. Melanoma mortality has steadily increased in the past decade, particularly in pale-skinned populations living in high-risk countries such as Australia and New Zealand (25). Similar trends have been reported in Scandinavian countries and the United Kingdom (25) and also in East Asian populations (26). Within an ethnically heterogenous population, non-white subgroups have higher mortality rates compared to their white counterparts, despite lower incidence reported in these subgroups. For instance, non-Hispanic African Americans have lower 5-year survival rates compared to white subgroups in the United States (27), with the discrepancy being attributed by some to socioeconomic inequalities (28). The type of melanoma suffered by non-white patients however is the histologically defined acral lentiginous melanoma (29), which is one of the melanoma subtypes with worse prognosis compared to cutaneous melanoma (30). This could also explain the higher melanoma mortality in non-white populations.

Similar to melanoma incidence, mortality is higher in males compared to females, across all races (31).This difference is also reported to be significant at all stages of the disease: even advanced stage IV (32). Annual melanoma mortality is also highest in individuals aged >70 worldwide (31): that is that increased age increases the risk of dying for every individual case.

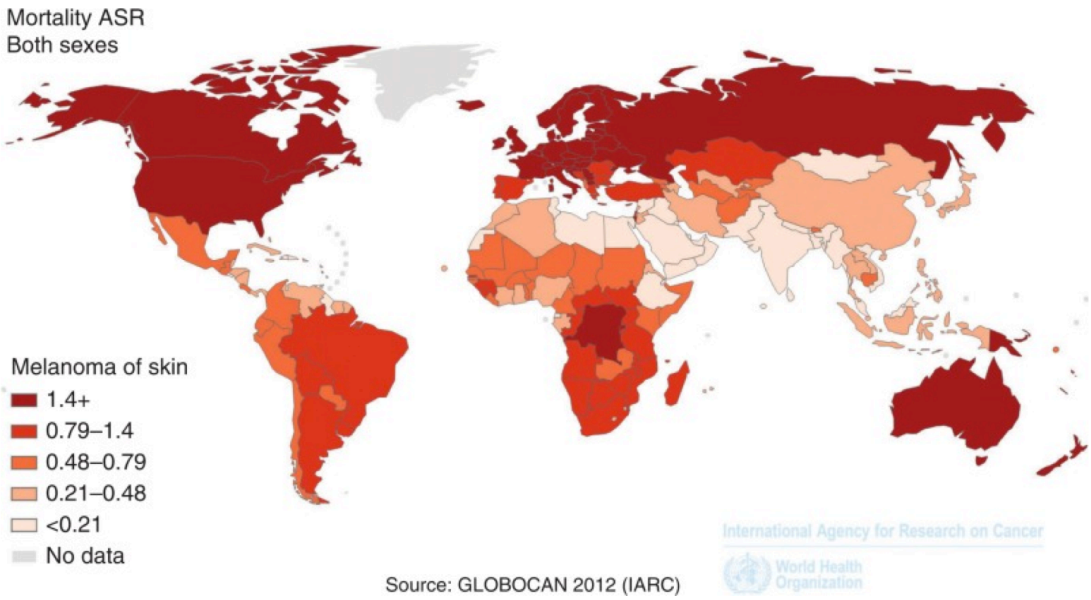


Figure 1.4 Worldwide mortality from melanoma, as per GLOBOCAN 2012 estimates

Age standardized Rate (ASR) is a summary measure of the rate that a population would have if it had a standard age structure. Source: international Agency for cancer Research-World Health Organisation (IARC-WHO) online Glossary of Terms: http://www-dep.iarc.fr/WHOdb/glossary.htm

1.3 Melanoma staging system: Factors predicting melanoma outcome

The melanoma staging system includes a combination of characteristics which have a significant effect on melanoma prognosis. The first multivariate analysis to identify these characteristics was based on insights from multiple institutions. It was published in 1981 and evaluated the effect of the following characteristics on melanoma prognosis: tumour thickness, Clark's level of invasion into the skin, number of mitoses, growth pattern, cell type, inflammatory reaction, vascular invasion and microscopic ulceration. Of these characteristics, tumour thickness was found to be the most significant predictor of prognosis, while ulceration and number of mitoses remained significant predictors after adjusting for tumour thickness (33). The significance of formal staging in estimating patient prognosis was evident and put forth the necessity of a unified staging system. This led to the formation of the American Joint Committee on Cancer (AJCC) in 1998 comprised of experts from North America, Europe and Australia. The AJCC set up a melanoma staging

database, for the continued collection and review of melanoma outcome data (34). The consensus staging system developed by the committee was subject to the following criteria: a) it should be evidence-based and reflect prognostic factors identified by multivariate Cox regression analyses, b) should be based on melanoma outcome results from multiple institutions and countries and c) should be practical and readily reproducible (35) using pathology reports from thousands of different histopathologists.

AJCC staging is based on the tumour-node-metastasis (TNM) categories to define the groupings: T (primary tumour characteristics), N (Regional Lymph node characteristics) and M (distant metastases characteristics). Characteristics pertaining to each category are described below, as per AJCC edition 7 (35-37):

"T" classification is based on:

- Tumour thickness (Breslow thickness): measurement (in millimetres) of thickness of tumour from top of epidermal granular layer to the deepest point of invasion
- Ulceration: the absence of intact epidermis above primary melanoma
- Mitotic rate: the number of mitotic cells per square millimetre of tumour.

"N" classification is based on

- number and type of regional lymph node metastases.

"M" classification is based on

- number and type of distant metastases.

The 7th Edition of the AJCC staging system was used to classify the melanomas used in this thesis, as this was active during the period of recruitment to the cohort and therefore that used by the pathologists reporting the histopathology slides. There is a new system in use since January 2018 (8th Edition of the AJCC staging system), but analyses performed by the research group showed that using only the broad staging, I, II, III of IV, there were only 3 participants with differences in the staging of disease in the data which formed the basis of this research.

1.4 **Melanoma aetiology**

Melanoma aetiology is multifactorial and involves a combination of genetic and environmental host factors. Epidemiological, *in vitro, in vivo* and most recently, omic studies have helped gain a deeper insight into this process. Melanoma risk factors can be broadly classified into environmental (UV/solar exposure) and host factors

(melanoma susceptibility genes, a phenotype characterised by increased numbers of melanocytic naevi, and pale skin type/pigmentation). However, the mechanisms underlying these factors are not mutually exclusive. The following sections will discuss each of these factors in detail, followed by a discussion of the molecular pathology of melanoma initiation.

1.4.1 Melanoma susceptibility genes

Melanoma susceptibility genes are those which when mutated or coded by polymorphisms (a mutation which is more common in the population having relatively minor effects biologically), confer an increased risk of developing melanoma. High-risk mutated genes have historically been identified using genetic studies of families with multiple family members with melanomas. Familial melanoma is considered to be the familial aggregation of melanoma, as defined by occurrence of melanoma in at least two relatives (either first degree or irrespective of degree of relationship) or families with three or more melanoma cases irrespective of degree of relationship (38). Melanoma susceptibility genes are normally classified as high-risk or low/moderate risk depending on the degree of risk they confer for developing melanoma, as indicated by the frequency of melanoma cases within an affected family. The first high-risk gene to be identified as associated with melanoma susceptibility was Cyclin-dependent kinase inhibitor 2A (*CDKN2A)* (39, 40) which was identified as result of linkage analyses of melanoma-affected families. *CDKN2A* encodes the tumour suppressor proteins p16INK4A and p14ARF: p16INK4A promotes cell cycle arrest by inhibiting retinoblastoma protein (RB) (41) and p14ARF induces cell cycle arrest and apoptosis via the p53 pathway (42). Hussussian et al first reported 8 different p16 germline mutations using a genotyping-based approach of melanoma-affected families. This included 1 nonsense mutation, 1 splice donor mutation and 6 missense mutations (Hussussian, 1994 #295). Since then, p16INK4A mutations have been predominantly reported to be loss-of-function missense mutations (39, 40, 43). For p14ARF, inactivation has been shown to be via whole gene deletions or splice mutations at the exon 1β (44-46). *CDKN2A* remains the most frequent high-risk melanoma gene with mutations detected in around 20-30% melanoma-prone families (47). However, other melanoma susceptibility genes have been identified since. Cyclin dependent kinase 4 (*CDK4*) was the second high-risk melanoma susceptibility gene identified in a candidate gene screening approach, an unsurprising finding given that CDK4 is the binding partner for p16INKA (48). The other high-risk melanoma susceptibility genes include the gene coding for Breast cancer associated protein 1 (*BAP1*) as well as genes involved in telomere

maintenance such as *TERT* (49), *POT1* (50, 51), *ACD* and *TERF2IP* (52), indicating the significance of telomere maintenance in melanoma susceptibility.

Among the intermediate-risk melanoma susceptibility genes is the Microphthalmia-associated transcription factor (*MITF*): a master regulator of melanocyte development and differentiation, (53, 54). The low-risk melanoma susceptibility genes include genes involved in a variety of biological processes such as pigmentation (*TYR, TYRP1, OCA2, MTAP*) (55), immune response (*HLA* class II genes, *IRF4*) (56, 57) and metabolism (*GSTM1, GSTT1, GSTP1*)(58). The predominance of pigment related genes as low risk melanoma susceptibility genes is related to the geographical variation in incidence described above. The single most important polymorphic gene associated with melanoma risk (*MC1R*) is most common in Northern Europe or in countries populated by migrants from Northern Europe such as Australia and New Zealand and is thought to underlie a significant proportion of susceptibility in those areas of the world.

While the genes mentioned above were discovered largely by Genome Wide Association Studies (GWAS), the role of other genes in melanoma susceptibility were identified using a 'candidate' approach. For instance, epidemiological studies to estimate the association of selected polymorphisms in the vitamin D Receptor (*VDR*) with risk of cutaneous melanoma have been reported (59).

1.4.2 Ultraviolet radiation and sun exposure

Ultraviolet/sun exposure is the most extensively studied environmental risk factor for melanoma, given the mutagenic role of ultraviolet radiation in skin cancers (60). Ultraviolet radiation is composed of Ultraviolet A (UVA: 315-400nm), Ultraviolet B (UVB: 280-315nm) and Ultraviolet C (UVC: 100-280nm), with 90-99% of radiation incident on earth's surface being UVA (60). The roles of UV radiation in initiating melanoma development and progression are discussed in the subsequent paragraphs. However, it is worth noting that incident UVB on the skin initiates a series of reactions leading to the synthesis of the essential hormone, vitamin D, which is known to have prognostic significance in melanomas: this aspect is discussed in detail in section 1.8. briefly, evidence described below indicating that sun exposure causes melanoma in the susceptible is complicated at least in public health terms because most people are dependent on sun exposure to manufacture enough vitamin D.

A combination of epidemiological and experimental evidence indicates the role of solar radiation in melanoma aetiology. In the case of epidemiological studies, the

association of melanoma incidence with various measures pertaining to sun exposure has been studied. For instance, mortality rates for malignant melanoma were shown to be inversely correlated with geographical latitude and annual ultraviolet radiation dose, in Canadian and American populations (61) as well as across 30 populations of European origin (62). It is to be noted that such studies are prone to population bias i.e. the effects associated with geographical gradient could alternatively be explained by gradient population mix and associated occupational sun exposure. Other studies have assessed the association of melanoma risk in relation to ambient solar radiation and length of residence in a specific geographical area. These studies collectively indicate that melanoma risk increases with increasing length of residence in a region of high ambient solar radiation (63). Because much of a person's sun exposure is during childhood and adolescence, studies have interrogated the effect of childhood sunburn on melanoma risk. However, there are no reported differences in melanoma risk among individuals who sunburn in childhood, adolescence or adulthood (64-66), suggesting that exposure is an important factor, irrespective of when it was acquired. In other words, the interpretation is that sunburn causes melanoma whenever it occurs. While the above studies have used incident solar radiation (in a particular geographical region) and sunburn history to gauge effects of radiation on melanoma risk, some others have used self-reported sun exposure as a measure of exposure to radiation.

For instance, a meta-analysis of 15 studies showed that reported recreational sun exposure was a risk factor for melanoma on the trunk and limbs, but not head and neck melanomas (67). Extensive epidemiological data linking UV radiation and melanoma risk is complemented by *in vitro* studies, wherein UV radiation has been shown to promote melanoma progression through several mechanisms (68). For instance, UVA-induced DNA damage is mediated by Reactive Oxygen Species (ROS) leading to intracellular oxidative damage (69). A recent study by Kemenisch et al report effects of that UVA treatment on melanoma cell lines from initial melanomas (vertical and radial growth phase). They demonstrated that UVA-treated early melanoma cells exhibit increased glucose uptake, lactate production and increased invasiveness. In other words, they provide *in vitro* evidence for UVA-induced Warburg effect mediated by oxidative stress in early melanomas, leading to an invasive phenotype (70). UV treatment (both UVA and UVB) have been shown to induce migration and invasion of not only melanoma cells *in vitro*, but also cells of the tumour microenvironment such as fibroblasts and untransformed melanocytes (71). Neonatal erythemagenic dose of UVB exposure has also been shown to increase melanoma progression, vascular invasion and an aggressive invasive

phenotype in 15 Hgf-Cdk4(R24C) mice, in which a DMBA-induced oncogenic CDK4 germline mutation leads to invasive melanomas as seen in patients (72). In the same mouse model, UV-induced immune suppression and subsequent tumour initiation was also shown. The same study also provided *in vivo* evidence for UV-induced TLR4/MYD88-driven neutrophilic inflammatory response, in addition to the UV-induced metastatic phenotype (73).

The advent of omic studies also provided significant insight into the signature of UV exposure in human melanomas. A study of 147 melanoma exomes (Whole Exome Sequencing- WES) identified excess C>T transitions in melanomas in sun-exposed body sites, which is an indicator of UV exposure and sun damage. They identified a motif (TTTCGT) to be enriched in genomic regions that are more likely to be mutated in sun-exposed melanoma. Given that this motif is a hotspot for creating UV-induced photoproducts, these findings argue for relevance of UV-induced DNA damage in melanoma initiation (74). Moreover, melanoma is known to harbour high mutation load compared to other cancer types (75, 76). In the TCGA melanoma dataset, 76% of primaries and 84% of metastatic melanomas harboured UV-driven C>T transitions (77). Notably, these studies have identified that the oncogenic *RAC1*^{P29S} mutation harbours the UV-induced C>T transitions in sun-exposed melanomas (compared to sun-shielded melanomas) (74, 77), thus accruing further evidence for UV-induced onset of melanoma.

1.4.3 Melanocytic Naevi

Just as UV radiation is a major environmental risk factor, increased numbers of melanocytic naevi is one of the major host-related risk factor for melanoma (pale skin with tendency to sunburn being the other major risk factor, being associated with inherited polymorphisms in pigment genes eg *MC1R*). Naevi were first observed and reported to be prevalent among melanoma-prone families (78), with the term Dysplastic Naevus Syndrome (DNS) being introduced to describe the naevus phenotype (79) characterised by a greater than average number of naevi and naevi which are individually unusual (dysplastic or atypical). Features such as increased size, border and pigmentation were used to define dysplastic naevi originally by Clark et al (78). However, there was historically a lack of consensus among dermatologists and dermatopathologists regarding the definition and classification of naevi: while ‘dysplastic naevi’ have a histologically-defined connotation, ‘atypical naevi’ appear to be more clinically-defined (80). A clinically defined atypical naevus is usually considered to be >5mm in diameter, with an irregular shape and colour. Despite differences in classification that thwart comparative/pooled analyses, the number of

naevi remains a significant factor for melanoma risk and hence has been extensively studied. Since the first report by Clark et al and coinage of the term 'dysplastic naevi', it has since been reported to occur frequently in Scottish (81), Dutch (82), English (83), Australian (84), Swedish (85), Italian (86), Spanish (87) and French (88) melanoma families. There is evidence for an autosomal dominant inheritance pattern of dysplastic naevi, as well as for a polygenic mode of inheritance (89). However, these naevi have been reported in sporadic cases as well, with occurrence in people with no family history of melanoma (90).

Despite differing criteria for classification of dysplastic naevi across epidemiological studies, it remains a consistent risk factor for melanoma (91). For instance, a pooled analysis of 15 case-control studies (across different latitudes) showed that higher whole-body naevus count was associated with significantly increased melanoma risk in participants aged<50 years and also those aged$\geq$50 years (92). In another meta-analysis of 47 case-control datasets, among which 27 studies had assessed melanoma risk, dysplastic naevi were highly significantly predictive of melanoma risk, despite differing criteria of clinical assessment (93). The phenotype is therefore a robust clinical marker of risk.

Though the aetiology of naevus development is not fully understood, there is some epidemiological and experimental evidence that offers insight. The role of sun exposure and ultraviolet radiation in naevus development has been widely explored. Studies have reported increased naevus counts in younger individuals living in sunnier regions (Australia) compared to less-sunnier regions (England), with no difference among older individuals (94). In Australian children, with very high naevus counts, there was associated self-reported sun exposure of >4 hours per day and family history of sunburn (95). Increased naevus counts (both dysplastic and common naevi) in sun-exposed compared to sun-shielded body sights have also been reported (96), adding further evidence for a role of solar radiation in naevus aetiology.

Genetic components associated with naevus aetiology have been identified from Genome Wide Association Studies (GWAS) of cutaneous naevus count (97) but not for dysplastic naevi. However, in a study of five melanoma families, family members with atypical mole syndrome were more likely to carry the *CDKN2A* mutation than a relative with no atypical mole syndrome (98). Linkage studies aiming to identify cause if dysplastic naevi have provided suggestive but inconclusive results for susceptibility loci on chromosomes 1,6, X (99) and 7 (100). The conclusion is therefore that the phenotype of increased numbers of naevi, which may or may not be associated with

clinically atypical naevi, is associated with risk and inherited melanoma susceptibility genes which may be highly penetrant or to have low penetrance.

Despite dysplastic naevi being a marker of melanoma risk, it is widely observed that these naevi rarely directly progress to melanoma. Follow-up studies of melanoma-prone families show that most naevi remained stable and rarely 'progressed' to a melanoma (101) (102). The study by Tsao et al estimated that the rate of transformation of any single naevus to a melanoma was $\leq$1 in 200,000 per year, in both men and women <40 years old. They also reported the lifetime risk of a 'naevus to melanoma transformation' as 0.03% for men and 0.009% for women. Though melanocytic lesions (dysplastic naevi, common naevi and/or atypical naevi) are considered to be 'growth-arrested', they harbour oncogenic *BRAF* mutations (103-106) and mutations in genes coding for the mitogen-activated protein kinase (MAPK) pathway (107). One explanation of this phenomenon is Oncogene Induced Senescence (108): for instance, overexpression of the activated oncoprotein $BRAF^{V600E}$ in cultured human melanocytes was shown to arrest growth and exhibit hallmarks of the senescent phenotype (109). In addition to the common *BRAF* mutation harboured by naevi, genetic alterations in *GNAQ, ROS, ALK, NTRK1, RET, HRAS* and *BAP1* have also been reported across the heterogenous spectrum of naevi (110).

Taken together, the development of benign proliferation of melanocytes which are naevi are most common in pale-skinned people and are associated with similar exposures and inherited genetic variation as melanoma. Moreover, it is not uncommon that melanomas arise in naevi. The initiating mutations e.g. in *BRAF* are not sufficient to cause a melanoma but proliferation occurring in a small percentage of melanocytes may result in a melanoma if additional changes occur which overcome the senescence normally associated with *BRAF* mutation.

1.4.4 Pigmentation and skin type

Pigmentation traits such as freckles, skin, hair and eye colour are known risk factors for skin cancers (111). In the case of melanoma, pigmentation has been identified as a risk factor in multiple epidemiological studies, with increased risk for pale-skinned individuals (91, 112). In other words, lower levels of pigmentation are associated with higher risk. Concordantly, melanoma burden is significantly higher in geographical regions with predominant pale-skinned populations (10).

Cutaneous pigmentation is of relevance to my thesis as most people derive vitamin D predominantly as a result of cutaneous synthesis in the presence of sun

exposure. That is that man evolved pale skin associated with an increased melanoma risk as the species migrated out of Africa, putatively as a result of the need to synthesise vitamin D (positive selection) and the loss of selection against the red hair gene coding for the MC1R "R" variant at the equator (Hochberg, 2010 #1096). The effects of vitamin D on melanoma are discussed separately in detail in section 1.8.

Skin pigmentation is determined by the amount and type of pigmentation produced by melanocytes, rather than the number of melanocytes. Eumelanin is black/brown pigment whereas pheomelanin is yellow/orange. The primary explanation for the association of reduced pigmentation and melanoma risk is that eumelanin produced by melanocytes scatters and absorbs 50-75% of UVR thus minimising DNA photo damage. In other words, eumelanin acts as a 'natural sunscreen' and thus prevent damage from UV exposure more efficiently than pheomelanin. Carcinogenicity of pheomelanin has also been postulated, with evidence for pheomelanin-dependent cellular oxidative stress (114): the presence of sulphur in the aromatic ring of pheomelanin makes it less stable and hence more efficient at producing reactive oxygen species (ROS) (115).

The eumelanin/pheomelanin ratio in melanocytes determines pigmentation and is regulated by the melanocortin 1 receptor (MC1R) signalling. The MC1R agonist melanocyte stimulating hormone-α (αMSH), upon activation leads to transcription of enzymes necessary for eumelanin signalling. On the other hand, the MC1R antagonist agouti signalling protein (ASIP in humans) promotes expression of enzymes for pheomelanin production and also inhibits eumelanin synthesis (116, 117). Thus, MC1R expression and signalling is a significant factor affecting pigmentation and consequently, UV-induced photodamage. *MC1R* is highly polymorphic with over 200 coding region variants described to date (118). Loss-of-function variants affecting the receptor's signalling ability lead to a shift away from eumelanin and towards pheomelanin synthesis (119, 120). This shift to pheomelanin is associated with the 'red hair colour' (RHC phenotype[ii]) which is characterised by pale skin, freckling and sun sensitivity (121). *MC1R* variants are classified according to the strength of associations with the RHC phenotype into "R" (strong association) (122, 123) or "r" (weaker association) alleles (122, 124).

[ii]RHC phenotype: The Red Hair Colour phenotype is characterised by the following pigmentary traits: pale skin pigmentation, red hair, lack of tanning ability and propensity to freckle.

MC1R variants associated with melanoma risk have been reported by several studies in different human populations. These studies have demonstrated that while some *MC1R* variants are associated with both melanoma risk and the RHC phenotype, others are associated only with melanoma risk, suggesting that the role of *MC1R* in melanoma development involves non-pigmentary routes/components (124-126) as well as pigmentary routes. Melanomas from individuals carrying germline *MC1R "*R" variants have a significantly higher somatic mutational burden, compared to those with no *MC1R "*R" variants. This effect was independent of confounders such as age, sex, site of melanoma, Breslow thickness and ulceration status (127). This study complements the notion of eumelanin (produced by functionally intact MC1R) being a 'natural sunscreen': loss-of-function *MC1R* variants, lead to reduced eumelanin production with consequent increase in susceptibility to UV-induced DNA damage and increased somatic mutation load.

The non-pigmentary effects of MC1R signalling have been reported to have a role in cutaneous immune responses and nucleotide excision repair. αMSH (the eumelanin producing agonist of MC1R), has been shown to modulate cutaneous immune response by inhibiting pro-inflammatory cytokines such as IL-1, IL-2 and interferon-gamma (128). It has also shown to have immunomodulatory effects via MC1R signalling in neutrophils and macrophages (129). However, α-MSH and MC1R are not necessarily dependent on each other for activity: α-MSH can signal via other melanocortin receptors (MC3R and MC5R) and MC1R can be stimulated by other agonists/antagonists. The ability of effective MC1R signalling (with the receptor being functionally intact) to boost UV-induced nucleotide excision repair (NER), via the cAMP production has been demonstrated in *in vitro* and *in vivo* models (130, 131). Thus, ineffective *MC1R* signalling can adversely affect cutaneous immune and DNA damage responses, thus offering additional explanations as to why *MC1R* variants (with partial/complete disrupted MC1R signalling) are associated with increased melanoma susceptibility (132).

Though *MC1R* variants have bene extensively studied, there are several other genes that contribute to pigmentation in humans. While some of these genes contribute to melanin production, others control the function of the primary melanin producing cells, the melanocytes. The genes that control melanin production encode enzymes regulating ratio of eumelanin to pheomelanin. Polymorphisms in these genes have been identified using GWASs, to be associated with increased melanoma risk: those coding for tyrosinase (133, 134), tyrosinase-related protein-1 (*TYRP1*), tyrosinase-related protein-2 (*TYRP2*), oculocutaneous albinism 2 (*OCA2*), solute carrier family 45, member 2 (*SLC45A2*), solute carrier family 24, member 4

(*SLC24A4*), and agouti signalling protein *ASIP* (135). These findings were confirmed in an independent GWAS study which agnostically identified the previously reported genetic loci associated with melanoma risk (136). The microphthalmia-associated transcription factor (MITF) is a melanocyte lineage-specific transcription factor which controls melanocyte migration and differentiation. The αMSH-MC1R-cAMP signalling axis leads to MITF-mediated transcription of target genes, which include eumelanin-synthesizing enzymes as well as pro-survival genes anti-apoptotic genes such as *BCL2A1*, *BCL2* and *BIRC7* (137), thus making MITF signalling a crucial link between pigmentation and tumour progression (138).

Taken together, the contribution of the 'pigmentation machinery' to melanoma risk is evident from genetic and functional studies, which have unravelled the mechanistic basis of the protective effect of melanisation.

1.5 Molecular pathogenesis of melanoma

The complex and multifactorial aetiology of melanoma has been described in the previous section. In addition to identifying the environmental and genetic factors that contribute to melanomagenesis, research efforts have focussed on understanding the molecular basis of melanomagenesis. The collective understanding of aberrant signalling pathways involved in melanomagenesis have also formed the basis of diagnostic and therapeutic approaches, which aid the management of the disease. A particular pathway or gene is considered likely to contribute to melanomagenesis if an aberration (mutation, genetic loss/gain, change in expression) is i) observed to exist consistently across melanoma stages or ii) be functionally proven to initiate melanoma development in *in vitro* and/or *in vivo* models. Based on this 'definition', various pathways and genes have been shown to contribute variably to melanomagenesis. This section will discuss the contributions of the major signalling pathways and genes whose dysregulation has been shown to contribute to melanomagenesis. The intent of this section is to give an overview of the molecular pathways and the consequent cellular processes, without being exhaustive.

1.5.1 The Ras-Raf-MEK-ERK pathway

The Ras-Raf-MEK-ERK pathway is a complex signalling cascade which responds to various hormones, differentiation and growth factors, to regulate crucial cellular functions. Aberrations in this pathway are frequently observed in many cancer types, with consequent effects on tumour proliferation, apoptosis and differentiation (139).

The three Ras proteins: H-Ras, N-Ras and K-Ras are small GTPases activated by the conformational change induced by the exchange of GDP for GTP. The active forms of the Ras proteins function as adapter molecules by binding to Raf kinases (three isoforms: A-Raf, B-Raf and C-Raf), which in turn lead to the sequential phosphorylation of the kinases MEK (MEK1 and MEK2) and ERK (ERK1 and ERK2). These kinases in turn phosphorylate and activate transcription factors which regulate crucial cellular functions (140). In sporadic cutaneous melanoma, two particular components of this signalling pathway are known to be dysregulated: the proto-oncogenes N-Ras and B-Raf (141). The mutually exclusivity of *NRAS* and *BRAF* mutations in melanoma has been reported (142). *NRAS* mutations are known to occur in approximately one-third of primary and metastatic sporadic cutaneous melanomas (143), the most common mutation being the Q61R (Glutamine- Arginine) mutation which impairs GTP hydrolysis and thus renders N-Ras constitutively active (144). *BRAF* mutations are known to occur in about 50-70% of melanomas, with the V600E (Valine to Glutamic acid) mutation accounting for 90% of all *BRAF* mutated melanomas. The V600E mutation affects the kinase domain leading to increased kinase activity of B-Raf (145). As discussed in section 1.4.3, *BRAF* mutations are observed in about 80% of benign naevi, including dysplastic naevi, suggesting that $BRAF^{V600E}$ is an early mutational event. Though, introduction of only the $BRAF^{V600E}$ in human melanoma cells has been shown to cause cell cycle arrest and senescence (109), expression of $BRAF^{V600E}$ combined with knockdown of the tumour suppressors *TP53* or *PTEN,* has been shown to lead to spontaneous melanoma development (146, 147) in animal models. This supports the view that the $BRAF^{V600E}$ mutation is an early oncogenic event and leads to malignant transformation by acquisition of additional mutations (148). Taken together, these and other studies implicate $BRAF^{V600E}$ mutations as a significant and early player in melanomagenesis.

1.5.2 PI3K, AKT and PTEN pathway

The PI3K-Akt-PTEN pathway is one of the downstream effector pathways of the Ras proteins: activation of Ras leads to phosphorylation of PI3K (phosphotidyl inositol-3-kinase) which in turn activates Akt, which is a key signalling molecule mediating processes such as angiogenesis, proliferation, apoptosis and cellular metabolism (149). This signalling axis begins with a ligand-dependent activation of tyrosine kinase receptors, G-protein coupled receptors or integrins (150). The tumour suppressor PTEN (Phosphatase and tensin homolog) inhibits the activation of PI3K, thus controlling the downstream effects of this signalling axis (151). In melanoma, loss or reduction of PTEN activity has been shown to eliminate Akt regulation,

resulting in melanoma progression and invasion (152-154). PTEN loss is considered an early event in melanoma development, as a consequence of single copy loss of chromosome 10, on which *PTEN* is located (155, 156). *PTEN* loss has been shown to bestow tumourigenic potential in melanoma cell lines by increasing Akt activity, thus circumventing anti-apoptotic signals (157). *PTEN* loss has also been shown to cooperate with $NRAS^{Q61K}$ mutations to initiate melanomagenesis, in a PI3K-independent and β-catenin dependent manner (158).

The mammalian Akt (also named protein kinase B-PKB) family is comprised of three highly homologous isoforms: Akt1, Akt2 and Akt3, all of which are crucial components of the PI3K-Akt-MEK-ERK signalling axis. Constitutive activation of Akt signalling, in particular *Akt3* (one of the three isoforms of Akt), has been shown to be prevalent in over 60% of melanomas (152, 154). This increased activity of Akt can be attributed to activating mutations of *AKT3* (159) or due to loss of its negative regulator *PTEN* (152). Thus, the deregulation of the PI3K-AKT-PTEN signalling axis has been shown to affect melanoma initiation and progression.

1.5.3 MITF

MITF (Microphthalmia-associated transcription factor) is the master transcription factor for the regulation of melanocyte differentiation and pigmentation. The latter has been discussed in the context of the role of pigmentation in the section: Melanoma Aetiology: Pigmentation and skin type (section 1.4.4). *MITF* amplification is more common in metastatic compared to primary melanomas and is associated with worse prognosis. However, MITF is considered to play a 'double role' of inducing and supressing cellular proliferation, depending on the level of *MITF* expression (160). *MITF* amplifications in conjunction with $BRAF^{V600E}$ has been shown to transform human melanocytes *in vitro.* Moreover, this study also identified *MITF* as a 'lineage survival oncogene': a gene required for the development of melanocytes, but amplification of which is maintained as a feature in melanoma initiation and progression (161).

The other genes implicated in the molecular pathogenesis of melanoma include *KIT* and *TP53.* Mutations or amplifications of *KIT* (CD117) have been reported in melanomas that occur in anatomic sites with little UV exposure such as oral mucosal melanomas, acral melanomas and anal melanomas (162, 163). The precise mechanism of how *KIT* contributes to melanomagenesis is unclear. However, studies show that constitute KIT activation in melanocytes leads to MITF and ERK2 activation, suggesting a potential route for melanoma initiation. p53 is a master

transcription factor which is activated in response to cellular stress cues such DNA damage and hypoxia (164). Inactivating mutations and deletions of the *TP53* gene have been reported in many cancer types. However, in melanoma *TP53* alterations are reported to be of low frequency, detected in 1-5% in primary (165) and 11-25% in metastatic melanomas (166), though variable frequencies have been reported.

In addition, a more recent study using whole exome sequencing (WES) of 213 melanomas identified inactivating mutations in the *NF1* gene in tumours that were wild-type for both *BRAF* and *NRAS*, the most common mutations in sporadic melanomas. Since loss of function of NF1 was accompanied by activation of the RAS pathway in a proportion of the samples, this study posited the role of *NF1* as a tumour suppressor gene in a subgroup of sporadic melanomas (167). *BRAF, NRAS* and *NF-1* are considered to be the most frequent melanoma driver mutations.

1.6 Tumour host interaction: role tumour microenvironment in melanoma development and progression

The tumour microenvironment has been accepted to be a significant contributor to the process of tumour progression. The 'seed and soil' hypothesis (168) suggests that tumour growth and survival is contextual: some physiological environments are more conducive for tumour development than others. The significance of the tumour microenvironment (TME) (169) in tumourigenesis and progression has since been extensively explored and remains an active area of cancer research in multiple cancer types. Components of the TME that have been implicated in tumour development include the extracellular matrix (ECM), fibroblasts (including cancer-associated fibroblasts- CAFs), immune and inflammatory cells, vascular networks and adipose cells. The crosstalk between these components and with tumour cells has been shown to influence the crucial characteristics of tumour development, as typified by the ten Hallmarks of Cancer (170). It is now known that melanoma cells actively interact with various components of their microenvironment, with significant implications on disease progression. Selected TME components and their role in melanoma progression are discussed below.

1.6.1 Immune response to melanoma

The role of the host immune response has been extensively studied and shown to be an important determinant of melanoma progression. Even prior to the era of molecular and genomic profiling, histological evidence of lymphocytic infiltration was

identified as a feature of spontaneous regression in primary melanomas (171). Multiple studies have since shown that tumour infiltrating lymphocytes (TILs) are of prognostic significance in independent cutaneous melanoma patient cohorts. Collectively, these studies have identified the following measures of immune infiltration to be of prognostic significance: intratumoural/peritumoural TIL density (172-175) and patterns of infiltration (brisk versus non-brisk infiltration) (176, 177). The vast majority of studies indicate that histological evidence of the presence of TILs is associated with improved prognosis and reduced chance of lymph node metastasis (178). A lack of prognostic significance has also been reported (179) in which the authors suppose that their findings suggest a role for TILs in vertical growth phase tumours rather than radial growth phase tumours. In the case of immune cell types, histological expression of markers for CD3, CD4 and CD8 cells (180, 181) and B cells (182, 183) have been associated with improved prognosis. On the other hand, markers for NK cells (184) and expression of FOXP3 (185) were associated with higher risk of relapse and worse progression free survival respectively. However, it has been shown that in the case of metastatic melanomas, tumours are not eliminated despite detectable immune infiltrate (186), indicating that immune evasion mechanisms could be involved. Taken together, this indicates that the interplay between melanoma cells and the host immune response is complex and extensive research efforts in the field are being focused on understanding this relationship better.

The apparent immunogenic behaviour of melanoma has been attributed partially to the high mutational burden in melanomas compared to other cancer types (75). Despite melanoma being one of the most immunogenic cancer types and increased immune infiltration predicting improved melanoma prognosis, metastatic melanoma mortality remains high (discussed in Melanoma Mortality section). This suggests that melanoma progression involves mechanisms that enable evasion of anti-tumour immune response. The response of the host immune response to tumour cells is a multi-step process and defects in these steps are characteristic of immune evasion. In the case of melanoma, the following steps give a brief overview of how melanoma cells evade the host immune response:

Defective antigen presentation: inefficient antigen presentation by melanoma cells is one of the causes of ineffective immune recognition. Expression of MHC Class I (187) and MHC Class II (188) may be downregulated in melanoma cells. Mutations in components of antigen presenting machinery in melanoma cell have also been reported (189). Moreover, since T cell cytotoxicity requires antigen presentation by mature dendritic cells (DCs), tumour-produced factors that prevent DC maturation or

switch to a tolerogenic phenotype have been described to be part of the immune evasion mechanism (190).

Defective priming and activation of T cells: cytotoxic T cell function is also inhibited by upregulation of immune checkpoint molecules. The role of immune checkpoint signals in normal physiology is to limit T cell responses in order to preserve self-tolerance during an immune response against a foreign antigen (191): to reduce tissue damage resulting from uncontrolled inflammation. However, upregulation of checkpoint molecules by melanoma cells dampens a cytotoxic T cell response against the tumour, thus becoming a mechanism of tumour immune evasion. Increased expression of the following checkpoint molecules has been shown to reduce the anti-melanoma immune response: CTLA4, PD1, LAG3, TIM3, VISTA and the checkpoint ligands PD-L1 and PD-L2 (192). Current immunotherapy approaches to treat metastatic melanoma aim to inhibit these checkpoint molecules, thus permitting an active T-cell mediated immune response against melanoma. This might be considered to be equivalent to removing the brakes on the immune responses. This is discussed in detail in subsequent section titled 'Melanoma immunotherapy'.

Reduced activity of anti-tumour immune populations: Inefficient antigen presentation impairs the ability of CD8 T cells to detect tumour-specific antigens. In addition to T-cell suppression, reduced NK cell responses have also been reported in melanoma patient-derived tumours as well as murine melanoma models. In normal physiology, NK cell activation and responses are mediated by activating receptors such as NKG2D, NKp46 and DNAM-1 (193). However, tumours from melanoma patients have reduced expression of these activating receptors (194) indicating NK cell-associated immune evasion (195). Moreover, NK cell functions are also inhibited by MDSC-mediated factors, as discussed below.

Upregulation of immunosuppressive immune populations: in normal physiology, cells such as T-regs and MDSCs function to balance immune responses and prevent 'excessive' immune response which could lead to autoimmunity. While T-regs are a specialised subpopulation of T cells that inhibit T-cell proliferation and cytokine production (196), MDSCs are a heterogenous population of myeloid origin cells which suppress various T-cell functions (197). In melanoma, T-regs-mediated over-production of factors which dampen activity of CD4 and CD8 T-cells and NK-cells (such as IDO and IL10) have been reported (198). The recruitment and stimulation of MDSCs to the TME has been shown to increase production of factors (such as nitric oxide, ROS-Reactive Oxygen Species and arginase-1) that inhibit anti-tumour

activity of T-cell and NK cells (199). A study by Jordan et al also reported increased MDSCs (defines as $CD3^-$ $CD19^-$ $CD56^-$ $CD11b^+$ HLA-DRA^- $CD33^+$ $CD14^+$ population) in the peripheral blood of stage IV melanomas compared to healthy donors (200), which indicates that MDSCs are associated with advanced melanomas.

1.6.2 Fibroblasts in the microenvironment

CAFs have been described as a subpopulation of functionally distinct fibroblasts that facilitate tumour promotion by enhancing pro-tumourigenic processes such as angiogenesis, inflammation and metastasis (201). CAFs are identifiable in the close vicinity of tumour cells and are functionally distinct from normal fibroblasts, in that they produce (and respond to) tumour promoting signals (202-204). CAFs have also been shown to selectively suppress CD8-mediated anti-tumour immune responses (205), thus highlighting the importance of this cellular subpopulation in tumour progression in general and melanoma in particular.

Melanoma-associated CAFs have been shown to mediate melanoma progression in both murine and human melanoma tumours. For instance, the NK cell-dependent cytotoxicity of fibroblasts derived from normal skin (normal fibroblasts) and primary melanoma tumours (CAFs) were compared: the CAFs decreased the susceptibility of melanoma cells to NK cytotoxicity by production of matrix metalloproteinases (MMPs) (206). Melanoma associated fibroblasts, in concert with fibronectin-rich matrices, have been shown to confer resistance to BRAF inhibitors, in the adjacent melanoma cells (207). This finding suggests that therapy resistance in melanoma patients to BRAF inhibitors is at least partially explained by the TME, making the TME an important factor in determining therapy resistance.

1.7 Melanoma therapy options: immunotherapy

Treatment options for melanoma are dependent on the stage of melanoma at diagnosis (Macbeth, 2015 #1107). The prevalent recommendation for stage I melanoma is surgery with wide excision, which involves the removal of the melanoma as well as the normal skin surrounding it. Stage II melanomas are also treated with surgery, but additional lymph node biopsies are also undertaken to gauge the spread of the disease. In the case of stage III and stage IV melanomas, surgical excision and lymph node biopsies are followed by adjuvant treatment with immunotherapies or targeted therapies (typically BRAF and MEK inhibitors). Radiation therapy is also

recommended in cases where lymph nodes are excised and found to contain malignant lesions.

Cancer immunotherapy is the collective term used to describe cancer therapies which aim to harness the 'cancer-eliminating' properties of the immune system. The concept of the immune system being capable of identifying and targeting transformed cancer cells stems from the Cancer Immunosurveilance model (208). This model describes the interaction between the host immune system and the tumour cell, wherein the tumour cell evades the attempts of the immune system to retard its progression (209). Simply put, immunotherapy aims to bolster the immune system's ability to detect and destroy transformed cancer cells. Various lines of immunotherapy exist which target various components of the immune system in order to increase anti-tumour immune response. Among them is the use of immune checkpoint inhibitors, in particular: anti-cytotoxic T lymphocyte antigen 4 (anti-CTLA4) and anti-programmed cell death protein 1 (anti-PD-1). CTLA4 and PD-1 are checkpoint molecules expressed on T-cells. PD-1, expressed on surface of T-cells, is the binding receptor for PD-L1/2 (expressed on tumour cells). In normal physiology, the binding of the receptor (PD-1) to the respective ligands (PD-L1/2) serves to regulate T-cell activation by competing with the co-stimulator CD28 to prevent persistent T-cell activation (210). The mechanism by which CTLA4 participates in regulation of T-cell activity is by competitive inhibition: CTLA4 competes with costimulatory molecule CD28 for binding to the B7 ligands B7-1 (CD80) and B7-2 (CD86). While CD28 binding to the B7 ligands provides positive costimulatory signals (and T cell activation), binding of CTLA4 to B7 ligands dampens T cell response (Linsley, 1994 #1098). Negative costimulation by CTLA4 is thus crucial in balancing T cell responses (Wei, 2018 #1097). Given both PD-1 and CTLA4 function as regulators of T cell response, led to the term 'checkpoint molecules'. The rationale behind using this mechanism for checkpoint immunotherapies is that blocking these checkpoint molecules would enable uninhibited cytotoxic attack by T-cells on the tumour. The checkpoint inhibitors that are currently used (FDA-approved) to treat metastatic melanoma are Ipilimumab, Nivolumab and Pembrolizumab and most recently adjuvant use of some has also been approved. While Ipilimumab targets CTLA-4, Nivolumab and Pembrolizumab target PD-1. Though CTLA4 and PD-1 are the active drug targets at the moment, other costimulatory and inhibitory ligands on cytotoxic T lymphocytes (CTLs) T-regs and other immune cells in the tumour microenvironment have been investigated. These include lymphocyte activation gene 3 (LAG3), whose binding ligand is MHC class II and is expressed on T-regs and inhibits T cell response (211).

The determinants of response to checkpoint therapies have been investigated in patient cohorts, in order to identify the factors that best predict response. Increased response rates to PD-1 inhibition in patients whose tumours or TILs express PD-L1 have been reported in some trials (212, 213). MHC Class II expression has also been associated with improved response to PD-1 inhibition (214). However, the following factors thwart the development of a reliable measure to predict response to checkpoint therapies: significant variation in the estimation of PD-L1 expression (using IHC), demonstrable response in some patients whose tumours do not express PD-L1 and lack of correlation between tumour PD-1 expression and CTLA4 response. More recently, researchers have taken a retrospective, genomic view of assessing differences between responders and non-responders to immunotherapy, i.e. identifying omic-based biomarkers to predict therapy response.

Whole Exome Sequencing (WES) has been used to identify associations between factors such as mutational burden and clinical benefit. This approach has been adopted in studies by Snyder at el (215) (64 patients treated with anti-CTLA4) and Van Allen et al (216) (110 patients treated with anti-CTLA4), who arrived at the consensus: high mutation burden is associated with clinical benefit to anti-CTLA4 treatment. Another study by Riaz et al (217) (WES on 174 patients treated with anti-CTLA4) reported that a subset (n=48) of patients with mutations in *SERPINB3* and *SERPINB4* had a higher likelihood of being responders. WES and RNAseq data from patients treated with anti-PD1 therapy were analysed by Hugo et al, who reported that mutation burden was not associated with therapy response. Albeit in a very small cohort (38 patients), whereas in a subset of patients (n=28), a gene signature of 26 genes was able to classify responders and non-responders (218). The lack of a validation data set in this study and in the other studies listed reflects the paucity of data yet available to address this need. Non-the-less these studies suggest that mutation load and specific, biologically relevant gene expression patterns underlie variation in response to immunotherapy in metastatic melanoma.

Other studies have explored the effect of factors such as the diversity of the T cell repertoire before treatment and its effect on response to immunotherapy (219-222), this area of research is actively expanding.

1.8 Vitamin D-VDR signalling

1.8.1 Components of vitamin D-VDR signalling

Vitamin D-VDR signalling refers to the signalling system activated in response to the ligand 1α,25-dihydroxyvitamin D_3 (vitamin D_3 henceforth), with consequent effects on transcription of target genes. The vitamin D_3 endocrine system is composed of two 'arms': a) vitamin D_3-metabolising enzymes which belong to the Cytochrome P450-CYP family of anabolic and catabolic enzymes and b) dedicated nuclear receptors such as VDR (Vitamin D Receptor) and RXR (Retinoid X Receptor), which upon activation bind to Vitamin D Receptor Element (VDRE)- containing regions of the chromatin and facilitate transcription of target genes (223). While the machinery to synthesis biologically active form of vitamin D_3 is found in yeast, plants and some invertebrates, the complete vitamin D_3 endocrine system (both the 'arms') is unique to vertebrates, indicating that vitamin D signalling gained importance as organisms evolved to cope with environmental stresses (223).

The biologically active form of vitamin D is vitamin D_3 (1,25 dihydroxyvitamin D_3), which is produced by a series of enzymatic reactions. Most humans obtain sufficient vitamin D primarily as a result of sun exposure, as there are few good dietary sources. While vitamin D_3 obtained from dietary sources enters circulation after processing in the intestine, cutaneously synthesised vitamin D_3 is metabolised in the dermis. The first cutaneous step involves the UVB-mediated non-enzymatic conversion of 7-DHC (7-dehydrocholesterol) to vitamin D_3 (also known as cholecalciferol). Intensity of UVB as well as skin pigmentation determine this first step (224), with increased melanin content retarding the UVB-mediated production of vitamin D_3 (225). Vitamin D levels in Northern Europe therefore are on average some 20 nmol/L lower in winter months where there is insufficient exposure to UVB for efficient synthesis (226, 227). The next enzymatic reactions are facilitated by the cytochrome P450 (CYP) family of enzymes, which specialise in catalysing various biological functions in human physiology. Six CYPs are involved in vitamin D metabolism via stepwise hydroxylation reactions of intermediate molecules. Among these, the primary CYPs associated with Vitamin D anabolism (synthesis) are CYP27A1 and CYP27B1, processes taking place in the liver and kidney respectively. Circulating cholecalciferol is converted to 25(OH)vitamin D_3 in the liver, which is then transported to the kidneys to be converted to 1α,25-dihydroxyvitamin D_3: the 'final' hormonal form of vitamin D. The catabolism (break-down) of vitamin D is mediated by CYP24A1 in the kidney. CYP24A1 responds to high levels of circulating vitamin D_3 by catalysing the break-down of 25-hydroxy vitamin D_3, to lower the level of hormonal vitamin D_3. Both

anabolic and catabolic CYPs are transcriptionally controlled: they contain the VDRE and hence are tightly regulated by the concentration of active vitamin D_3 at any given point (228). This process is summarized in Figure 1.5 (229).

The Vitamin D Receptor (VDR) and Retinoid X Receptor (RXR) are members of the Nuclear Hormone Receptors (NHR) family- which includes other members such as the Retinoid Activated Receptor(RAR), Peroxisome Proliferator-activated Receptor (PPAR), Tyrosine Receptor (TR) and Liver X Receptor (LXR), to name a few (230). Because VDR is the canonical receptor of vitamin D signalling and mediates its genomic effects, it is discussed in detail below. However, the role of RXR and its canonical ligand Retinol (vitamin A) in cancer and metabolic disease have also been extensively researched (231). In my thesis, I do account for the effects of RXR, when interrogating the effects of VDR on the LMC melanoma transcriptome but I have not explored this literature in the thesis.

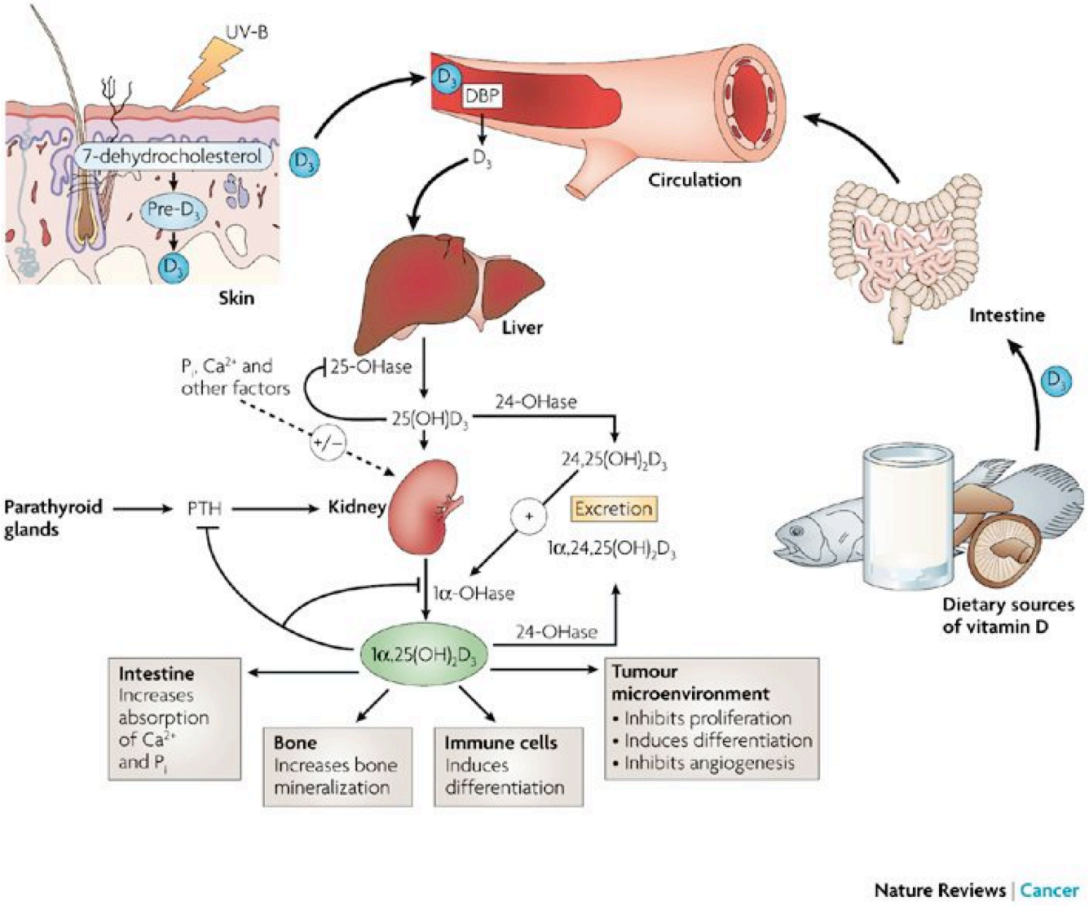


Figure 1.5 Summary of the vitamin D-VDR signalling axis.

Adapted from: Deeb KK, Trump DL, and Johnson CS. Vitamin D signalling pathways in cancer: potential for anticancer therapeutics. Nature reviews Cancer. 2007;7(9):684-700.

1.8.2 The Vitamin D Receptor (VDR)

The Vitamin D Receptor (VDR) is the high-affinity Nuclear Hormone Receptor (NHR) of hormonal vitamin D_3. Being a transcription factor, VDR dimerizes with RXR (Retinoid X Receptor) to enable transcription of genes whose promoters contain VDRE (Vitamin D Responsive Elements) (232). VDR belongs to the Nuclear Receptor superfamily whose members contain a conserved DNA-binding and ligand-binding domains as well as variable C-terminal and N-terminal regions (233).

1.8.2.1 VDR structure: functionally relevant features of VDR

The cloning of the human VDR (hVDR) in 1988 (234) paved the way for numerous studies that unravelled the structure-function relationship of VDR (235). Human VDR is a 423 amino acid protein composed of multiple structural domains that have been ascribed specific functional relevance based on structure/function data. VDR can thus be split into three functional domains described below (236).

The N-terminal DNA binding domain (DBD) is the most conserved among nuclear receptors and is composed of three α-helical sub-domains (P-box, D-box and T-box) which enable stable DNA-binding (positively charged amino acids bind strongly to negatively charged phosphate backbone of DNA) as well as response element specificity.

The C-terminal ligand-binding domain (LBD) is composed of α-helices and β-strands that encompass a lipophilic ligand-binding pocket. In addition to ligand-binding, the LBD also mediates heterodimerization with RXR and ligand-activated recruitment of coregulatory complexes.

The Hinge region links the DBD and LBD and offers rotational freedom for binding various response elements.

1.8.2.2 VDR-mediated regulation of target genes: The Vitamin D Response Element (VDRE)

The tissue and gene specificity of ligand-bound VDR-mediated transcription is regulated by two factors: DNA binding specificity (presence of VDRE) within regulatory regions of target genes and recruitment of ancillary transcription complexes. These factors are crucial, given that they determine the specificity and intensity of VDR-mediated transcription.

Like other members of the nuclear receptor family, ligand activated VDR-RXR dimers have affinity for specific DNA binding sequences. In addition to the actual repeat sequence, the nature of the repeat (direct, everted, inverted or palindromic) and spacer length (+3 nucleotides, +4 nucleotides etc) determine binding specificity among HREs (Hormone Response Elements). For instance, two direct hexameric repeats of AGGTCA form the binding site for VDR (known as DR3-type VDRE), all-trans retinoic acid receptor and 9-cis retinoic acid receptor. However, these response elements differ in the number of spacer nucleotides between the half sites: 1nt for the 9-cis retinoic acid receptor, 5nt for the all-trans retinoic acid receptor and 3nt for the VDR (237). This highlights the potential of these unique repeat sequences to control transcriptional diversity. The VDREs are typically composed of the DR3-like elements (described above), other VDRE types exist- ER9-type (Everted Repeat with 9nt spacer) and even a transrepressor VDRE (nVDRE- negative VDRE) (238). The first genes to be identified as having a proximal VDRE that facilitated their transcription were Osteocalcin (239) and Osteopontin (240), which were specifically studied owing to the known role of vitamin D-VDR signalling on bone remodelling. This was followed by discovery of this specific VDRE sequence in regulatory elements of other genes, thus putting forth a conclusive VDRE sequence. This eventually lead to identification of VDRE across the genome with the advent of high throughput technologies such as ChIP-Seq (Chromatin Immunoprecipitation). I have used data from these studies to support my own research, which I have discussed in detail in Section 3.4.5.3.

1.8.2.3 Role of coregulatory complexes in VDR function

Ligand binding to VDR causes a series of conformational changes, one of which is an alteration to the AF2 domain of the LBD, making available hydrophobic docking surfaces for cofactor binding. A typical example is the SRC1 (Steroid Receptor Coactivator-1) which has autonomous transcription activation potential, histone acetylase activity capable of directly altering chromatin structure (241) and is known to bind and facilitate VDR-mediated transcription activity (242, 243). A more specialized coactivator complex for VDR is the DRIP (VDR-interacting proteins) complex that has no homology with SRC family coactivators and yet recognizes the AF-2 transactivation motif of VDR. DRIP complex proteins were the first VDR-specialised coactivators to be discovered, owing to their ability to selectively enhance transactivation of ligand bound VDR-RXR dimers (244). Thus, coactivator complexes make the chromatin environment more conducive to transcription initiation, while also offering selectivity for VDR-mediated transactivation.

1.8.2.4 Ligand-induced activation of VDR

Ligand binding to nuclear VDR prompts a change in the co-factors bound to VDR: co-repressors are now released and replaced by co-activators. This leads to a more 'receptive' chromatin confirmation prompting mobilization of transcription machinery and transcription of target genes. The negative feedback loop keeps this process in check by a) 'supply' of the ligand vitamin D is controlled by anabolic CYPs-transcribed by vitamin D-VDR binding and b) resultant unliganded cytoplasmic VDR is degraded by proteosomal and polyubiquitination events (245)

Taken together, the above section gives an overview of the different mechanisms by which VDR exerts transcriptional control over target genes, highlighting the complexity and yet the specificity of this transcription factor.

1.8.3 Vitamin D signalling targets

The most well-established physiological effect of Vitamin D_3 is the effect on bone mineral homeostasis, where vitamin D_3 is known to mediate the balance of calcium and phosphate-associated signal transduction pathways. vitamin D_3-bound VDR induces expression of calcium channel proteins such as TRPV6 in the small intestine, as well as calcium absorption-enabling enzymes such as calbindin (246). Additionally, vitamin D_3 synthesis is stimulated in response to reduced circulating phosphorous, leading to vitamin D-VDR-mediated transcription of phosphorous-absorbing enzymes and membrane proteins (247). In the case of inadequate serum vitamin D_3 (vitamin D deficiency), this leads to decreased calcium and phosphorous absorption by the bone chondrocytes, i.e. bone resorption. This weakens the bone and cartilage architecture (248). However, VDR expression is not restricted to intestines and skeletal tissue, but is also expressed in the brain, skin, muscle, heart, stomach, pancreas, mammary glands, testes and in activated T and B lymphocytes (249, 250). The enzymatic machinery needed to metabolize vitamin D_3 is expressed in colon, prostate, breast, and skin (250). In addition to the classic genomic response, vitamin D_3 is also known to elicit certain non-genomic responses that are rapid (within seconds to minutes, depending on tissue type) and transcription-independent, such as the rapid intestinal absorption of calcium (transcaltachia), opening of voltage-gated Ca^{2+} and Cl^- channels in osteoblasts and rapid migration of endothelial cells (251). Taken together, it can be deduced that most human tissue are responsive to vitamin D_3 (249, 252), indicating the ubiquity of this signalling axis, with potential tissue-specific effects.

1.8.4 Vitamin D signalling and the immune system

Evidence linking vitamin D and the immune system originates back to when cod liver (which is a rich source of dietary vitamin D) was used to treat tuberculosis, albeit without attributing the effects to vitamin D itself (253). The understanding of the relationship between vitamin D and the immune system has since burgeoned, with evidence indicating both 'pro-immune' and 'anti-immune' effects of vitamin D. The evidence that vitamin D is associated with an increased (pro-immune) response stems from both epidemiological observations and functional studies. Multiple cross-sectional studies have reported low vitamin D to be associated with increased reports of infections such as upper respiratory tract infection (254, 255), influenza (256) and HIV infection (257). An association between vitamin D deficiency and infections suggests therefore a role for the hormone in mounting immune responses to infections: or that vitamin D deficiency is associated with immune failure. Epidemiological evidence indicates that vitamin D deficiency is also associated with development of autoimmune diseases such as multiple sclerosis (258), diabetes mellitus (259) and rheumatoid arthritis (260) which might suggest that upregulation of immune responses to "self" tissues is also related to vitamin D deficiency. There is extensive literature indicating that lupus patients have lower vitamin D levels compared to healthy controls (261). Moreover, the canonical receptor for vitamin D activity: VDR, is expressed on lymphocytes (262), suggesting that immune cells can respond to vitamin D stimulus. The role of vitamin D in anti-bacterial immunity has been shown to be mediated by components of the innate immune system. Binding and activation of toll-like receptors (TLRs) by APC cells has been shown to induce the expression of both VDR and the anabolic enzymes for vitamin D synthesis (263). Microarray-based studies have revealed that gene targets of $1,25D_3$-VDR signalling include anti-bacterial proteins: cathelcidin (264) and β-defensin-2 (*DEFB4*)(265).

The inhibitory role of vitamin D ('anti-immune') have also been reported. Two independent studies have demonstrated the role of vitamin D_3 as an inhibitor of T-cell proliferation that blocks transition from early G_1 phase to late G_1 phase (266, 267). In addition, $1,25D_3$ has been shown to promote regulatory T-cells (T-regs) by mechanisms that involve Antigen Presenting Cells (APCs) (268). In the context of B-cells, $1,25D_3$ has been shown to suppress B-cell maturation into plasma cells and class-switching memory cells as well as to regulate IL-10 (269) and CCR10 (270).

In summary, the current evidence suggests that there is indeed interaction between hormonal vitamin D and components of the immune system. However, the direction of interaction could be dictated by cellular context and a complex set of

feedback loop mechanisms, which could explain the apparent contrast in effects on the immune system. There are strong data suggesting vitamin moderates and interacts with immune cells both epidemiological and *in vitro*/*in vivo* but the literature describes enormous complexities which I feel are as yet unresolved.

1.8.5 Vitamin D signalling in cancer

1.8.5.1 Epidemiological evidence

The scepticism around the associations between high vitamin D levels and better outcomes is generally based upon the observation that people who are leaner, fitter and wealthier in many countries. Furthermore, that supplementation trials for a number of conditions were negative except for a general increased overall survival in a meta-analysis performed by Autier (271). The suspicion was that the association between lower vitamin D levels and reduced risk of a variety of conditions had occurred because of reverse causality[iii]. In particular, the negative correlation between low vitamin D levels and high levels of C-reactive protein (CRP) (273) suggested the association between low vitamin D levels and ill health, rather than low vitamin D levels were acting as a biomarker of systemic inflammation. Despite this scepticism, epidemiological studies have reported associations of serum vitamin D levels with cancer risk and mortality in colorectal, breast, prostate, bladder, lung, melanoma and other skin cancers (274).

Colorectal cancer: the strongest consensus evidence that low serum/plasma 25(OH)vitamin D is associated with increased cancer risk exists in the case of colorectal cancer. This evidence stems from meta analyses (275-277) as well as nested case-control studies (278, 279).

Breast cancer: the evidence for 25(OH)vitamin D and risk of breast cancer is not as consistent compared to colorectal cancer. However, a meta-analysis of 8 studies showed that higher circulating 25(OH)vitamin D was associated with lower risk of breast cancer incidence (280). However, there is lack of concordance of findings between retrospective and prospective studies: the inverse association of circulating 25(OH)vitamin D_3 with breast cancer risk seemed to be restricted to retrospective studies such as by Shao et al (280), with null association for prospective studies (275,

[iii] Reverse Causality: reverse causality is when the exposure-disease process is reversed. In other words, instead of the exposure causing the disease (causality), the event of having the disease (e.g. Being diagnosed with a disease) can cause a change in the pattern of the exposure (e.g. Lifestyle changes in relation to the disease).(272. To SH. Statistics How To. https://www.statisticshowto.datasciencecentral.com/reverse-causality/.

281). This discordance between retrospective and prospective studies indicates the possibility of bias, in particular of reverse causality as mentioned above.

Prostate cancer*:* among the various cancer types whose association with circulating vitamin D have been investigated, prostate cancer is the only malignancy with evidence for a positive association with cancer risk i.e. higher serum vitamin D has been reported to be associated with increased risk of prostate cancer. For instance, a meta-analysis of 21 studies reported that in 16 of the studies, men with higher serum levels of 25(OH)vitamin D_3 had an increased risk of prostate cancer, compared to men with lower 25(OH)vitamin D_3. On the other hand, prostate cancer incidence and mortality in African-American males, who have reduced cutaneous synthesis of vitamin D owing to skin pigmentation, is higher than their Caucasian counterparts (282). But this effect could be confounded by the disparity in medical care or socioeconomic status. Taken together, the association between prostate cancer incidence and circulating vitamin D levels is still unclear and the focus of ongoing research efforts.

Bladder cancer: there is evidence for inverse relationship between high serum 25(OH)vitamin D_3 and reduced risk of bladder cancer, based on meta-analyses (283, 284).

Lung cancer: two meta-analyses assessing the association between circulating 25(OH)vitamin D_3 and risk of lung cancer indicate an inverse relationship i.e. patients with low circulating 25(OH)vitamin D_3 had a higher risk of lung cancer (285, 286).

Skin cancers: when considering the relationship between skin cancer and vitamin D, it is important to note an important factor common to both: UVB radiation. While UVB is an important part of skin cancer aetiology, it is also necessary for the cutaneous synthesis of vitamin D. This relationship has been discussed in section 1.4.4.

Lower incidence of non-melanoma skin cancers (NMSC) has been reported in subjects with highest serum vitamin D concentrations, compared to those with lowest serum vitamin D concentrations (287).

Melanoma: several studies have investigated the role of serum vitamin D in melanoma. Our research group has particularly focussed on this aspect. In a small retrospective studied carried out in Leeds, designed to identify lifestyle factors associated with late relapse from melanoma, participants who had suffered a relapse were less likely to report taking vitamin D supplements than controls who were melanoma cases who had not relapsed after 5 years (256). The group then carried

out a prospective study of 872 patients in the Leeds Melanoma Cohort, in which higher 25(OH)vitamin D_3 (henceforth referred to as vitamin D) levels at recruitment were found to be associated with lower Breslow thickness and better Melanoma Specific Survival (MSS) (288). Subsequently a similar relationship between vitamin D levels and stage at diagnosis was reported in studies from France (289) Germany (290), America (291) and Australia (292). Fang et al reported a similar study which also addressed the issue of the degree to which vitamin D levels might merely be a surrogate for higher levels is systemic inflammation. In that study Fang et al showed that low serum vitamin D levels were associated with higher C-Reactive Protein levels (CRP-a marker of systemic inflammation, frequently higher in patients with the metabolic syndrome/obesity), higher ulceration, increased tumour thickness and poor MSS. The crucial result was that the relationship between vitamin D levels and survival was independent of CRP level.

In summary, there is strong evidence that low levels of vitamin D at diagnosis are associated with thicker melanomas and poorer prognosis in 4 continents. This effect seems to be independent of the CRP level. The strong negative correlation between CRP and vitamin D levels is however of interest and it has been suggested by Amer (273) and others that some of the beneficial effects of vitamin D might be a result of suppression of systemic inflammation.

1.8.5.2 In vitro evidence

The first study to indicate the anti-proliferative activity of $1,25(OH)_2D_3$ was the study by Colston et al, who elegantly demonstrated that $1,25(OH)_2D_3$ had high affinity for VDR protein and decreased the doubling time of melanoma cells in a dose-dependent manner (293). This was followed by a mounting evidence for anti-tumour activity of vitamin D-VDR signalling from *in vitro* studies performed in the 1990s and early 2000s. These studies were designed to assess a specific aspect of the anti-tumour effects of vitamin D treatment, such as its effect on proliferation, apoptosis, invasive potential and DNA repair. In the context of proliferation, a 'targeted' approach was taken, wherein the effect of $1,25(OH)_2D_3$ (or its analogues) on a particular cell cycle and/or proliferation-associated pathway was assessed. For instance, $1,25(OH)_2D_3$ was shown to induce cell cycle arrest in squamous cell carcinoma (294), breast cancer (295), leukaemia, myeloma and colon cancer (296). Efforts were also focused on identifying gene targets of the signalling axis. A study using the myelomonocytic cell line U937 demonstrated that *p21 (waf1/cip1)* was differentially expressed in response to $1,25(OH)_2D_3$ treatment and also contained a

VDRE (297). The gene was later confirmed to be a primary VDR target gene using ChIP-Seq (298). $1,25(OH)_2D_3$ treatment has also been shown to cause stress-induced apoptosis by upregulation of *VDUP* (Vitamin D Upregulated Protein- a vitamin D transcription target (299)) which neutralizes thioredoxin (300, 301). It was recently demonstrated that VDR expression in various normal as well as cancer cells played a role in avoiding impaired mitochondrial function and eventual cell death (302). The effect of vitamin D-VDR signalling in tumour invasion and metastasis has been explored by Munoz et al who have used a colorectal cancer cell models to describe a mechanistic basis for this effect. $1,25(OH)_2D_3$ was shown to repress *DKK4* (303) as well as Wnt/β-catenin signalling (304), both of which were shown to promote invasion and angiogenesis. Expression of SNAIL, which promotes invasiveness, progression and poor prognosis was shown to inhibit VDR expression and response to $1,25(OH)_2D_3$, indicating that vitamin D-VDR signalling was inversely related to invasion-associated prognosis (305). Studies to assess the role of VDR in DNA-damage response stemmed from the link between VDR and the p53 pathway, in that they're known to physically interact and share 'common' target genes (306).

In addition to the studies described in the previous paragraph which have investigated the genomic and functional impact of vitamin D-VDR signalling, there are many others which have investigated the epigenetic effect of this signalling axis. This effect on the epigenome is of significance, given its ability to affect key biological processes (307). $1,25(OH)_2D_3$ treatment has been shown to decrease promoter methylation of E-cadherin and thus increase expression in MDA-MB-231 breast cancer cells (308)

In the case of melanoma, one of the earliest *in vitro* experiments to demonstrate the anti-proliferative effect of vitamin D was done in human melanoma Hs695t cell line, whose doubling time was significantly increased upon vitamin D treatment (293). More recently Reichrath et al tested the effect of vitamin D on a panel of metastatic melanoma cell lines and showed that only some of the cell lines were responsive to the anti-proliferative effects of vitamin D, putting forth the notion that vitamin D signalling exerts its impact on a subtype of metastasizing cell types (309). In addition, the pro-differentiation roles of vitamin D analogue were demonstrated in the human metastatic cell line SKMEL-188 (310). Efforts to investigate the prospective role of VDR in melanoma include studies that look into polymorphisms in the gene that codes for VDR. The study conducted in Leeds reported VDR alleles that were associated with increased or decreased risk of MSS (311).

Since one of the focuses of my thesis is to assess the role of vitamin D in melanoma using transcriptomic and clinical data, I have elaborated more on this topic in a dedicated chapter: ‘Chapter 3: Vitamin D-VDR signalling in melanomas’.

Chapter 2
Description of data and Methods

This chapter includes a description of two main data sets used in this thesis: the Leeds Melanoma Cohort (LMC) and The Cancer Genome Atlas (TCGA) melanoma datasets. This chapter also details the methods and relevant materials used throughout the thesis while more details pertaining to Chapter-specific analyses are detailed in the Methods section of that chapter.

2.1 The Leeds Melanoma Cohort (LMC)

The Leeds Melanoma Cohort (LMC) is composed of 2184 population-ascertained participants of primary melanoma from the North of England. Invitations to participate in the study were extended 3 months after initial diagnosis of disease. If the invitation was accepted, the intent was to interview and obtain biological specimens (such as blood, excised tumour) within 3-6 months after diagnosis. The time to interview (from the initial invitation to participate) was variable, with a median of 5.2 months. The total recruitment period extended between 2001-2013. Participants who consented to participate ('participants' henceforth) completed detailed questionnaires, which queried various lifestyle factors. The parts of the questionnaire which are relevant to my research projects are: age, sex, dietary supplement intake and sun-exposure information. Variables pertaining to the melanoma tumour itself which were obtained from histopathology reports include Breslow thickness, ulceration status, vascular invasion, tumour site and tumour mitotic rate. Patient survival information was also recorded and periodically updated using national databases such as the ONS (Office for National Statistics) and the National Cancer Registry.

Approvals for the study have been granted by the Multicentre Research Ethics Committee (MREC) (1/3/057) and the Patient Information Advisory Group (PIAG) (3-09(d)/2003).

2.1.1 Subset of 703 participants from the LMC used in this thesis

The initial Leeds Melanoma Cohort is composed of 2184 melanoma participants (as described above). For a subset of these participants (n=703), the primary melanomas which were excised at diagnosis were processed into Formalin-fixed paraffin-embedded (FFPE) tumour blocks. The tumour blocks were then used to generate

Tissue Microarray (TMA) cores, from which mRNA was extracted and used to generate transcriptomic profiles (described in section 2.2). For the purposes of this thesis: the clinical, histopathological and transcriptomic data pertaining to this subset of 703 participants will be referred to as the 'LMC dataset', despite the 'original' Leeds Melanoma Cohort being composed of 2184 participants. **To this effect, any reference to 'LMC dataset' in this thesis, refers to the subset of 703 participant.** Lifestyle, histopathological and survival variables pertaining to these 703 participants, which have been used for analyses in this thesis are summarised below.

Treatment status: Only 10 of the 703 participants had been treated with BRAF inhibitors, 10 with Ipilimumab and 2 with Pembrolizumab.

Age: Age at diagnosis, in years. The median age of the 703 participants was 58.37 years and Standard Deviation (SD) was 12.88 years.

Sex: Sex at diagnosis, indicated as Male or Female (self-reported in questionnaire)

Breslow thickness: the measurement of Breslow thickness is described as:

"Breslow thickness is measured from the top of the granular layer of the epidermis (or, if the surface is ulcerated, from the base of the ulcer) to the deepest invasive cell across the broad base of the tumour (dermal/subcutaneous) as described by Breslow." (1, 2).

In the LMC dataset of 703 participants, the minimum and maximum Breslow thickness were 0.33 mm and 20 mm respectively, with a median of 2.3 mm and SD of 2.30 mm.

Ulceration: melanoma ulceration is defined as : "full thickness absence of an intact epidermis above any portion of the primary tumour with an associated host reaction (characterized by a fibrinous and acute inflammatory exudate) above the primary tumour based on histopathological examination." (2).

In the dataset of 703 participants, ulceration status was assigned 'yes' or 'no' by the reviewing pathologist.

Mitotic rate: Mitotic rate is defined as the number of tumour cells which are mitotically active per square millimetre of tumour area, as estimated by the reviewing pathologist. In the dataset of 703 participants, mitotic rate data was available for 595 tumours, in which the minimum and maximum mitotic rate were 0 and 83 respectively, with median value of 3 mitoses per square millimetre of tumour area and SD of 9.18 mitoses per square millimetre of tumour area.

AJCC stage: All tumours were classified according to the recommendations of the 7th edition of the AJCC (312) Melanoma staging system (313). Though a more recent (8th edition) of the AJCC staging system was published, the LMC cohort participants were diagnosed (and subsequently treated) based on the 7th edition's recommendation and hence this classification was retained. However, a reclassification of the 703 participants by 8th edition's recommendation revealed that only 3 participants were discordantly classified compared to the 7th edition-based classification (verified by Mrs. Joanne Gascoyne in the Leeds group). Taken together, the classification by 7th edition still remains relevant despite the recently updated recommendations and hence was used as such in all analyses in this thesis.

Vascular invasion: vascular invasion indicates the presence (or absence) of tumour cells which are fixed to the walls and within the lumens of lymphatic or blood vessels (3). In the dataset of 703 participants, vascular invasion status was available for 626 tumours, for which vascular invasion status was assigned 'yes' or 'no' by the reviewing pathologist.

Melanoma survival/death: survival information for participants was obtained both directly (by annual re-contact) and indirectly from review of national cancer registries and the ONS. In the case of deceased participants, the cause of death was obtained from death certificates and medical records. This was reviewed by research nurses in the Leeds Melanoma Research group led Prof. Julia Newton-Bishop, to generate Melanoma Specific Survival (MSS).

Distribution of 703 LMC participants based on survival data and clinicopathological features defined above is detailed in Table 2.1. Because of missing data for some variables, information is not complete for some participants, but overall the coverage is very good.

Table 2.1 Distribution of 703 LMC participants based on age, sex and histopathological variables

Variable	Number of participants
Age at diagnosis (in years, median= 58.37 years)	
≤58.37 years	352
>58.37 years	351
Total	703
Sex	
Females	385
Males	318
Total	703
Breslow thickness (in mm, median= 2.3mm)	
≤ 2.3 mm	357
> 2.3 mm	346
Total	703
Ulceration status	
Yes	235
No	468
Total	703
Mitotic rate (number of mitoses/square mm of tumour, median= 3 mitoses/square mm of tumour)	
≤ 3 mitoses/square mm of tumour	301
> 3 mitoses/square mm of tumour	402
Total	703
AJCC Stage	
Stage I	233
Stage II	355
Stage III	107
Stage unavailable	8
Total	703
Vascular invasion status	
Yes	69
No	558
Unavailable	76
Total	703
Melanoma survival	
Alive	470
Dead	233
Death from melanoma-specific causes (MSS)	196
Death from non-melanoma causes	36

2.2 Generation of the LMC transcriptome

Primary melanoma tumours excised from participants and stored as FFPE (Formalin Fixed Paraffin Embedded) blocks in various hospitals in the North of England were mailed to our lab, with participants consent, as per the standard operating procedures (SOP) developed by the Leeds Melanoma Research group, in compliance with the ethical approvals and the Human Tissue Act. The FFPE blocks were then sectioned, H&E stained and reviewed for sampling. Tumour cores were selected after examination under microscope then sampled using a TMA (Tissue Microarray) needle. RNA was extracted and used to generate transcriptomic data which will be referred to as the 'LMC transcriptome' henceforth. The steps involved in generating and pre-processing the LMC transcriptome are detailed below. These steps and were performed prior to the commencement of my PhD project.

2.2.1 Tumour sampling, expression profiling and processing

FFPE tumour blocks were sectioned to produce 5μm sections mounted on glass slides to be used for Haematoxylin and Eosin (H&E) staining. Protocol for H&E staining is described in section 3.3. The H&E stained slides were reviewed by Prof. Julia Newton-Bishop and Dr Jonathan Laye (Senior Histopathologist) to identify a region of the tumour that was suitable for sampling. The intent was to identify the deepest part of the tumour which had the highest tumour cell content and least stromal invasion. The tumour region satisfying these criteria were marked as a core with a marker on the H&E slides of primary tumours obtained from the 703 participants. The sampling of the core tumour was done on the TMA apparatus. Each marked H&E slide was aligned over its respective tumour block and the TMA needle was used to guide a horizontal 0.6mm core through the block. The contents of the core were then stored at 4^{o}C before subsequent RNA extraction. The tissue cores were dewaxed in xylene and absolute ethanol (two changes) after which RNA was extracted using High Pure Paraffin RNA kit (Roche Diagnostics) according to the manufacturer's protocol and eluted in nuclease-free water. Transcriptomic expression was quantified using the Illumina DASL (cDNA-mediated Annealing, Selection, extension and Ligation) Human HT12 v4 array (whole genome) by a service provider: Service XS (Leiden, Netherlands). The processing and normalisation of the microarray data was performed by Dr Jeremie Nsengimana (Senior Statistician). Briefly: the microarray image data files (quantifies gene expression as fluorescence intensity) was processed in the Illumina proprietary software GenomeStudio to obtain the raw numerical data. The R package Lumi (314) was used to background-correct and quantile-normalise the data. Technical

variabilities which could confound 'true' variability of gene expression such as batch, chip, age of the FFPE block and RNA concentration were identified using the R package Swamp (function used: quickadjust.zero) (315) and adjusted out. In addition, outliers were identified by examining raw and normalised density plots. Post these quality control processes, expression values pertaining to each probe was standardised to mean 0 and variance 1. These data have been published (316) and have been deposited in the European Genome-Phenome Archieve (EGA) at the European Bioinformatics Institute (EBI), with accession number EGAS00001002922.

2.2.2 Choice of probes, probe to gene mapping strategy

The Illumina DASL Human HT12 v4 array includes multiple probes which are designed to span different regions of a particular gene, allowing isoform specific analyses. Studying isoform specific mechanisms can be of interest for certain genes if taken as candidates, but it is less informative in agnostic whole-genome studies, such as those conducted in this thesis. It was thus necessary to select one probe per gene. To this effect, a 'probe-to-gene' mapping strategy was implemented. Information regarding the number of isoforms 'covered' by each probe was downloaded from the Illumina product support website (https://support.illumina.com/downloads.html). This document matched each probe as:

- A = All isoforms. The probe is designed to hit all splice isoforms of a gene.
- I = Isoform specific. The probe is designed to hit a specific splice isoform of a gene, for which multiple isoforms are known to exist.
- S = Single isoform. The gene has only one known splice isoform and our probe hits it.
- M = Multiple isoforms. This gene has multiple isoforms. The probe targets more than one and fewer than all of them.

The proportion of samples for which a particular probe was detected with reliable intensity (above the background noise) was also known for each probe, referred to as 'Proportion Detected'. Among the total of 29,354 probes on the array, those which were isoform specific ('I') and detected only one splice isoform of the gene, despite the gene being known to have multiple isoforms were dropped (4644 probes). Among the remaining probes, if a gene still mapped to more than one probe, then the probe with lowest Proportion Detected was dropped. This resulted in 20,560 unique probe-gene pairs. This strategy of probe filtration ensured that a gene was not represented by probes which i) spanned only one isoform of the gene, with no

information regarding the other isoforms, despite being detected reliably in a majority of the samples (represented by high Proportion Detected) and ii) spanned multiple isoforms of the gene but detected unreliably in a majority of the samples (represented by low Proportion Detected).

2.3 Copy Number Alteration (CNA) data in the LMC

Copy number alternation profiles were estimated from Next-Generation Sequencing (NGS) data generated from DNA extracted from 303 LMC tumours. However, only 276 tumours (of these 303 tumours) are a subset of the 703 samples for which transcriptomic data was generated. In other words, 'matching' NGS and transcriptomic data was available for a set of 276 tumours.

Quality control of the NGS data included normalisation to control samples from the 1000 Genomes Project (Phase 3, 1KGP, n=312, ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase3/data/) that matches our experimental set up (Illumina platform, low coverage, paired end library layout). These data were generated by Dr Anastasia Filia, a former PhD student in the Leeds Melanoma Research group and were in part pre-processed by Joey Mark Diaz, a colleague Marie-Curie Research Fellow in the same group. Joey used the package *bamwindow* (https://github.com/alastair-droop/bamwindow) to create bins or windows of size 10k across the genome, identified and excluded blacklisted regions (those known to generate unreliable sequences. QDNASeq package in R was used to identify highly variable regions in the genome which were added to the blacklist and to adjust the read counts on each valid window based on the interaction between GC content and Mappability (317). Log2 ratio of window read counts between each LMC sample and the median in each window of the 1KGP normal samples were obtain and segmented. Segmentation of copy number was performed using circular binary segmentation available in the R package *DNAcopy* (318). Gistic2.0 (319) was used to identify significant copy number aberrations peaks and gene level copy number estimate. These DNA Copy Number Alteration (CNA henceforth) data were used in my thesis in conjunction with transcriptomic profiles to investigate or validate mechanisms of gene regulation.

2.4 Statistical methods

Statistical tests performed for analyses pertaining to relevant chapters are detailed in the Methods section of that chapter. All statistical analyses were performed in Stata/SE 14.2 (320) unless otherwise specified.

2.5 Gene and pathway enrichment analyses

All gene and pathway enrichment analyses were performed in Reactome FIViz (321) on the Cytoscape (version 3.5.1) (322) desktop feature. Reactome FIViz is an app designed to discover pathways and network patterns associated with gene expression profiles from cancer and other types of disease. The app allows access and use of the Reactome Functional Interaction (hence the name Reactome FI) network, which is a manually curated database. The unique feature of the Reactome Functional Interaction network is that it is constructed from two types of data sources:

i. High-coverage, pairwise networks derived from experimental sources (such as yeast two-hybrid techniques, mass spectrometry pull down assays and DNA microarrays), databases such as BioGrid, Database of Interacting Proteins (DIP), the Human Protein Reference Database (HPRD), I2D, IntACT, and Molecular Interaction Database (MINT) and expression data sets from the Stanford Microarray Database and the Gene Expression Omnibus. While these datasets provide ample pairwise interaction data, interaction does not necessarily indicate a biologically-relevant relationship. As a consequence, these data sources tend to have high degree of false positivity.
ii. Low-coverage, highly-curated pathway data derived form databases such as Reactome, IntAct, Protein Analysis Through Evolutionary Relationship (PANTHER), Kyoto Encyclopaedia of Genes and Genomes (KEGG) and BioCarta. These databases are based on manually-curated interactions between proteins/genes.

The authors use a naïve Bayes classifier approach to identify high-likelihood functional interactions from non-functional interactions and false positives (321, 323). Taken together, Reactome Functional Interaction network was created by combining the Interaction data (in the form of interaction between two genes/proteins) from both data sources listed above i.e. combining high-coverage, unreliable pairwise data and low-coverage, curated pathway data.

It is for this reason that I chose to use Reactome FIViz (the web-based tool which works on the basis of the Reactome Functional Interaction network) for my pathway enrichment analyses.

In order to identify the pathways which are enriched in any given list of genes (e.g. list of genes that correlate positively with tumour *VDR* expression), the following steps were performed:

i) The Cytoscape app was launched and the Reactome FI plugin was chosen in the 'Apps' tab. From among the features of Reactome FI, the 'Gene set/Mutation analysis' option was chosen. The latest version Reactome FI network version (2015 update) was chosen to be used as the basis of the enrichment analysis.

ii) The list of gene names which needs to be queried is entered manually or uploaded as a text file. Upon choosing the 'perform Pathway Enrichment' option from the dropdown options, the list of genes is compared with the 'background' Reactome FI network. The output from this comparison is represented by a P-value (from a hypergeometric test) and FDR (Benjamini-Hochberg) for each identified pathway.

    For instance, for an input query gene list, *m* genes were identified to belong to a certain *Pathway-x*, which is composed of *n* genes (as defined by the Reactome FI network) then the following values are computed (done automatically by by Reactome FIViz):

    - *Ratio of protein in gene set*: ratios of numbers of genes contained in pathways to total genes in the Reactome FI network (*n/total genes in Reactome FI 2015 version*)
    - Number of protein in gene set: numbers of genes in pathways (n)
    - *Protein from network*: numbers of hit genes from the query gene list (*m/n*)
    - *Nodes:* nodal genes from my input query gene list, which 'matches' with the genes in *Pathway-x.* in other words, these are the names of the *n* genes.

In addition, the source database for each of the enriched pathways are indicated by a letter in parentheses after each pathway gene set name. The source database annotations are:

- C - CellMap,
- R – Reactome,
- K – KEGG,

- N – NCI PID,
- P - Panther, and
- B – BioCarta.

In the case of my analyses, I restricted the results to pathways which had FDR<0.05, which I used for further interpretation. Upon filtering by this threshold, I exported the 'pathway table' which was composed of the names of enriched pathways, the nodal genes pertaining to each pathway, *Ratio of protein in gene set, Number of protein in gene set* and *Protein from network.* For all the enrichment analyses I have performed, the exported pathway table with all of the aforementioned information is attached in the Appendix. In the main text of the Results section, I have included a concise table of the top pathways, the respective FDR values and nodal genes.

2.6 TCGA data

The Cancer Genome Atlas (TCGA) metastatic melanoma data such as transcriptomic (RNA-Seq), clinical, methylation and copy number data were downloaded from cBioPortal (http://www.cbioportal.org/). Of note, the TCGA expression profiles were generated from fresh-frozen tissue using RNAseq platform and they were all from highly selected and advanced tumours (as opposed to FFPE, array platform and unselected population design in the LMC). For all the analyses, the same statistical tests and software/packages were used to analyse the LMC and the TCGA datasets. Since the TCGA data is a composed of participants from multiple clinical sites, there was the issue of incomplete or unreliable data for some variables. In my thesis, I have used transcriptomic (RNASeq), CNA (NGS) and methylation (HM450K array) data pertaining to the melanomas samples from the TCGA melanoma cohort. The availability of each of these data in subsets of the TCGA cohort are described in Table 2.2.

Table 2.2 Table indicating availability of data pertaining to the TCGA melanoma cohort

	Transcriptomic data available?	CNA data available?	Methylation data available?
Primary melanomas (n=103)	Yes	No	Yes
Metastatic melanomas (n=365)	Yes	Yes	Yes

2.7 Haematoxylin and eosin staining

FFPE sections from the LMC primary melanomas (Chapter 3) and mouse lungs (Chapter 4) were counterstained with Haematoxylin and Eosin (H&E), after staining with the respective antibodies. The protocol used for H&E staining is as follows:

1. The slides after treatment with the secondary antibody step, were washed 3x in wash buffer. This was followed by submerging the slides in Meyer's haematoxylin for 30 seconds
2. Slides were then rinsed in tap water till the water went clear
3. Slides were then submerged for 2 minutes in Scott's tap water
4. Slides were washed in tap water for 1 minute
5. Slides were dabbed with paper towels to remove excess water, but without dehydrating the slides. This was followed by sequential hydration in gradient ethanol solutions followed by xylene treatment.
   100% ethanol for 15 seconds → 100 % ethanol for 2 minutes → 80% ethanol for 3 minutes → 80% ethanol for 3 minutes → 100% xylene for 3 minutes → 100% xylene for 3 minutes → 100% xylene for 3 minutes → slides were mounted with DPX slide-mountant → Slides were air dried overnight and used for subsequent review.

Chapter 3
Vitamin D-VDR signalling in melanoma

3.1 Introduction

3.1.1 Vitamin D-VDR signalling: an overview

The hormone 1α,25-dihydroxyvitamin D_3 (vitamin D_3 henceforth) is the biologically active form of vitamin D and is the ligand for the dimeric Vitamin D Receptor (*VDR*) and Retinoid X Receptor (*RXR*). The 'classic' model of vitamin D signalling involves ligand binding to the receptor dimer, facilitating transcription of target genes that contain the Vitamin D Response Element (VDRE) (230, 324). However, both genomic and non-genomic actions of vitamin D_3 have been described, as discussed below and are also pictorially depicted in Figure 3.1. The VDRE consists of two hexanucleotide repeats with a nucleotide triplet sandwiched in between (GGTCCA-NNN-GGTCCA): this is referred to as the DR3 sequence, which is present in the promoter region of target genes. While VDR occupies the 3' half site, RXR occupies the 5' half site of the double hexanucleotide in the promoter region of target genes to facilitate transcription (325). Vitamin D_3 has also been shown to repress gene expression in a mechanism that involves binding of the VDR-RXR dimer to negative-VDREs comprised of a CANNTG-like motif (referred to as VDR-interacting repressor: VDIR) of genes such as *PTH* (238). Vitamin D_3 has also been shown to induce association of VDR and VDIR, leading to recruitment of HDAC co-repressor to repress *CYP27B1* expression (326, 327). On the other hand, transcriptional activation by vitamin D_3 involves binding of activated VDR-RXR dimer to the DR3 complex (as described above) to facilitate transcription of targets such as *CYP24A1* (328), *BGLAP* (240) and *CDKN1A (297).* In addition to the classic genomic response, vitamin D_3 is also known to elicit certain non-genomic responses that are transcription-independent and rapid (within seconds to minutes, depending on tissue type) such as the rapid intestinal absorption of calcium (transcaltachia), opening of voltage-gated Ca^{2+} and Cl^- channels in osteoblasts and rapid migration of endothelial cells (251). This has been shown to be mediated by signalling cascades such as Protein Kinase C (PKC), leading to increase in intracellular Ca^{2+} (329). Increased intracellular Ca^{2+} can in turn lead to activation of pathways such as the Raf-MEK-MAPK-ERK cascade, which mediates proliferation in skeletal muscles (330) and normal colon cells (329). In addition, non-genomic activation of PKC has been shown to increase transcriptional activity of VDR by phosphorylation-dependent stabilisation (331, 332). This demonstrates the overlap in cell signalling components that are influenced by

genomic and non-genomic effects of vitamin D_3 and suggests a cooperative relationship that impacts the overall downstream effect of vitamin D_3 (summarized in Figure 3.1).

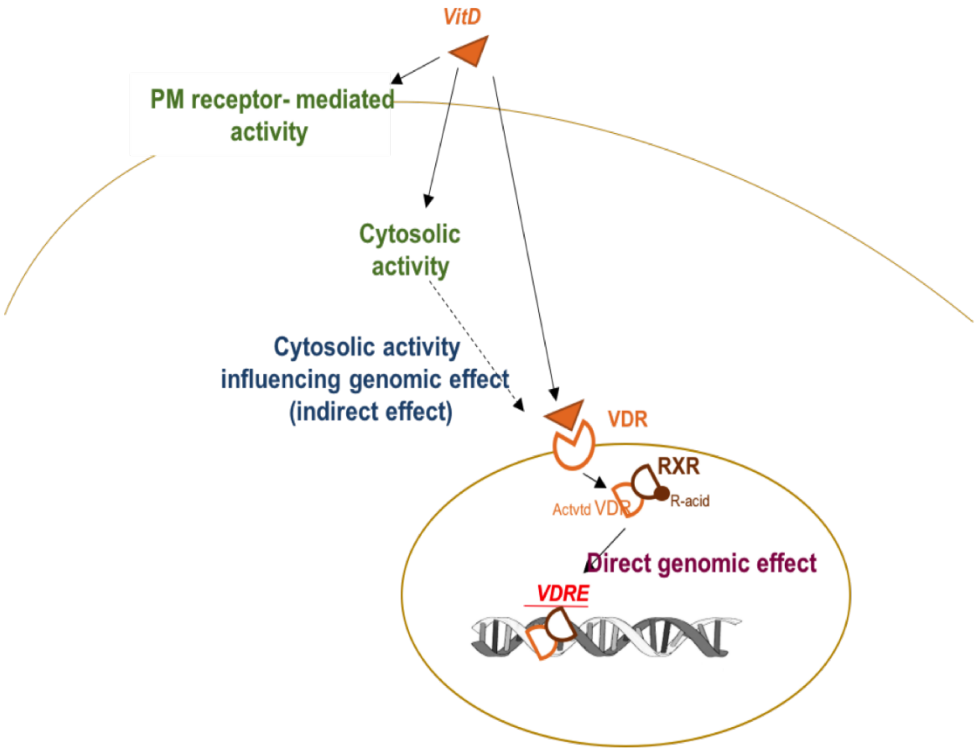


Figure 3.1: The possible modes-of-action of vitamin D

PM: Membrane level effect where vitamin D could bind and signal via non-canonical receptors, Cytosolic effects include possible influence of vitamin D on cytosolic signalling factors that could affect downstream signalling pathways and direct genomic effect is the vitamin D-VDR signalling axis via VDRE-containing genes. RXR-Retinoid Receptor, R-acid-Retinoic acid. Adapted from (333).

3.1.2 Physiological effects of vitamin D-VDR signalling

The physiological effect of the vitamin D-*VDR* signalling axis is context specific: in the intestines and bones it is known to mediate calcium and phosphate homeostasis (334) while in the epidermis it contributes to stratum-specific keratinocyte differentiation (335, 336), anti-microbial innate immune response (337, 338) and reduced keratinocyte proliferation (339). Vitamin D deficiency has historically been of interest in the context of several aspects of public health (340-342) given its reported association with risk of tuberculosis (343), autoimmune disorders such as SLE (344), arthritis (260), type I diabetes (345), and the incontrovertible evidence for its role in bone and muscle health.

It is worth noting that *in vitro* studies indicate a synergistic relationship between vitamin D and *VDR*, wherein both vitamin D_3 treatment and intact *VDR* expression are necessary for downstream effects, as would be expected between a ligand and its canonical receptor. However, this does not preclude ligand-independent effects of *VDR.* Studies to understand the ligand-independent effects of *VDR* have been used to disentangle the vitamin D- *VDR* functional relationship and to provide a mechanistic basis for '*VDR*-specific' physiological functions (346). For instance, unliganded *VDR* has been shown to bind to a novel corepressor Hairless (*HR*) (347) to affect keratinocyte/hair-follicle homeostasis (348). Notably, a study by Trivedi et al demonstrated a ligand-independent effect of VDR on the proliferation of MCF-7 breast cancer cells (349), suggesting that cancer cells are amenable to ligand-independent effects of VDR. Taken together, these data suggest that the vitamin D-*VDR* signalling axis may consist of both an independent and synergistic relationship between the ligand and the receptor, with distinct physiological and pathological consequences.

3.1.3 Vitamin D signalling in cancer: summary of *in vitro* and *in vivo* evidence

Studies that demonstrate the effect of vitamin D_3 signalling on tumour progression have used *in vitro* cancer cell line models and *in vivo* VDR knockout mouse models to understand the mechanism of action of vitamin D signalling. Genomic profiles of *in vitro* responses to vitamin D_3 treatment in cell lines of prostate cancer (350), breast cancer (351), leukaemia (352), colon carcinoma (353) and squamous cell carcinoma (265) indicate that the genes/pathways corresponding to the following cellular processes are perturbed in response to vitamin D treatment: DNA repair, cell cycle progression, apoptosis, cell adhesion, metastatic potential, differentiation, membrane transport and growth regulation. Vitamin D_3 has been shown to exert its anti-proliferative effects by inhibiting cell cycle progression into S-phase (354), inhibiting pro-proliferative pathways such as TGF-β (355) and Wnt/β-catenin (356) and mitogenic pathways involving EGF/MAPK/MEK/ERK (357) and IGF (358). VDR-null mice showed colorectal cell hyperproliferation in the absence of vitamin D_3-mediated anti-proliferative effects (359), thus corroborating the aforementioned *in vitro* observations. Vitamin D_3 has also been shown to have a potent anti-angiogenic effect on tumour-induced angiogenesis in BALB/c mice (360) and in tumour-derived endothelial cells (361). In addition, vitamin D_3 has been shown to diminish invasive phenotype and lung metastasis by decreasing activity of cell-adhesion proteins in VDR-null mice (362) and in mammary epithelial cells (363).

3.1.4 Vitamin D signalling in melanoma: *in vitro* and *in vivo* evidence

The relationship between vitamin D and melanoma is particularly interesting given that ultraviolet light is responsible not only for vitamin D production but also for melanoma carcinogenesis. One of the earliest *in vitro* experiments to demonstrate an anti-proliferative effect of vitamin D_3 was done in human melanoma Hs695t cell line, whose doubling time was significantly increased upon vitamin D_3 treatment (293). More recently Reichrath et al tested the effect of vitamin D_3 on a panel of metastatic melanoma cell lines and showed that only some of the cell lines were responsive to the anti-proliferative effects of vitamin D_3, putting forth the notion that vitamin D_3 signalling exerts its impact on a subtype of metastasizing cell types (309). In addition, the pro-differentiation roles of vitamin D_3 analogues were demonstrated in human metastatic cell line SKMEL-188 (310).

Multiple research groups have also investigated the *in vitro* effects of non-calcemic vitamin D metabolites, as an alternative to vitamin D_3. The primary rationale behind this being: the hypercalcemic effects (elevated calcium levels in the blood stream) of vitamin D_3 can cause medical complications as severe as organ failure (364). Non-calcemic metabolites which have comparable physiological potency could potentially circumvent this issue and hence have been extensively explored as therapeutic options. One such non-calcemic metabolite is 20-hydroxyvitamin D_3. *In vitro* studies by Slominski et al have demonstrated that the vitamin D metabolite 20-hydroxyvitamin D_3 has an anti-proliferative effect on melanoma cells (310, 365) by inhibition of NFκB signalling (366). A more recent study by the same group elegantly exhibited that 20(OH)D_3 inhibits cell migration and cell-cell adhesion *in vitro* while reducing tumour load *in vivo* with no evidence of toxicity (367).

As for the effect of *VDR* expression on melanoma*,* Brożyna et al used an immunostaining-based approach to show that high *VDR*-expressing tumours were less advanced (decreased ulceration, Breslow thickness and increased immune infiltrate) and had better prognosis compared to low *VDR*-expressing tumours (368). Though the effects of VDR expression on tumour progression have been explored extensively in colon cancer, the evidence for melanoma are currently limited to the study by Brożyna et al. Thus, one of the aims of my PhD project (and this chapter in particular) is to interrogate the role of *VDR* in melanoma progression using melanoma transcriptomic and clinical data.

3.1.5 Vitamin D signalling in melanoma: epidemiological evidence

The current evidence for a protective role of vitamin D-*VDR* signalling in melanoma survival comes from both epidemiological studies and *in vitro* studies. In a prospective study of 872 patients in the Leeds Melanoma Cohort, higher 25-hydroxyvitamin D_3 levels were associated with lower Breslow thickness and better Melanoma Specific Survival (MSS) (288). This was followed by independent studies in American (291), Australian (292) and German (290) patient cohorts. In the American cohort (1,042 patients diagnosed with primary melanoma), lower vitamin D was associated with increased tumour thickness, advanced melanoma stage and high CRP (C-Reactive Protein) levels. Importantly, the associations with melanoma thickness and stage persisted after adjusting for CRP, which is a marker of systemic inflammation and high CRP levels predict poor prognosis. This finding indicated that though vitamin D and CRP are correlated, the association of vitamin D with tumour stage and thickness is not simply a reflection of systemic inflammation. In the Australian cohort (100 patients diagnosed with primary melanoma), serum vitamin D of <50nmol/L was associated with four-fold increase in risk of having a thicker tumour, compared to patients with ≥50nmol/L. In the German cohort (764 patients diagnosed with metastatic melanoma), once again, low serum vitamin D levels were associated with increased tumour thickness and stage. Taken together, it is worth noting that these studies were performed in populations from 3 different continents. Though the levels of serum vitamin D were variable amongst the different populations, low serum vitamin D was associated with increased tumour stage and thickness, when performing 'intra-population' comparisons.

Efforts to investigate the prospective role of VDR in melanoma include studies that look into polymorphisms in the gene that codes for VDR. The study conducted in Leeds reported VDR alleles that were associated with increased or decreased risk of MSS (311).

3.1.6 Relevance of my research in understanding the role of vitamin D signalling in melanoma

While previous studies have assessed the individual effects of either vitamin D or *VDR*, my research used both tumour *VDR* expression and corresponding serum vitamin D levels at diagnosis to assess their synergistic as well as individual contributions to melanoma survival and their genome-wide effects. The data set used was a unique set of 703 formalin fixed paraffin embedded (FFPE) primary melanomas, which are a subset of tumours from the Leeds Melanoma Cohort (LMC) (316). For these tumours, tumour core-derived transcriptomic data, extensive clinical data (including serum 25-

hydroxyvitamin D at diagnosis and accurate melanoma-specific survival), single observer histopathological data (such as quantified TILs and tumour mitotic number) and whole-genome Copy Number Alteration (CNA) data were used to assess and validate the pan-genome effects of vitamin D-*VDR* signalling.

The primary questions addressed in this chapter of my thesis are:

1. Are serum vitamin D levels associated with increased tumour thickness and stage in the cohort of 703 LMC primary melanomas, as has been previously reported by others?
2. Is transcriptomic tumour *VDR* associated with melanoma survival in the LMC primary melanomas?
3. Are serum vitamin D levels associated with melanoma survival, within the context of tumour *VDR* expression?
4. Given that tumour *VDR* expression is prognostically significant, what are the signalling genes, pathways and processes that are significantly associated with tumour *VDR* expression?
5. Can these findings be validated *in silico* using independent data sets and reported melanoma molecular phenotypes?
6. Is there causal evidence for the transcriptome-derived findings?

This is addressed in a separate chapter: Chapter 4, which is composed of the functional validation for findings from Chapter 3.

To address these questions, I have used a combination of statistical models and bioinformatic tools to interrogate the LMC transcriptome, which are detailed in the sections below.

3.2 Chapter aims and overview

Chapter aims	Section
To assess the association of vitamin D and *VDR* with clinicopathological features and expression of other vitamin D-VDR signalling pathway genes	3.4.1
To assess the association of vitamin D-VDR signalling with melanoma prognosis	3.4.2
To assess the effect of serum vitamin D on melanoma prognosis, within the context of tumour *VDR* expression	3.4.3
To identify the transcriptomic characteristics of melanomas pertaining to participants with very high serum vitamin D levels (>115nmol/L)	3.4.4
To identify the transcriptomic correlates of tumour VDR expression	3.4.5
To validate *in silico* the transcriptomic correlates of tumour VDR expression	3.4.6

Contributions to this chapter:

- Sathya Muralidhar performed statistical and bioinformatics analyses described in this chapter, under the supervision of Dr Jeremie Nsengimana (senior statistician in the group) and guided by Prof. Julia Newton-Bishop and Prof. Tim Bishop.
- Imputed immune cell scores (described in 3.4.3.2) were generated by Ms. Joanna Pozniak (PhD student in the group). This extensive work is described in detail in the publication 'Genetic and environmental determinants of immune response to cutaneous melanoma' Pozniak et al, in press, Cancer Research, January 2019.
- Histopathological measures of immune infiltrate (described in 3.4.6.2) in the LMC primary melanomas was derived from extensive work done by Dr Sally O'Shea who reviewed all the histological slides according to protocol
- Immunohistochemical staining of the LMC primary melanoma sections (described in 3.4.6.2) for VDR expression was optimised and performed by Dr Jonathan Laye (Senior Histopathologist in the group).
- Review and quantification of the extent and localisation of VDR expression the LMC primary melanoma sections (described in 3.4.6.2) was jointly performed by Dr Jonathan Laye, Sathya Muralidhar and Prof. Julia Newton-Bishop.

The work described in this chapter has been presented by me at the following scientific conferences/meetings:

- Selected abstract- Melanoma discovery and medicine session, NCRI annual meeting, Glasgow 2018
- Invited speaker- Ulster University, Coleraine, September 2017
- Selected abstract- GenoMel/MELGEN annual meetings held in Leeds (2016), Genoa (2017) and Essen (2018)
- Selected abstract- Vitamin D and analogs in Prevention and Therapy annual meeting, Homburg 2017

3.3 Methods

3.3.1 Variables used in this chapter

The following lifestyle, clinical and histopathological variables which have been used in this chapter are detailed below.

3.3.1.1 Season-adjusted serum vitamin D

For 549 of the 703 LMC participants, concentrations of cryopreserved serum 25-hydroxyvitamin vitamin D_2 and D_3 (nmol/L) were measured in 100 μL by Liquid Chromatography Tandem Mass Spectrometry (LC-MS/MS) by the NHS laboratory in the Leeds Teaching Hospitals Trust as described previously (288). Concentration of serum 25-hydroxyvitamin vitamin D_2 and D_3 for the remaining 154 samples was not available because: i) the initial batch of serum samples (from participants recruited early in the study) were stored in -20°C. Later in the study, it came to light that storage at -80°C was the most commonly reported storage temperature for sera from which 25-hydroxyvitamin vitamin D_2 and D_3 concentrations were measured. So, the samples saved at -20°C were not used for analyses and only samples stored at -80°C were deemed suitable for analysis ii) laboratory results of 25-hydroxyvitamin vitamin D_2 and D_3 concentrations were misplaced by the NHS laboratory in the Leeds Teaching Hospitals Trust, thus making these concentration values unavailable.

Of the available data, levels of D_2 were in the range of <10 nmol/L, which despite being detectable, is not quantifiable. Therefore, D_2 and D_3 levels were summed. Even though 1,25 dihydroxy vitamin D_3 is the metabolite which binds and signals via VDR-RXR, 25-hydroxyvitamin vitamin D_2 and D_3 (commonly referred to as 25(OH)D in publications) are the immediate precursors of 1,25 dihydroxy vitamin D_3 and are more stable in storage (369). Hence, they are widely used as proxy measures of 1,25

dihydroxy vitamin D_3 in the serum. Thus, these measured levels of 25-hydroxy vitamin D_2/D_3 were used for subsequent seasonal adjustment.

The blood samples (from which serum 25-hydroxy vitamin D_2/D_3 was measured) had been drawn at diagnosis, which was distributed across various seasons. Thus, the effect of the season had to be adjusted for, in order to make the samples comparable across the dataset without the confounding effect of seasonal variation in vitamin D synthesis. This was done using a linear regression model. Diagnoses made in January-March were considered as winter-diagnosed, April-June as spring-diagnosed, July-September as summer-diagnosed and October-December as autumn-diagnosed. A dummy variable was created for each season and Winter was set as baseline. After adjusting out the season effect on vitamin D, residuals were added to the intercept to obtain the corresponding winter vitamin D for each patient. **This value of season-adjusted levels of 25-hydroxy vitamin D_2/D_3 will be used for all analyses henceforth and will be referred to as 'serum vitamin D'.**

3.3.1.2 Tumour VDR expression

All references to '**tumour *VDR* expression**' refers to the log-normalised expression value of *VDR* expression in the LMC transcriptome. This value pertains to the probe designed to detect the *VDR* transcript (probe ID: ilmn_2319952). The 'Proportion detected' value (described in section 2.2.2) for this probe was 0.93. This meant that this probe was detected in 93% of samples with reliable intensity, above the background noise. In addition, this probe is designed to hit all splice isoforms of the *VDR* gene.

3.3.1.3 Tumour anatomical site

The anatomical site from which the primary tumour was excised from was recorded as part of clinical data. This was necessary because: i) the anatomical site on which the melanoma occurs is a prognostic factor (370) and ii) anatomical site of melanomas is also associated with sun-exposure patterns. One of the hypotheses that I tested in this chapter is the association of tumour *VDR* expression and sun-exposure, which I performed using both anatomical site and a direct measure of sun-exposure. **The anatomical sites were grouped into the following broader categories (Table 3.1) and referred to as 'tumour site' henceforth.**

Table 3.1: Distribution of 703 LMC participants based on the anatomical site from which their tumour was extracted

Tumour anatomical site group	Anatomical site	Number of tumours
Head/neck	Head/neck	81
Limbs	Arm unspecified	2
	Dorsal foot	2
	Elbow	5
	Knee	11
	Lower arm	34
	Lower leg	102
	Leg unspecified	1
	Thigh	74
	Upper arm	68
	Total on Limbs	299
Trunk	Abdomen	34
	Back	157
	Chest	34
	Shoulder	2
	Buttock	6
	Total on Trunk	233
Rare (sun protected sites)	ENT	5
	Acral	19
	Anal	5
	Cervix	1
	Foot	18
	Inside of hand	5
	Nodal with no known primary	3
	Penis	3
	Perineal	1
	Subungual	14
	Vaginal	3
	Vulval	13
	Total on Rare sites	90

3.3.1.4 Mitotic rate

The number of tumour cells in which mitoses can be detected per square millimetre of tumour area was counted and referred to as 'mitotic rate'. This was done by the histopathologist (Dr Sally O'Shea) using a light microscope. The 703 LMC melanomas were categorised into two categories, to enable certain analyses described in this chapter, for example: comparing *VDR* expression between tumours with 'low' and 'high' mitotic rate. To this effect, tumours with $\leq$ 1 mitoses (0 or 1 mitoses) per square millimetre of tumour area (n=346) were categorised as 'low' mitotic rate. Tumours with

> 1 mitoses per square millimetre of tumour area (n=356) were categorised as 'high' mitotic rate. **These grouped values will be referred to as 'mitotic rate' henceforth.**

3.3.1.5 Histopathological measure of tumour immune infiltrate

A categorical variable describing 3 categories of tumour-infiltrating lymphocytes within the whole tumour was categorised as follows:

- Absent (where there were either no TILs or TILs did not infiltrate the melanoma) (n=45)
- Non-brisk (where there were TILs but they either focally infiltrated the melanoma and/or sub totally infiltrated the invasive margin/base of the melanoma) (n=493)
- Brisk (where TILs infiltrated the entire base of the melanoma and may even have infiltrated the majority of the tumour) (n=64)

These grouped values of pathologist-graded measure of tumour immune infiltration in the whole tumour section will be referred to as 'histological TILs' henceforth

3.3.1.6 Self-reported supplement intake

Data had been collected on the regular use of vitamins, minerals, fish oils, fibre or food supplements in the period of 1 year prior to the interview. Details pertaining to dietary supplement intake are in Table 3.2.

Table 3.2: Distribution of 703 LMC participants based on self-reported dietary supplement data

Variable	Number of participants
Response if participant to '*A year ago, did you regularly take dietary supplements*?'	
Answered 'yes'	269
Answered 'no'	392
Total	661
If 'yes', what supplements were taken (self-reported)?	
Cod liver oil	103
Fish oil	22
Multivitamins/ vitamin C, E, B complex/ vitamin D	66
Other	78

3.3.1.7 Sun exposure

For a proportion of the 703 participants, the following sun exposure measures were recorded via the questionnaire.

- Average weekday exposure in warmer months: median= 1.75 hours per day
- Average weekday exposure in cooler months: median= 1.20 hours per day
- Average weekend exposure in warmer months: 4.30 hours per day
- Average weekend exposure in cooler months: 2.83 hours per day
- Average weekday exposure overall (combining warmer and cooler months): median=1.5 hours per day
- Average weekend exposure overall (combining warmer and cooler months): 3.6 hours per day

Details pertaining to sun exposure data are in Table 3.3

Table 3.3: Distribution of self-reported sun exposure patterns in the LMC dataset

Self-reported sun-exposure measure	Number of participants
Average weekday exposure in warmer months	349
≤ 1.75 hours per day	174
> 1.75 hours per day	175
Average weekday exposure in cooler months	345
≤ 1.2 hours per day	173
> 1.2 hours per day	172
Average weekday exposure overall	344
≤ 1.5 hours per day	172
> 1.5 hours per day	172
Average weekend exposure in warmer months	353
≤ 4.3 hours per day	176
> 4.3 hours per day	177
Average weekend exposure in cooler months	352
≤ 2.83 hours per day	176
> 2.83 hours per day	176
Average weekend exposure overall	351
≤ 3.6 hours per day	175
> 3.6 hours per day	176

The other variables used in this chapter, which have already been described in detail in 'Chapter 2: Methods' are: sex, age at diagnosis, AJCC stage, ulceration, Vascular invasion and Melanoma Specific Survival (MSS).

3.3.2 Univariable and multivariable regression analyses

Univariable linear regression was used to assess the correlations of serum vitamin D or *VDR* expression with clinicopathological variables. The clinicopathological variables which were found to correlate significantly, were used in a subsequent multivariable linear regression model to identify clinicopathological variables which correlated independently with serum vitamin D or *VDR* expression.

While the Regression coefficient (R) represents the strength of correlation, with R>0 denoting increased expression compared to baseline and R<0 denoting decreased expression compared to baseline. ^{B} indicates the group used as the baseline for comparison with other groups.

The Stata function 'regress' was used for univariable regression analysis.

3.3.3 Univariable and multivariable survival analysis

A univariable Cox Proportional Hazards model was used to estimate the effect of serum vitamin D or *VDR* expression on melanoma survival. The Stata function used for this analysis was 'stcox'. The time to failure was set as the time to death i.e number of years from date of diagnosis to date of death (melanoma-specific death). The failure event in this case was melanoma-specific death.

3.3.4 Creating 3 VDR-groups using X-tile

VDR expression and MSS were used as input for X-tile (264), which used this information to categorise the 703 tumours into 3 groups with most divergent melanoma specific survival. To avoid overfitting, this approach was trained in randomly selected 2/3 of the samples and validated in the remaining 1/3. In both sets, tumours with the lowest 17%, middle 66% and highest 17% of *VDR* expression were identified as low, intermediate and high-VDR groups respectively. These percentiles were applied to the TCGA melanoma dataset to identify low, intermediate and high-VDR groups. The statistical significance of the difference in prognosis between the 3 groups was tested using Cox Proportional Hazards model in the Leeds data and TCGA melanoma data.

3.3.5 Whole-transcriptome correlations

A linear regression analysis (STATA command 'regress') was used to test the correlation of each gene with *VDR* expression, as a result of which each gene was assigned a regression coefficient and p-value (P) to measure the strength of the correlation. The linear regression model is used to describe the relationship between a predictor variable (say 'x', expression value of *VDR*) and a response variable (say 'y', expression value of a gene). In this case, the equation to describe the relationship would

be: $y=\beta x+c$, where β is the regression coefficient and c is the intercept value. The null hypothesis (H_0) in this model is that there is no association between the predictor variable and response variable i.e $\beta=0$. The p-value from this model is the probability that the null hypothesis is rejected i.e $\beta\neq0$. Since a p-value was generated to test the correlation of the expression of each gene with *VDR*, it was necessary to perform multiple correction, in order to eliminate identification of 'true positive' correlates by chance. To this effect, multiple testing (Benjamini-Hochberg False Discovery Rate-FDR) was used to identify the most significant correlates, thus minimizing selection of genes correlated by chance. Additionally, only genes with $|R|>0.2$ were selected to exclude genes with 'weak' correlations. This pipeline was used for whole-genome correlation in both the LMC and TCGA data. Genes which passed the aforementioned significance threshold were plotted in a volcano plot ('plot' function in R) and used for functional enrichment analysis. In the volcano plots, the regression coefficient (x-axis) versus p-values (y-axis) were plotted, in order to graphically represent significantly correlated genes (as indicated by p-value) along with strength of correlation (as indicated by regression coefficient) with the predictor variable (VDR expression).

FLG-adjusted correlations with VDR: A sensitivity analysis was conducted adjusting the expression of *FLG2* to account for any bias in *VDR* expression which might have originated from keratinocytes. The rationale for this being: *FLG2* is the gene marker of differentiated keratinocytes (371) and hence *FLG2* expression was considered to be a proxy for a melanoma with 'high epidermal content'.

3.3.6 Gene and pathways enrichment analyses

Functional enrichment of significant *VDR*-correlated genes was carried out using Reactome FIViz (321) in Cytoscape. The 'Gene set/mutation analysis' feature was used to produce a Functional Interaction Network and perform pathway enrichment of a given input gene list. Significantly enriched pathways were identified as those with a FDR <0.05 (hypergeometric test computed by Reactome FIViz).

3.3.7 Replicating TCGA and Lund molecular phenotypes in the LMC melanoma transcriptome

A classification of 208 LMC primary melanomas (subset of the 703 used in this thesis) as high immune, normal-like, proliferative or pigmentation subtypes (defined by Jonsson et al. 2010 and referred to as Lund molecular phenotypes) has been published by our group (372). The same classification approach (nearest centroid) was used to assign each of the 703 tumours to one of these subtypes and similarly to one of the 3

subtypes (Immune, Keratin and *MITF*.low) defined by the TCGA signature (TCGA, 2015).

3.3.8 Imputed immune cell scores

As described by Pozniak et al (in press, 2019, Cancer Research). Briefly, genes that were expressed among the top 25% across the whole genome in a melanocyte and melanoma cell lines were removed from the initial list of the genes provided by Angelova et al (373). A particular cell type was eliminated if less than 10% of its original genes remained after this filtering process. For each immune cell type, the negatively correlating genes were removed. This produced 26 immune cell types, for which the immune cell scores were calculated as the mean of expression values of all genes attributed to specific cell types after z-score normalization.

3.3.9 VDR genomic binding regions

3.3.9.1 ChIP-Seq analysis

The genomic regions identified as having *VDR*-binding peaks across 6 tissue types (accessible data from (374)) were downloaded as a BED file and analysed in GREAT 3.0.0 (375). 'Basal plus extension' approach was used with Human GRCh37 species assembly, whole genome as background regions and the gene regulatory domain set to $\pm$20 kb upstream and $\pm$400 kb distal. The genes mapped to these regions ('region-gene associations') were exported and overlap with *VDR*-correlated genes in the LMC (at FDR<0.05) were assessed for each tissue-type. The process of estimating 'region-gene associations' is as follows: first, every gene is assigned a regulatory domain and then each genomic region is associated with all genes whose regulatory domain it overlaps.

3.3.9.2 VDR-binding motifs

The database MotifMap was used, from which data pertaining to the transcription factor binding sites (based on hg19 build) was downloaded. From this, the following data specific to VDR binding motifs was extracted: start site, end site and chromosomal location. These regions (containing the VDR0binding motif) were mapped to genes using GREAT. Similar to the process described above, 'Basal plus extension' approach was used with Human GRCh37 species assembly, whole genome as background regions and the gene regulatory domain set to $\pm$20 kb upstream and $\pm$400 kb distal. The genes mapped to these regions ('region-gene associations') were exported and overlap with *VDR*-correlated genes in the LMC (at FDR<0.05) were assessed.

3.3.10 Immunohistochemistry of VDR in the LMC tumours

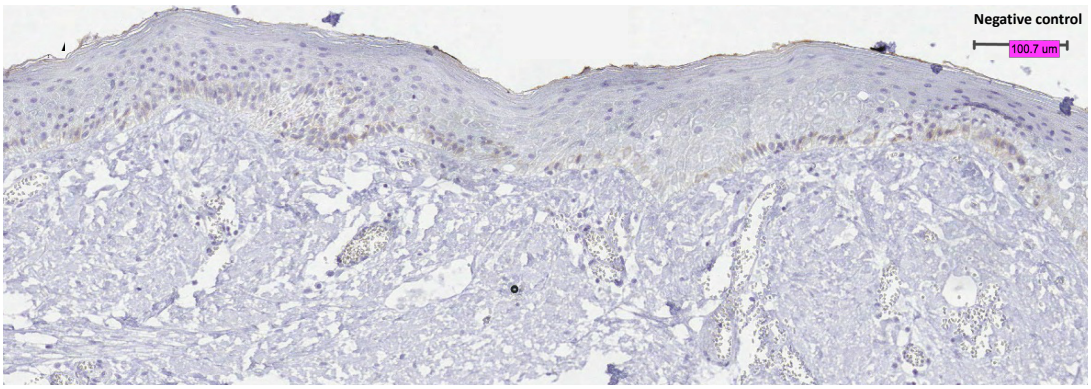
3.3.10.1 Optimisation of the anti-VDR antibody

The anti-VDR antibody (D-6 Sc-13133, Santa Cruz) was optimised on FFPE-derived sections of normal (no melanoma) human skin, in order to identify the optimal experimental conditions which would enable quantification of VDR expression in the melanoma tissue sections from the LMC participants. The following protocol was used for the optimisation process:

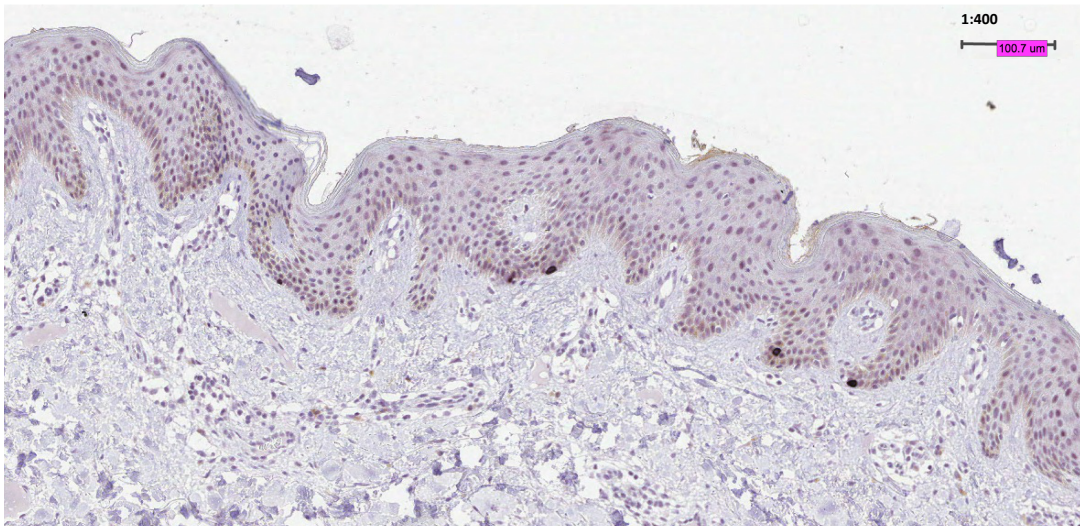
1) The FFPE blocks were sectioned using a microtome to produce 5μm sections
2) The tissue sections were deparaffinised on a hotplate at 70°C followed by antigen retrieval in a pressure cooker, with the slides submerged in 1x antigen retrieval solution (Menapath technologies). After antigen retrieval, the slides were washed in wash buffer (1x, Menapath technologies) followed by rinsing in running tap water.
3) The slides were then marked around the tissue region of interest, with a wax marker. Slides were kept in a humidity chamber after this step to ensure tissue hydration.
4) The tissue was then treated with 100μl peroxide blocking solution (Menapath technologies) for 11 minutes. The solution was then washed away with wash buffer (1x, Menapath technologies)
5) The tissue was then treated with 100μl Casein blocking solution (Menapath technologies) for 10 minutes
6) The tissues were then treated with the desired primary antibody concentrations, diluted in antibody-diluent solution (Menapath technologies). In this case, 5 concentrations were performed on 5 consecutively sectioned human skin sections: 1:50, 1:100, 1:200, 1:330, 1:400 and negative control (no antibody, only the diluent solution). The primary antibody treatment was allowed to incubate at 4°C overnight.
7) This was followed by secondary antibody treatment using ImmPRESS HRP reagents (MP-7452, Vector Laboratories) and visualised using purple Vector VIP substrate (SK-4600, Vector Laboratories).
8) The sections were left to dry overnight, after which they were reviewed.
9) Figure 3.2 includes representative images of the human skin sections stained with VDR as per the above protocol. Of the 5 concentrations of antibody that were used, the use of 1:200 concentration appeared to have the most lucid staining as evaluated by the following criteria: i) positive staining for VDR in the keratinocytes, which is distinguishable from the remaining epidermal tissue. This is because

VDR is known to be expressed in epidermal keratinocytes ii) negative staining in the dermal tissue. While the higher concentrations (1:50 and 1:100) stained positive for VDR in the epidermal keratinocytes, the staining was not completely negative in the dermal tissue. On the other hand, the lower concentration (1:300 and 1:400) produced staining that was too weak to be able to identify distinct keratinocyte positivity. However, the concentration of **1:200** satisfied both criteria and was thus chosen to be used for staining the melanoma sections from the LMC.

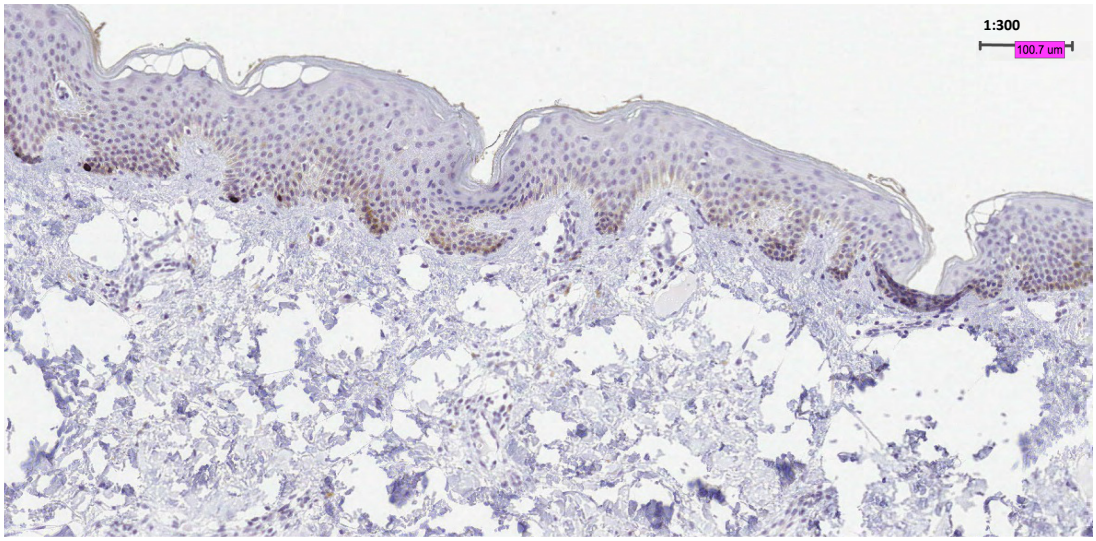
A



B



C



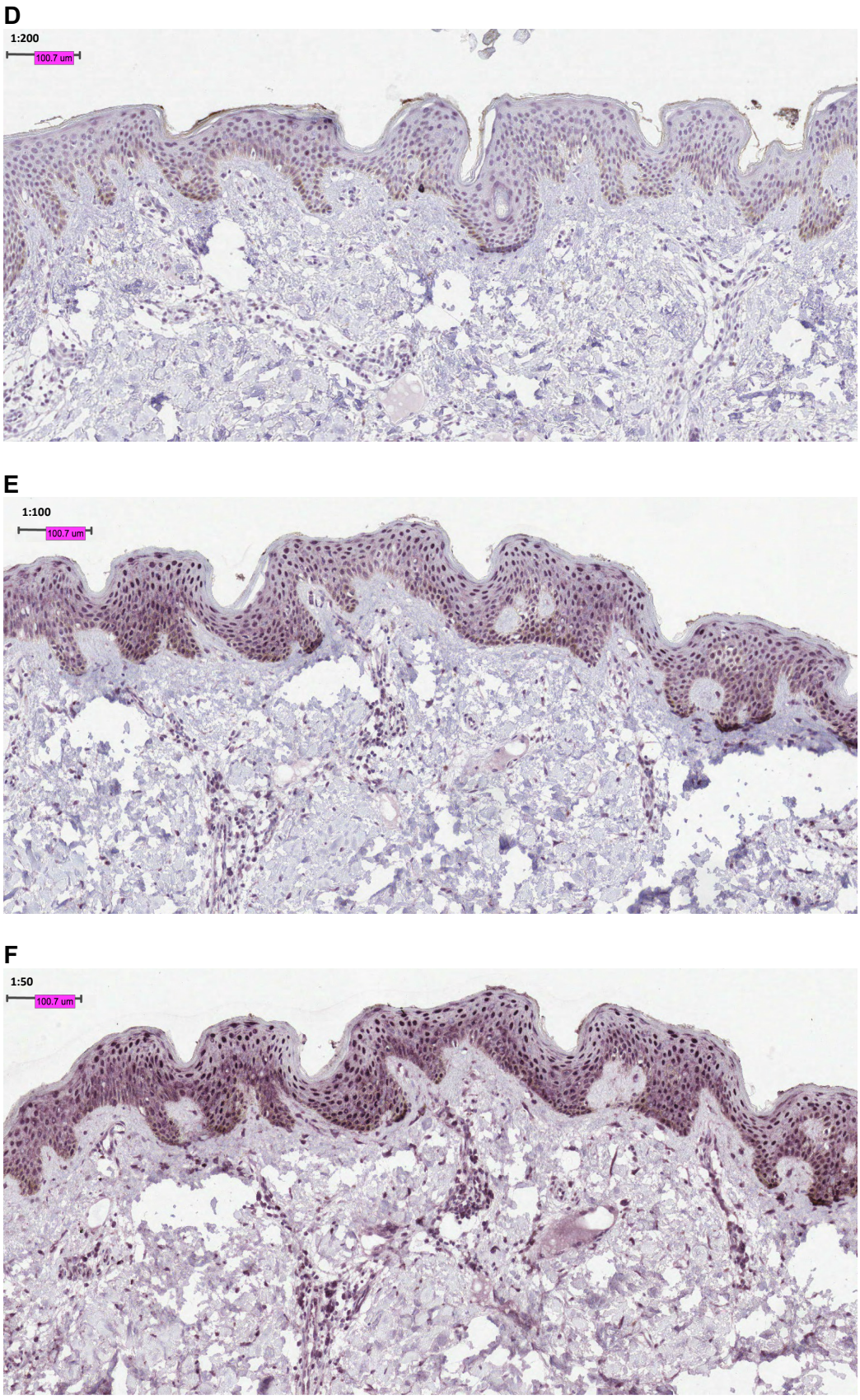


Figure 3.2: Optimisation of immunohistochemical staining of VDR using sections of human skin.

The concentrations tested are indicated inset: negative control (A); 1:400 (B); 1:300 (C); 1:200 (D); 1:100 (E); and 1:50 (F).

3.3.10.2 Immunohistochemical staining and evaluation of tissue sections from LMC melanomas

A subset of the LMC tumours (n=30) were sectioned and stained for VDR expression using the anti-VDR antibody (D-6 Sc-13133, Santa Cruz). The reason for the availability of a limited number of tumour sections for IHC staining was because only tumour blocks from deceased participants could be used. The protocol detailed above was used for this purpose and the chosen antibody concentration was 1:200. The stained sections were then reviewed by myself and Dr. Jon Laye (senior histopathologist in the Leeds group). We reviewed each stained tumour section as follows:

- Positive staining of keratinocytes was evaluated, as an 'internal control' to ensure each section was adequately stained. To this effect, all 30 sections showed keratinocyte VDR positivity.
- A focal region was defined to be that around the 'tumour core': the part of the tumour from which the TMA needle had been used to extract cells from which the transcriptomic profiles were generated. In this focal region around the tumour core (5-10 cell width) the following measures were evaluated:
    i) number of tumour cells which stained positive for VDR in the cytoplasm,
    ii) number of tumour cells which stained positive for VDR in the nucleus, and
    iii) presence/absence of TILs which stained positive for VDR in the nucleus.

Representative images from this experiment are presented and discussed in section 3.4.5.2.

3.4 **Results**

3.4.1 Serum vitamin D and *VDR* in the LMC

This section will begin with a description of the distribution of serum vitamin D and tumour *VDR* expression in the LMC dataset. This is followed by the assessment of individual correlations of serum vitamin D level and tumour *VDR* expression with

i) each other i.e. the correlation between *VDR* and serum vitamin D levels
ii) clinicopathological features that characterise the progression of primary cutaneous melanoma
iii) measures of self-reported sun-exposure in participants
iv) expression of genes that code for key components of the vitamin D-VDR endocrine system such as the CYP metabolic enzymes and *RXR* and
v) expression of genes that code for the other NR1L family of nuclear receptors.

The clinicopathological factors tested include AJCC stage, mitotic rate, age at diagnosis, sex and anatomic site of melanoma. Age at diagnosis and sex were tested because melanoma diagnosis has been reported to vary with sex and age at diagnosis (376), both of which are risk factors associated with melanoma survival. Anatomic site of melanoma at diagnosis has also been reported as primary prognostic factor for primary cutaneous melanoma (377), and hence was tested.

Though the LMC transcriptomic data set is derived from 703 primary melanomas, this section (3.4.1) is based on analysis carried out on data from 700 primary melanoma tumours, after excluding 3 samples from participants with vitamin D levels greater than 115nmol/L (considered outliers). The significance of these participants is discussed in section 3.4.4.

3.4.1.1 Distribution of serum vitamin D levels and VDR expression in the Leeds Melanoma Cohort

In the 700 LMC primary melanomas, *VDR* expression was normally distributed, with a log-normalised mean expression value of 7.7, minimum of 2.9 and maximum of 9.5 (Figure 3.3A). Figure 3.3B represents the distribution of serum vitamin D levels of all 549 participants, highlighting the 3 participants with serum vitamin D levels >115nmol/L and the remaining 546 participants with serum vitamin D levels <115nmol/L. In the 546 participants, the mean serum vitamin D level was 41.95 nmol/L, the minimum value was -2.8 nmol/L and maximum was 105 nmol/L. The negative minimum value of serum vitamin D is a consequence of the 'season-adjustment' process, to account for the 20nmol/L average variation of vitamin D levels between summer and winter in the UK. This season-adjustment is explained in detail in the methods section.

3.4.1.2 Correlation of serum vitamin D levels with tumour VDR expression

The vitamin D-VDR signalling axis exerts transcriptional control over target genes, which include genes whose protein products control the metabolism of vitamin D, thus maintaining equilibrium of this signalling pathway. Owing to this negative feedback loop, the correlation between serum vitamin D (vitamin D in circulation) and tumour VDR expression was assessed.

Serum vitamin D did not correlate significantly with tumour *VDR* expression (R=0.36, P=0.74) in the subset of participants for whom serum vitamin D levels was available (n= 549; Figure 3.3C). The correlation between serum vitamin D and VDR expression was also not significant in the 546 participants after exclusion of the 3 samples from participants with outlier serum vitamin D levels >115nmol/L (R=0.34, P=0.73; Figure 3.3D).

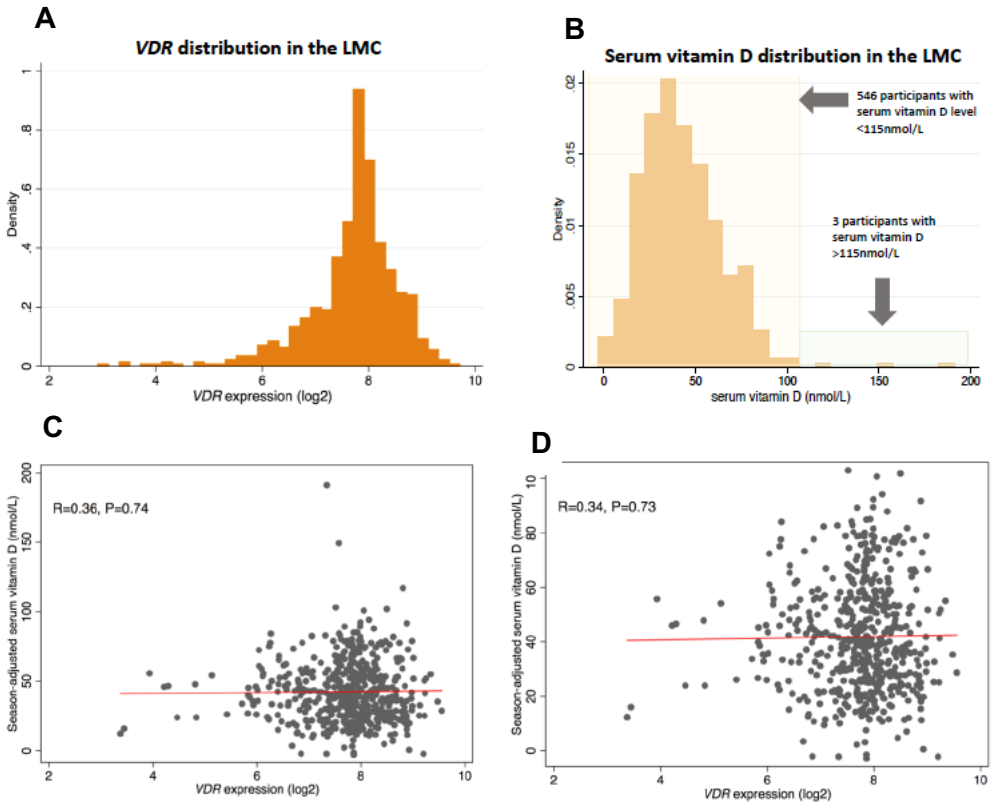


Figure 3.3: Distribution and correlation of serum vitamin D and VDR expression in the LMC

A: Distribution of tumour VDR expression in the 703 LMC primary melanomas;

B: Distribution of serum vitamin D levels in the 549 LMC primary melanomas, highlighting the 546 participants with serum vitamin D levels<115nmol/L (yellow box) and 3 participants with serum vitamin D>115nmol/L (green box);

C: Correlation between serum vitamin D and tumour VDR expression in 549 LMC primary melanomas (including 3 participants with serum vitamin D>115nmol/L);

D: Correlation between serum vitamin D and tumour VDR expression in 546 LMC primary melanomas (excluding the 3 participants with outlier serum vitamin D levels >115nmol/L)

3.4.1.3 Clinicopathological Correlates of serum vitamin D

In the 700 LMC primary melanomas, univariable regression analyses revealed that serum vitamin D did not correlate significantly with sex (P=0.77), AJCC stage (P>0.07), mitotic rate (P=0.26), tumour site (P>0.26) or ulceration (P=0.92) (Table 3.4).

Table 3.4: Association of season-adjusted serum vitamin D with clinicopathological variables in 703 LMC participants

[B] indicates the group used as the baseline for comparison with other groups.

	Univariable		
Season-adjusted serum vitamin D association with	**Reg Coef**	**Std. Error**	**P-val**
Age (years)	0.14	0.07	0.04
Sex			
Females[B]			
Males	0.55	1.89	0.77
AJCC Stage			
Stage I[B]			
Stage II	-3.05	2.04	0.136
Stage III	-5.28	2.96	0.075
Mitotic rate			
<1 mitoses/mm^2 tumour[B]			
>=1mitoses/mm^2 tumour	2.10	1.86	0.26
Tumour site			
Head[B]			
Limbs	-3.44	3.05	0.26
Trunk	-1.52	3.16	0.63
Rare (sun-protected sites)	-3.65	3.91	0.35
Ulceration			
Ulcerated[B]			
Non-ulcerated	-0.00009	0.0009	0.925

Members of the Cytochrome P450 (CYP) enzyme superfamily play a role in maintaining vitamin D homeostasis. The primary CYPs associated with Vitamin D anabolism (synthesis) are CYP27A1 and CYP27B1 in the liver and kidney respectively. The catabolism (break-down) of vitamin D is mediated by CYP24A1 in the kidney. To investigate if serum vitamin D levels are associated with the expression of CYP genes in the LMC primary melanomas, a correlation analysis was done. To this effect, serum vitamin D did not correlate significantly with CYP27A1 (R=-0.72, P=0.57), CYP27B1 (R=-0.85, P=0.28) or CYP24A1 (R=-0.53, P=0.46) (Figure 3.4).

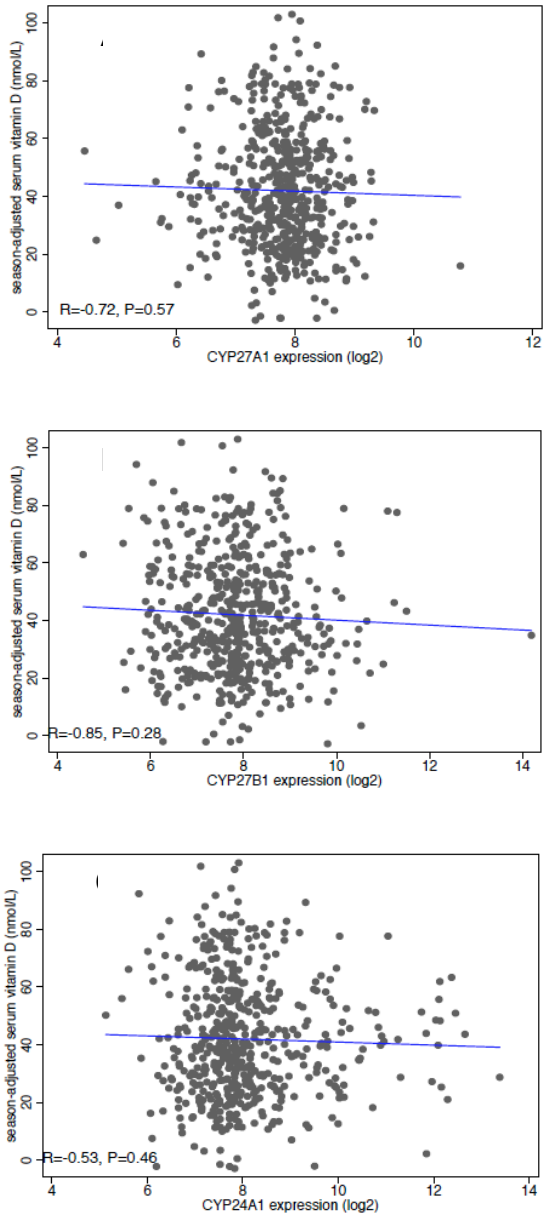


Figure 3.4: Correlation of serum vitamin D with expression of Cytochrome P450 (CYP) genes involved in vitamin D metabolism

Correlation of serum vitamin D with *CYP27A1* (A), *CYP27B1* (B) and *CYP24A1* (C) expression in the 546 LMC primary melanomas.

3.4.1.4 Clinicopathological correlates of tumour VDR expression

Univariable regression analyses identified the clinicopathological features that significantly correlated with *VDR* expression. The significant correlates were then used in a multivariable regression model to identify the independent correlates of *VDR* expression.

The univariable regression analysis revealed the following observations (summarised in Table 3.5). ^{B} indicates the group used as the baseline for comparison with other groups.

- *VDR* expression was marginally significantly correlated with age at diagnosis (R=-0.005, P=0.04)
- *VDR* expression was significantly higher in females^{B} compared to males (P=0.003).
- *VDR* expression was significantly higher in AJCC Stage I^{B} tumours compared to AJCC Stage II (R=-0.17, P=0.012) and AJCC Stage III (R=-0.25, P=0.009) tumours.
- *VDR* expression was significantly higher in tumours with <=1 mitosis/sq.mm^{B} than those with >1 mitoses/sq.mm (R=-0.23, P=0.0004).
- VDR expression did not vary significantly between tumours on the head/neck^{B} and limbs (R=-0.04, P=0.63). However, tumour *VDR* expression was significantly lower in truncal tumours (R=-0.35, P=0.001) and Rare tumours (R=-0.44, P=0.001) compared to head/neck tumours. Rare tumours are those on sun-protected sites.

Taken together, the univariable analyses identified the following clinicopathological features to be significantly associated with *VDR*: AJCC stage, mitotic rate, sex and tumour site. However, it is important to assess which of these characteristics were independently associated with *VDR* expression. Multivariable analysis estimates the significance of the association of one variable with VDR expression, after adjusting for all other variables.

In the multivariable regression analysis, age at diagnosis (P=0.14) and sex (P=0.14) were not significantly associated with *VDR* expression, suggesting that age and sex were not independently associated with *VDR* expression. In other words, age and sex were confounded by AJCC stage and tumour site. On the other hand, AJCC stage, mitotic number and tumour site remained significantly associated with *VDR* expression in the multivariable analysis. *VDR* expression was significantly higher in stage I tumours compared to stage II (P=0.08) and stage III tumours (P=0.06), after adjusting for age, sex and tumour site. *VDR* expression was also independently

inversely correlated with mitotic number (P=0.001). Similarly, *VDR* expression was significantly higher in head tumours compared to truncal (P=0.001) and rare tumours (P=0.003), after adjusting for age, sex and tumour site. Taken together, these results indicated that AJCC stage, mitotic number and tumour site were independently associated with tumour *VDR* expression whereas age and sex were not.

Table 3.5: Association of tumour VDR expression with clinicopathological variables in 703 LMC tumours

[B] indicates the group used as the baseline for comparison with other groups.

	Univariable			Multivariable		
***VDR* association with**	**Reg Coef**	**Std. Error**	**P-val**	**Reg Coef**	**Std. Error**	**P-val**
Age (years)	-0.005	0.002	0.04	-0.003	0.002	0.14
Sex						
Females[B]						
Males	-0.18	0.06	0.003	-0.09	0.06	0.16
AJCC Stage						
Stage I[B]						
Stage II	-0.17	0.07	0.012	-0.12	0.07	0.08
Stage III	-0.25	0.09	0.009	-0.18	0.09	0.06
Mitotic rate						
<1 mitoses/mm^2 tumour[B]						
>=1mitoses/mm^2 tumour	-0.23	0.06	0.0004	-0.20	0.06	0.001
Tumour site						
Head[B]						
Limbs	-0.04	0.10	0.63	-0.11	0.10	0.28
Trunk	-0.35	0.10	0.001	-0.36	0.10	0.001
Rare (sun protected sites)	-0.44	0.12	0.001	-0.38	0.13	0.003

Additionally, *VDR* has been shown to be higher in primary melanomas with increased immune infiltrate (368), which by itself is a predictor of improved melanoma prognosis (171, 172). Hence, the association of *VDR* expression with pathologist-graded measure of tumour immune infiltration was assessed. *VDR* expression was significantly lower in tumours with no immune infiltrate ('absent') compared to those with non-brisk (P=0.02) and brisk (P=0.006) immune infiltrate (Figure 3.5).

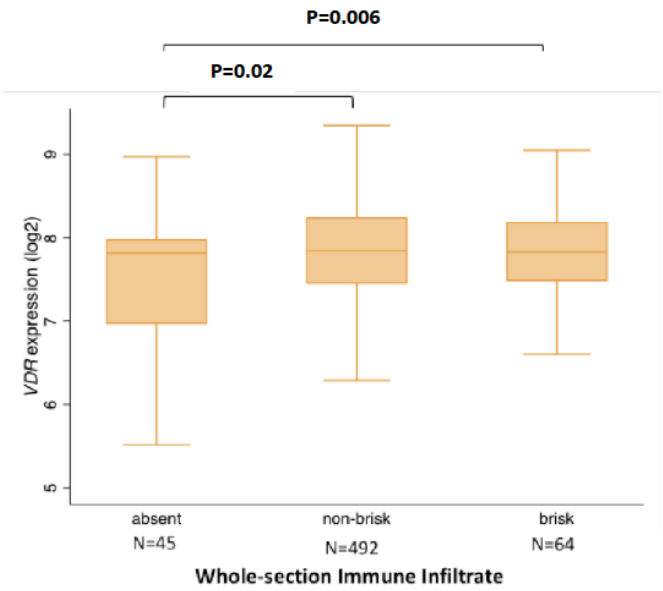


Figure 3.5: Association of tumour VDR expression with pathologist-graded measure of tumour immune infiltration in the whole tumour section

VDR expression was compared between tumours with 'absent' (n=45), 'non-brisk' (n=492) or 'brisk' (n=64) immune infiltrate. P-values from linear regression model.

3.4.1.5 Tumour VDR expression and self-reported sun exposure

Since tumour *VDR* expression varied between tumours diagnosed in different anatomical sites, being significantly lower in sun-protected sites compared to those on the head, it was hypothesised that *VDR* expression in the tumour would be associated with the participants' sun-exposure pattern. The association of tumour *VDR* expression with each of these sun-exposure measures was estimated using a linear regression model.

To this effect, tumour *VDR* expression was significantly associated with only one of the sun exposure measures: average measure of sun exposure in the weekend in both warmer and cooler months (measured in hours per week, P=0.04). This measure is a combined measure of two other measures: average measure of sun exposure in the weekend in warmer months and average measure of sun exposure in the weekend in cooler months, both of which also have borderline significant associations with *VDR* expression (P=0.06) (Table 3.6).

Table 3.6: Association of sun-exposure measures with tumour VDR expression

Self-reported sun-exposure measure	Regression Coefficient	P-value
Average weekday exposure in warmer months	0.04	0.17
Average weekday exposure in cooler months	0.03	0.32
Average weekday exposure overall	0.04	0.23
Average weekend exposure in warmer months	0.06	0.06
Average weekend exposure in cooler months	0.06	0.06
Average weekend exposure overall	0.07	0.04

3.4.1.6 Correlation of tumour VDR expression with expression of NR1L superfamily genes

DR belongs to the NR1L superfamily of nuclear receptors which consists of other nuclear hormone receptors such as LXRB, FXR1, RXR, FXR2 and PXR. Of these NR1L superfamily members, RXR is of particular interest since vitamin D-activated VDR forms a dimeric complex with RXR, which together facilitate transcription of target genes. Thus, the correlation of *VDR* with the expression of three RXR isoforms as well as the other NR1L family members was checked. *VDR* correlated significantly with *RXRB, LXRB, PXR, FXR1* and *FXR2* (Table 3.7)

Table 3.7: VDR correlation with expression of NR1L superfamily members

Correlation of *VDR* with	Regression Coefficient	P-value
RXRA	0.01	0.73
RXRB	-0.4	2.9×10^{-6}
RXRG	0.002	0.90
LXRB	0.27	7×10^{-14}
PXR	-0.13	0.0004
FXR1	0.13	0.0003
FXR2	-0.12	0.001
CAR1	0.01	0.75

3.4.2 Vitamin D-VDR signalling and melanoma prognosis

This section will focus on assessing the individual effects of serum vitamin D and tumour *VDR* expression on melanoma prognosis in the LMC primary melanomas. To validate the findings from the LMC, the prognostic significance of tumour *VDR* expression was

also evaluated in the metastatic melanomas from the TCGA melanoma cohort. It was not possible to assess the prognostic effect of serum vitamin D in the TCGA cohort, owing to absence of data for serum vitamin D in this cohort of patients.

3.4.2.1 Serum vitamin D levels and melanoma prognosis in the LMC

High serum vitamin D levels at diagnosis have been associated with improved prognosis in patients from 6 independent melanoma cohorts, including the larger 'original' Leeds Melanoma Cohort of 2184 participants. Hence, the effect of higher serum vitamin D in the 546 LMC primary melanomas was assessed. There was no significant association of serum vitamin D with prognosis (HR=0.99, P=0.06). This was analysis was done using serum vitamin D on a continuous scale, indicating that for every 1nmol/L increase in serum vitamin D, the estimated risk of death fell by 1%. However, various thresholds of vitamin D levels to indicate deficiency and sufficiency have been proposed, with different cut-off limits in different human populations. The clinical definition of deficient winter vitamin D set by the UK Government Scientific Advisory Committee on Nutrition (378) was 25 nmol/L. So, the LMC participants were categorised based on their vitamin D levels into 'low' serum vitamin D (<25nmol/L, n=116) and 'high' serum vitamin D ($\geq$25nmol/L, n=430) groups and will be referred to as such henceforth. Interestingly, this threshold of 25nmol/L was very close to the optimum cut-off identified as best predicting survival in our data set (22nmol/L) using X-tile (see section 3.3). In comparing the prognosis of participants with categorised levels of vitamin D, there was borderline significant variation in melanoma prognosis between the participants with high and low serum vitamin D at diagnosis (HR=0.72, P=0.07).

3.4.2.2 Tumour VDR expression and melanoma prognosis

VDR loss (as measured by IHC) has been reported to be associated with cutaneous melanoma progression (368) and *VDR* polymorphisms have been shown to be associated with melanoma survival (379). So, the prognostic significance of tumour *VDR* expression was assessed in the LMC primary melanomas. Higher *VDR* expression was significantly correlated with clinicopathological features such as lower AJCC stage, lower mitotic rate, increased immune infiltrate and tumours of the head/neck (compared to truncal and sun-protected tumours) (section 3.4.1.4), which by themselves are significant predictors of melanoma outcomes. In order to ensure that the effect of *VDR* expression on melanoma survival was not confounded by the aforementioned clinicopathological features, a multivariable analysis was performed to assess if tumour *VDR* expression was an independent predictor of melanoma survival.

Tumour *VDR* expression was found to be protective for melanoma death (univariable model, HR=0.75, P=0.0001). A multivariable model revealed that tumour

VDR expression was protective for melanoma death independent of AJCC stage, mitotic rate, tumour site and tumour immune infiltrate (HR=0.8, P=0.008) (Table 3.8). In other words, the protective effect of tumour *VDR* (on melanoma death) persisted even after adjusting for clinicopathological features, which are also significantly correlated with *VDR*.

Table 3.8: Association of VDR expression with death from melanoma in the LMC

***VDR*: association with death from melanoma**	**Hazard Ratio**	**Std. Error**	**P-value**
Univariable	0.75	0.05	0.0001
Adjusted for AJCC stage, mitotic rate, tumour site and tumour immune infiltrate	0.80	0.06	0.008

VDR correlated significantly with expression of some members of the NR1L superfamily of nuclear receptors (see section 3.4.1.4). This put forth the possibility that the prognostic significance of tumour *VDR* expression could be confounded by the expression of these genes. This was tested using a multivariable survival analysis, in which the prognostic significance of *VDR* was estimated after adjusting for each of the significantly correlated NR1L family genes. Tumour *VDR* expression was significantly protective for melanoma death after adjusting for *RXRB, LXR, PXR, FXR1* or *FXR2* (Table 3.9). This indicated that the association of *VDR* with improved melanoma prognosis was independent of the expression of other NR1L superfamily members.

Table 3.9: Association of VDR expression with melanoma death after adjusting for expression of NR1L family genes

***VDR*: association with death from melanoma**	**Hazard Ratio**	**Std. Error**	**P-value**
Univariable	0.75	0.05	0.0001
Adjusted for *RXRB* expression	0.76	0.05	0.00006
Adjusted for *LXR* expression	0.76	0.05	0.0002
Adjusted for *PXR* expression	0.74	0.05	0.00003
Adjusted for *FXR1* expression	0.76	0.05	0.00003
Adjusted for *FXR2* expression	0.77	0.05	0.00009

Collectively, in the 700 LMC primary melanomas, *VDR* expression but not serum vitamin D level was significantly associated with improved melanoma survival. In order to assess if serum vitamin D offered a prognostic benefit within the context of *VDR* expression, the 700 tumours were stratified into 3 *VDR*-groups. To this effect, a survival-

based stratification using X-tile (see Methods 3.3) was performed, which identified three tumour groups: low-VDR (17% of tumours with lowest *VDR* expression, n=119), high-VDR (17% of tumours with highest *VDR* expression, n=119) and intermediate-VDR (middle 66% of tumours, n=462) (Figure 3.6A). Compared to the low-VDR group (which had the worst prognosis and hence used as baseline) the intermediate-VDR group (HR=0.51, P=0.0003) and the high-VDR group (HR=0.25, P=5.4 x 10^{-8}) had significantly improved prognosis (Figure 3.6B). These *VDR*-groups were used subsequently in a 'vitamin D-*VDR* subgroup analysis', to explore the association of serum vitamin D within the context of tumour *VDR* expression: discussed in section 3.4.3.

Having identified 3 *VDR*-groups in the LMC primary melanomas, it remained to be tested if this *VDR*-associated improvement in survival was also relevant in metastatic melanomas. To this effect, the TCGA metastatic melanoma data set was used. When the same *VDR* expression cut offs (lowest 17%, middle 66% and highest 17%) were applied to the TCGA metastatic melanoma data, the replication produced three *VDR*-based tumour groups: low-VDR (n=71), intermediate-VDR (n=270) and high-VDR (n=69) (Figure 3.6C) with a progressive improvement in melanoma survival with increasing *VDR* expression (overall HR=0.774, P=0.05) (Figure 3.6D).

Taken together, increased tumour *VDR* expression was associated with improved prognosis in both primary and metastatic melanomas, as evidenced in the LMC and TCGA datasets respectively.

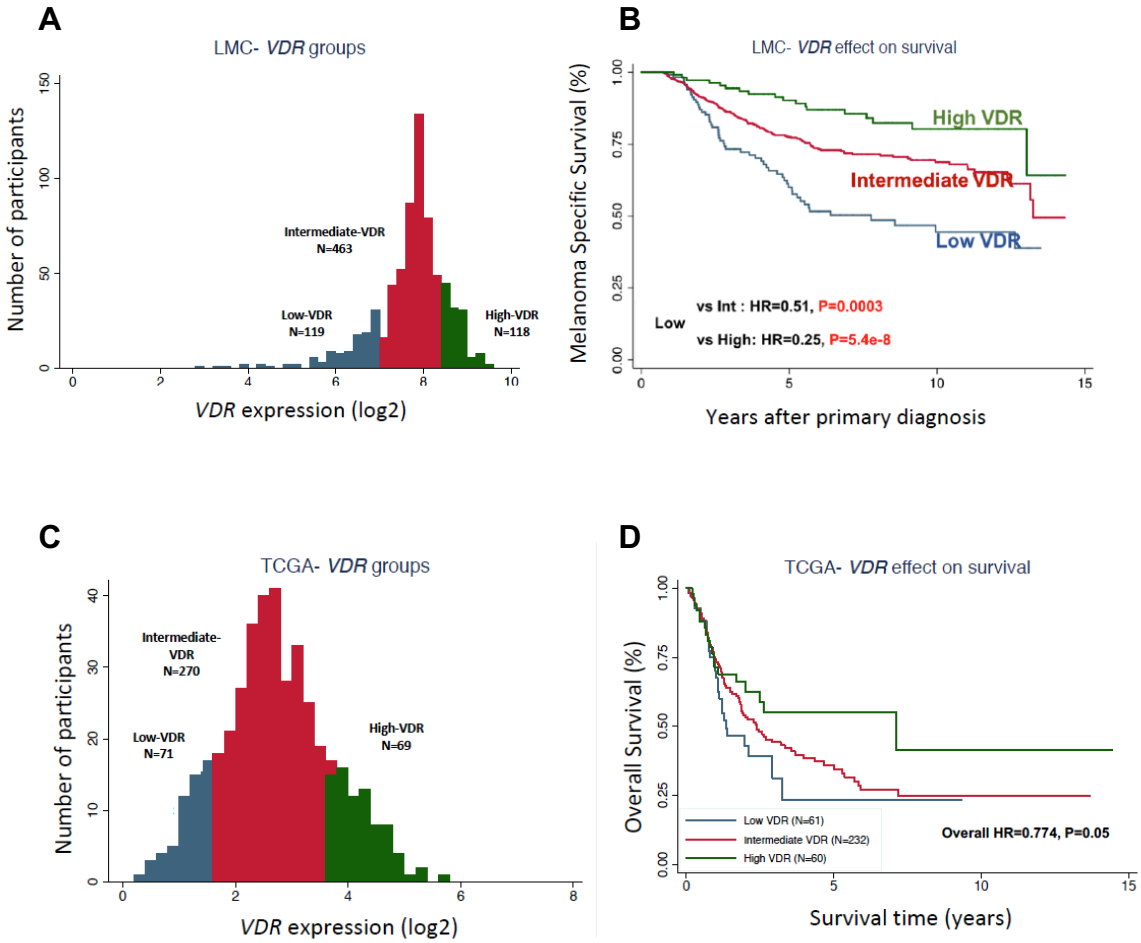


Figure 3.6: Three VDR-groups in the LMC primary melanomas

A: Stratification of the 700 LMC primary melanomas into low-VDR (n=119), intermediate-VDR (n=463) and high-VDR (n=119) groups using X-tile;

B: Difference in survival of the 3 VDR-groups in the LMC dataset. Cox Proportional Hazards model was used to identify HR (Hazard Ratio) of intermediate- and high-VDR groups, relative to the low-VDR group;

C: Stratification of the TCGA metastatic melanomas into low-VDR (n=71), intermediate-VDR (n=270) and high-VDR (n=69) upon applying the proportions derived from the LMC analyses (lowest 17%, middle 66% and highest17% of *VDR* expression);

D: Difference in survival of the 3 VDR-groups in the TCGA dataset. Cox Proportional Hazards model was used to identify HR (Hazard Ratio) and P-value.

3.4.3 Context-specific protective effect of serum vitamin D on melanoma survival

This section will focus on the vitamin D- *VDR* subgroup analysis, wherein each of the three *VDR* -groups (described in 3.4.2.2) were further subdivided based on the serum vitamin D level (low or high serum vitamin D) at diagnosis of the participants. This produced 6 subgroups: high-VDR tumours with low (n=20) or high (n=73) vitamin D, intermediate-VDR tumours with low (n=73) or high (n=288) vitamin D and low-VDR tumours with low (n=20) or high (n=73) vitamin D. The melanoma prognosis of participants with low or high serum vitamin D within each of the three *VDR* groups was compared. In other words, the prognostic significance of serum vitamin D within the context of tumour *VDR* expression was assessed. Serum vitamin D was not significantly associated with melanoma prognosis in the low-VDR (P=0.66) and high-VDR (P=0.57) groups. However, in the intermediate-VDR group, participant with low serum vitamin D levels had a significantly worse prognosis compared to those with high serum vitamin D (HR=1.73, P=0.02) (Figure 3.7). Based on this observation, it was postulated that the intermediate-VDR group would be an optimal subset of primary melanomas to identify factors associated with the protective of vitamin D on melanoma survival. To this effect, the two intermediate-VDR subgroups (with low/high serum vitamin D) were used for subsequent comparative analyses. Differences in clinicopathological features as well as transcriptomic differences were assessed between participants with low or high serum vitamin D in the intermediate-VDR group.

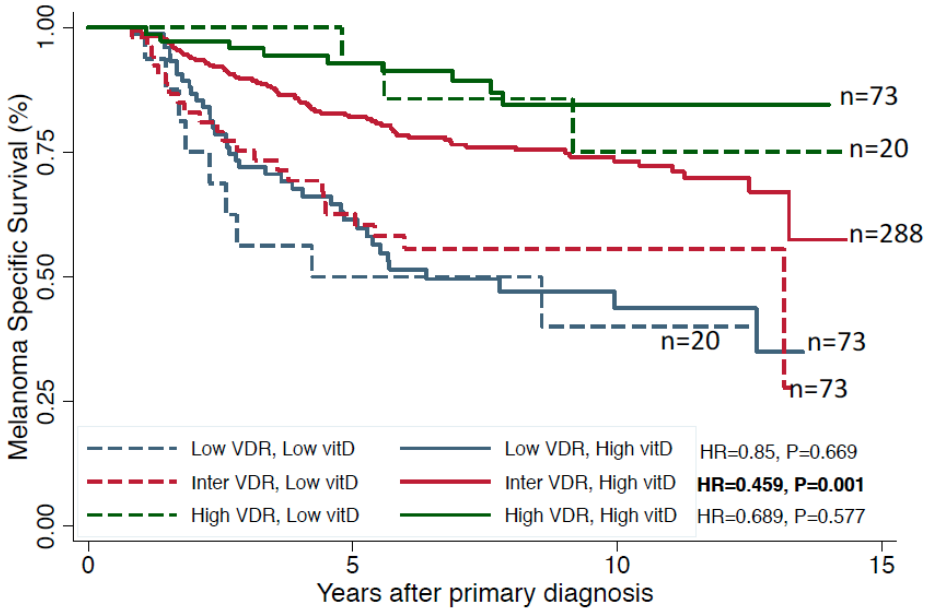


Figure 3.7: Context-specific protect effect of serum vitamin D on melanoma death

Vitamin D-VDR subgroup analysis splitting each of the three VDR-groups into two based on their serum vitamin D level (low or high serum vitamin D) to produce 6 groups. The Kaplan-Meier plot depicts the difference in survival of these 6 groups. Participants in the intermediate-VDR group who had higher serum vitamin D level at diagnosis having significantly improved prognosis compared to those with low vitamin D.

3.4.3.1 Clinicopathological features associated with serum vitamin D level in the intermediate-VDR group

There was no significant variation in age at diagnosis, sex, tumour mitotic rate, site of melanoma and ulceration status between participants with low (n=73) or high (n=288) serum vitamin D levels in the intermediate-VDR group (Table 3.10). However, participants with low serum vitamin D in the intermediate-VDR group had significantly higher Breslow thickness (linear regression test, P=0.02, Figure 3.8A, Table 3.10) and higher frequency of AJCC stage II tumours (Pearson's Chi-squared test, compared to AJCC stage I tumours: P=0.01, Figure 3.8B, Table 3.10), when compared to participants with high serum vitamin D. Additionally, histopathological measure of vascular invasion was available for a subset of patients (n=326) for whom a measure of serum vitamin D levels at diagnosis was available. The frequency of tumours with vascular invasion was significantly (Pearson's Chi-squared test, P=0.01) higher in

participants with low serum vitamin D (20% had vascular invasion) compared to those with high serum vitamin D (10% had vascular invasion) Figure 3.8C, Table 3.10).

Table 3.10: Association of serum vitamin D with clinicopathological features in melanomas from the intermediate-VDR group

[B] indicates the group used as the baseline for comparison with other groups.

Association of serum vitamin D (in the intermediate-VDR group) with	Reg Coef	Std. Error	P-val
Age (years)	-0.002	0.001	0.20
Sex			
Females[B]			
Males	0.03	0.03	0.31
AJCC Stage			
Stage I[B]			
Stage II	-0.10	0.04	0.01
Stage III	-0.03	0.06	0.60
Mitotic rate			
<1 mitoses/mm^2 tumour[B]			
>=1mitoses/mm^2 tumour	0.01	0.03	0.79
Breslow thickness (mm)	-0.68	0.29	0.02
Ulceration status			
No[B]			
Yes	-0.05	0.06	0.43
Site of melanoma			
Head[B]			
Limbs	-0.09	0.06	0.11
Trunk	-0.04	0.06	0.46
Rare (sun protected sites)	-0.06	0.11	0.58
Vascular invasion			
No[B]			
Yes	-0.16	0.06	0.01

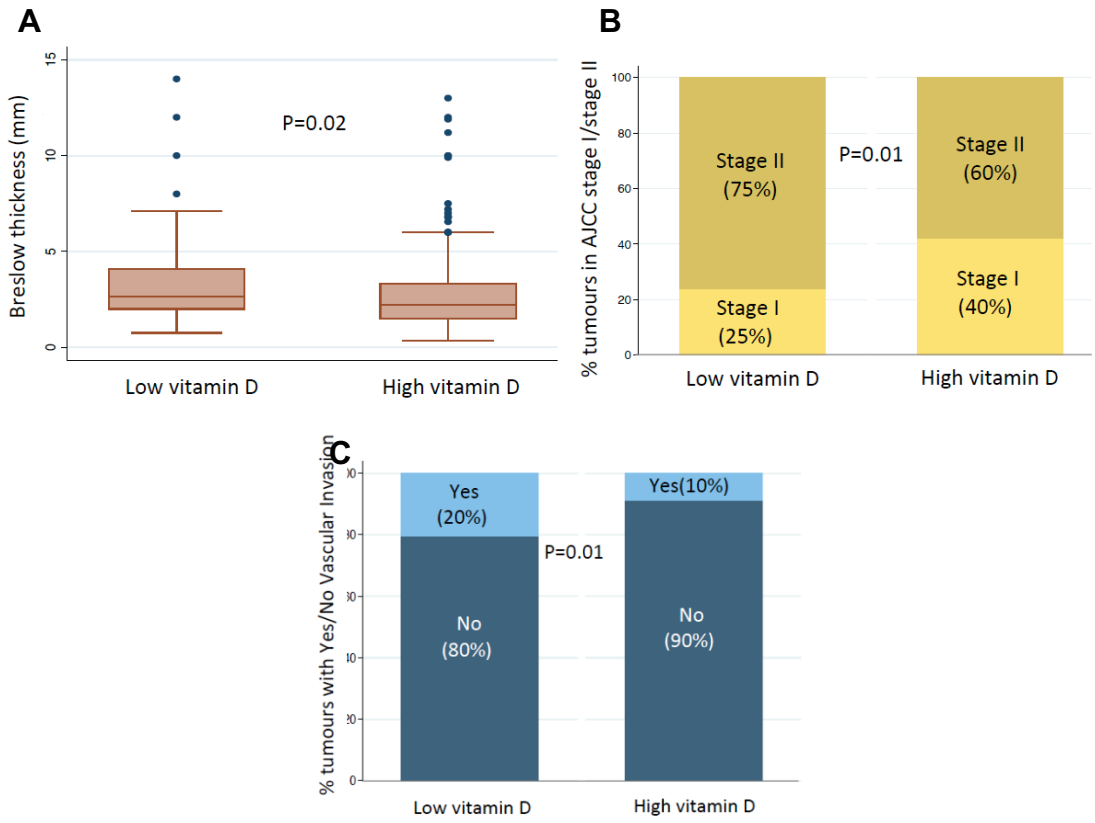


Figure 3.8: Clinicopathological features which varied significantly between tumours from participants with low (n=73) and high (n=288) serum vitamin D in the intermediate-VDR group

A: Comparison of Breslow thickness (mm) between participants with low/high serum vitamin D in the intermediate-VDR group. P-value from linear regression model;

B: Comparison of AJCC stage between participants with low/high serum vitamin D in the intermediate-VDR group. P-value from Pearson's Chi-squared test;

C: Comparison of vascular invasion between participants with low/high serum vitamin D in the intermediate-VDR group. P-value from Pearson's Chi-squared test

In an effort to identify the potential reason behind the low or high levels of serum vitamin D in the intermediate-VDR participants, self-reported supplementation data were used. Among the intermediate-VDR group, those with high serum vitamin D were more likely to have taken dietary supplements (46%) compared to those with low serum vitamin D (20%) (P=0.0004, Figure 3.9). However, in the low-VDR groups, the proportion of patients who had reported to have taken dietary supplements was borderline significantly associated with their serum vitamin D levels (P=0.062) (Table 3.11). In the

high-VDR groups, the proportion of patients who had reported to have taken/not taken dietary supplements was not significantly associated with their serum vitamin D levels (P=0.23) (Table 3.12). This observation could be reflective of the fact that vitamin D intake is proportional to the serum vitamin D levels, but this effect is apparent in the intermediate-VDR group because this group has the highest number of participants.

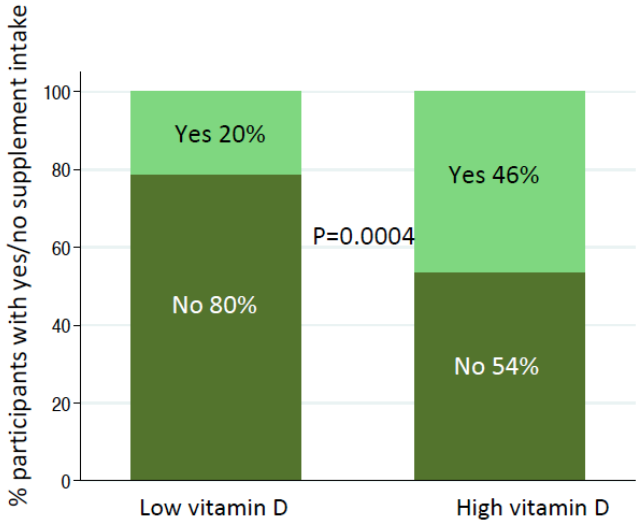


Figure 3.9: Association of self-reported dietary supplement intake with serum vitamin D in participants of the intermediate-VDR group

Comparison of supplement intake ('Yes' or 'No' in the year preceding questionnaire) between participants with low/high serum vitamin D in the intermediate-VDR group. P-value from Pearson's Chi-squared test

Table 3.11: Association of supplement intake in participants with low and high serum vitamin D in the low-VDR group

Supplement intake in the past 1 year	Low vitamin D	High vitamin D
Yes	4 (25%)	38 (51%)
No	12 (75%)	37 (49%)

Table 3.12: Association of supplement intake in participants with low and high serum vitamin D in the high-VDR group

Supplement intake in the past 1 year	Low vitamin D	High vitamin D
Yes	3 (20%)	27 (36%)
No	12 (80%)	48 (64%)

3.4.3.2 Imputed immune cell scores, cytokine and chemokine gene expression associated with serum vitamin D level in the intermediate-VDR group

Since vitamin D has been shown to affect components of the immune system (as discussed in 3.1), the immune landscape of tumours in participants with high (≥25nmol/L) or low (<25nmol/L) serum vitamin D levels in the intermediate-VDR group was compared. This was done using imputed immune cell scores and the expression of cytokine and chemokine genes. The imputed immune cell scores (see 3.3), henceforth referred to as Angelova immune cell scores were used. Briefly, each tumour sample was assigned 26 Angelova immune cell scores (pertaining to 26 immune cell types) based on expression of genes deemed to be uniquely expressed by the particular cell type. Angelova immune cell scores represent a quantitative, *in silico* measure of the tumour's immune landscape. In comparing the 26 Angelova immune cell scores between the two groups, 2 of the 26 imputed immune cell scores were significantly lower in tumours of participants with high serum vitamin D: scores for neutrophils (P=0.02) and monocytes (P=0.04) (Table 3.13).

The differential expression of 154 cytokine and chemokine genes was compared between the high and low vitamin D participants in the intermediate-VDR group (Appendix Table T3-1). After adjusting for multiple correction, only the expression of *CXCL2* was significantly lower in the high vitamin D participants compared to low vitamin D participants (adjusted P=0.03).

Table 3.13: Differences in imputed immune cell scores between participants with low (<25nmol/L) or high (>25nmol/L) serum vitamin D in the intermediate-VDR group

Imputed immune cell score	P-val (from T-test)	Mean expression in low vitamin D group	Mean expression in high vitamin D group
Neutrophils	0.02	0.16	-0.01
Monocytes	0.04	0.16	-0.01
T Gamma Delta cells	0.07	0.18	0.02
Macrophages	0.07	0.13	-0.01
Natural Killer cels	0.08	0.20	0.04
Cytotoxic cells	0.09	0.24	0.05
Natural Killer T cells	0.11	0.21	0.05
Central memory CD8 cells	0.14	0.15	0.02
Mast cells	0.17	0.11	0.02
Immature B cells	0.17	0.16	0.02
T cells	0.18	0.16	0.05
Myeloid Derived Suppressor Cells	0.18	0.16	0.04
Th17	0.19	-0.09	-0.02
Dendritic cells	0.19	0.12	0.04
NK56 dim	0.19	0.14	0.03
Th1	0.24	0.16	0.06
Activated B cells	0.25	0.17	0.06
Effector_memory_CD8	0.26	0.13	0.04
Central memory CD4	0.26	0.13	0.05
T-regulatory cells	0.32	0.11	0.03
T Follicular Helper cells	0.36	0.17	0.07
NK56 bright	0.67	0.13	0.08
Plasmacytoid Dendritic Cells	0.74	-0.03	0.01
Immature Dendritic Cells	0.79	0.06	0.04
Th2	0.85	0.08	0.06
Eosinophils	0.90	0.00	-0.01
Activated CD4 cells	0.91	0.01	0.00
Memory B cells	0.98	-0.03	-0.03

3.4.3.3 Whole-transcriptome correlations with serum vitamin D level in the intermediate-VDR group

Given that serum vitamin D levels ≥25nmol/L were protective for melanoma death only in the intermediate-VDR group of tumours, a whole-transcriptome correlation analysis was performed to identify the genes that were significantly differentially expressed between participants with low serum vitamin D (n=58) and high serum vitamin D (n=303) in the intermediate-VDR subgroup. To this effect, a linear regression model was used to estimate the correlation between each gene and serum vitamin D. The strength of correlation was gauged by the P-value (produced by the linear regression model), which was subsequently adjusted for multiple correction to produce the FDR (False Discovery Rate). Of the 20,560 genes tested, none of the genes reached the significance threshold of FDR<0.05. This is depicted in a histogram of FDR values corresponding to each gene, with the majority of the genes having FDR>0.05 (Figure 3.10).

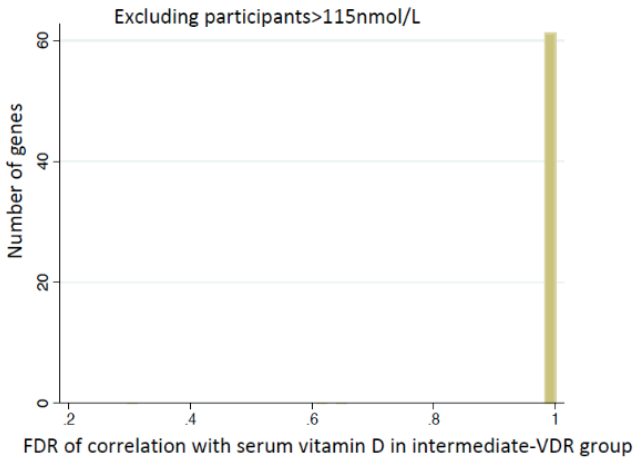


Figure 3.10: Frequency of genes that are significantly associated with serum vitamin D in the intermediate-VDR group, excluding participants with serum vitamin D >115 nmol/L

FDR values generated from P-values (Benjamini-Hochberg FDR) produced from a linear regression model, regressing each gene (of the 20,560 genes) with serum vitamin D in the intermediate-VDR group are plotted in a histogram. Most of the genes have FDR close to 1 as indicated by the histogram.

In summary: in the 700 LMC participants, high serum vitamin D levels (>25nmol/L) offer a survival benefit only in a subgroup of patients belonging to the intermediate-VDR group. An agnostic whole-transcriptome correlation analysis of the intermediate-VDR tumours revealed that there were no genes that were significantly differentially expressed between participants with low or high serum vitamin. However, imputed immune cell scores for neutrophils and monocytes were inversely associated with serum vitamin D levels, as was the expression of *CXCL2.*

As mentioned earlier in this section, there were 3 participants with serum vitamin D levels >115 nmol/L, who were not included for analyses thus far. Though there are only 3 participants, their levels of serum vitamin D are still within the physiological range. Thus, it was of interest to assess the transcriptomic characteristics of melanomas >115nmol/L, which is described in the following section 3.4.3.4.

3.4.3.4 Transcriptomic characteristics of melanomas from participants with serum vitamin D levels >115nmol/L

Among the 703 LMC primary melanomas, season-adjusted serum vitamin D levels were available for 549 participants. However, 3 of the 549 patients had serum vitamin D levels greater than 115nmol/L. The analyses described in section 3.4.3 were done excluding these three patients. However, this section (3.4.4) focuses on analyses done including these three patients i.e. using all 549 participants for whom serum vitamin D levels were available. The three participants with serum vitamin D greater than 115nmol/L belonged to intermediate-VDR group with high serum vitamin D. When a whole-transcriptome correlation was performed with serum vitamin D in the intermediate-VDR group, 441 genes were found to correlate significantly with serum vitamin D after adjusting for multiple correction (FDR<0.05, Figure 3.11B).

At the same multiple correction threshold, there were no genes that correlated significantly with serum vitamin D in the low-VDR (Figure 3.11A) and high-VDR (Figure 3.11C) groups. These findings are depicted using volcano plots.

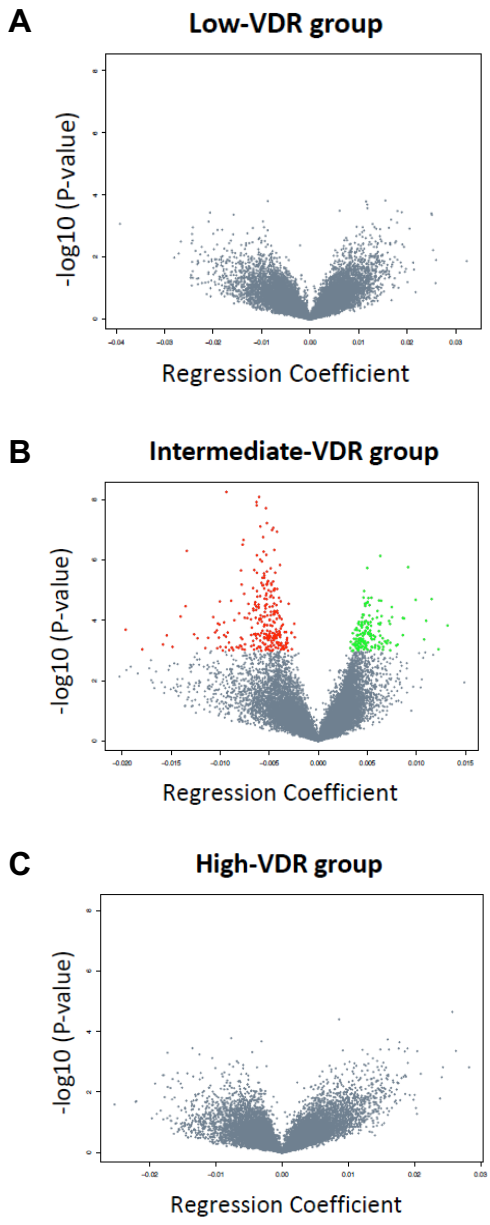


Figure 3.11: Volcano plots of genes that are significantly associated with serum vitamin D in the low-, high- and intermediate-VDR groups (including participants with serum vitamin D >115 nmol/L)

A: Volcano plot of genes which correlate with serum vitamin D in the low-VDR group (n=119);

B: Volcano plot of genes which correlate with serum vitamin D in the intermediate-VDR group (n=465). Red dots denote significantly negatively correlated genes and green dots denote significantly positively correlated genes;

C: Volcano plot of genes which correlate with serum vitamin D in the high-VDR group (n=119).

Thus, the inclusion of participants with serum vitamin D greater than 115nmol/L significantly affects the whole-transcriptome correlates of serum vitamin D in the intermediate-VDR subgroup of tumours. This highlights that serum vitamin D levels greater than 115nmol/L are associated with variation in gene expression in the intermediate-VDR group. This prompted further investigation to identify the biological function of these genes associated with 'very high' serum vitamin D. Among the 441 genes that correlated with serum vitamin D in the intermediate-VDR group, 283 correlated negatively whereas 158 correlated positively with serum vitamin D. The negatively correlated genes were enriched for pathways such as T cell activation, mitochondrial translation, MHC class II antigen presentation, HIF-1 pathway and Renal cell carcinoma (Table 3.14). The nodal genes pertaining to these pathways were *PPP3CB, PPP3CC, HLADOA, HLADQA1, BRAF, MRPS5, MRPS33, MRPS30, ARNT, PLCG1, HMOX1, EGLN1* and *PFKL.*

The positively correlated genes were enriched for pathways such as retinol metabolism, Cytochrome P450-drug metabolism, fatty acid degradation and glycolysis/gluconeogenesis, to name a few (Table 3.15). The nodal genes pertaining to these pathways were *ADH1A*, *TUBA3C*, *UGT1A10*, *DYNC1LI1* and *MIP.*

Because the inclusion of only 3 participants has a radical effect on the transcriptomic correlates of serum vitamin D, these participants were deemed to be outliers and excluded for subsequent analyses.

Table 3.14: List of pathways enriched for genes negatively correlated with serum vitamin D in the intermediate-VDR group without exclusion of the 3 samples from participants with levels >115nmol/L.

Pathways	P-value	FDR	Nodes
T cell activation(P)	1.15E-05	3.95E-03	PPP3CB, PPP3CC, HLA-DOA, PLCG1, HLA-DQA1, BRAF
Mitochondrial translation(R)	1.94E-05	3.95E-03	MRPL4, MRPL40, MRPL9, MRPS5, MRPS33, MRPS30
MHC class II antigen presentation(R)	2.32E-04	0.0313	CLTA, HLA-DOA, KIFAP3, DCTN2, HLA-DQA1
HIF-1 signalling pathway(K)	4.51E-04	0.0445	ARNT, HMOX1, PLCG1, EGLN1, PFKL
Regulation of Hypoxia-inducible Factor (HIF) by oxygen(R)	5.84E-04	0.0445	ARNT, EPAS1, EGLN1
HIF-1-alpha transcription factor network(N)	7.67E-04	0.0445	ARNT, HMOX1, EGLN1, PFKL
Renal cell carcinoma(K)	7.67E-04	0.0445	ARNT, EPAS1, EGLN1, BRAF

Table 3.15: List of pathways enriched for genes positively correlated with serum vitamin D in the intermediate-VDR group

Pathway	P-value	FDR	Nodes
Retinol metabolism(K)	4.00E-04	3.71E-03	UGT1A10, ADH1C
Drug metabolism - cytochrome P450(K)	4.37E-04	3.71E-03	UGT1A10, ADH1C
Metabolism of xenobiotics by cytochrome P450(K)	5.17E-04	3.71E-03	UGT1A10, ADH1C
Chemical carcinogenesis(K)	6.19E-04	3.71E-03	UGT1A10, ADH1C
Phagosome(K)	2.18E-03	0.0109	DYNC1LI1, TUBA3C
Passive transport by Aquaporins(R)	3.91E-03	0.0156	MIP
downregulated of mta-3 in er-negative breast tumours(B)	8.78E-03	0.0204	TUBA3C
stathmin and breast cancer resistance to antimicrotubule agents(B)	0.0107	0.0204	TUBA3C
Ascorbate and aldarate metabolism(K)	0.0131	0.0204	UGT1A10
Signalling by Retinoic Acid(R)	0.0165	0.0204	ADH1C
Tyrosine metabolism(K)	0.017	0.0204	ADH1C
Pentose and glucuronate interconversions(K)	0.0175	0.0204	UGT1A10
Protein folding(R)	0.0185	0.0204	TUBA3C
Porphyrin and chlorophyll metabolism(K)	0.0204	0.0204	UGT1A10
Vasopressin-regulated water reabsorption(K)	0.0214	0.0214	DYNC1LI1
Fatty acid degradation(K)	0.0214	0.0214	ADH1C
Drug metabolism - other enzymes(K)	0.0223	0.0223	UGT1A10
Pathogenic Escherichia coli infection(K)	0.0266	0.0266	TUBA3C
Starch and sucrose metabolism(K)	0.0271	0.0271	UGT1A10
Steroid hormone biosynthesis(K)	0.0281	0.0281	UGT1A10
Glycolysis / Gluconeogenesis(K)	0.0324	0.0324	ADH1C
Phase 1 - Functionalization of compounds(R)	0.0338	0.0338	ADH1C
Salmonella infection(K)	0.0414	0.0414	DYNC1LI1
Gap junction(K)	0.0423	0.0423	TUBA3C

Since these pathways were identified in the whole-transcriptome correlation including the 3 participants with >115nmol/L serum vitamin D (549 participants), it was of interest to test if these genes/pathways also correlated with serum vitamin D in the

intermediate-VDR group, excluding the 3 participants (546 participants). To this effect, none of the negatively correlated nodal genes (from analysis including 3 outliers) were differentially expressed between participants with low serum vitamin D and high serum vitamin D (in cohort excluding 3 outliers): *PPP3CB* (P=0.67), *PPP3CC* (P=0.11), *HLADOA* (P=0.06), *HLADQA1* (P=0.10), *BRAF* (P=0.70), *MRPS5* (P=0.23), *MRPS33* (P=0.23), *MRPS30* (P=0.40), *ARNT* (P=0.29), *PLCG1* (P=0.86), *HMOX1* (P=0.27), *EGLN1* (P=0.13) and *PFKL* (P=0.33). Among the positively correlated genes, there was no significant variation for *ADH1A* (P=0.16), *TUBA3C* (P=0.89) and *UGT1A10* (P=0.13) between participant with low or high serum vitamin D (in cohort excluding 3 outliers). However, *DYNC1LI1* (P=0.02) and *MIP* (P=0.001) were expressed significantly higher in participants with high serum vitamin D compared to those with low serum vitamin D. *DYNC1LI1* is a coding gene which codes for the Dynein Cytoplasmic 1 Light Intermediate Chain 1 protein, whose Gene Ontology (GO) terms include RNA binding and microtubule motor activity. *MIP* codes for the Major Intrinsic Protein: a water-transporting aquaporin, whose Gene Ontology (GO) terms include calmodulin binding and water channel activity.

In summary: Among the genes identified in the whole-transcriptome correlation with serum vitamin D in cohort of patients with serum vitamin D >115nmol/L, only *MIP* and *DYNC1LI1* are of relevance in the dataset of patients excluding these outliers.

3.4.4 Tumour *VDR* expression: transcriptomic characteristics

The prognostic significance of tumour *VDR* expression in both the LMC primary and TCGA metastatic melanomas has been discussed in section 3.4.2.2. The significant association of tumour *VDR* expression with improved melanoma survival necessitated the identification of genes that correlate significantly with tumour *VDR* expression. The rationale was that the significantly correlated genes and the correspondingly enriched biological pathways would offer insights into the mechanistic basis of the protective effect of *VDR* on melanoma death. To this effect, an agnostic correlation analysis was used to identify genes and signalling pathways that most significantly correlated with *VDR* expression in the LMC primary melanomas. This agnostic correlation with *VDR* was also performed on TCGA transcriptomes from metastatic melanomas, for comparison. The concordance between genes agnostically identified to be correlated with *VDR* in the LMC and genomic regions reported to have a VDR binding site (based on ChIP-Seq experiments) was also assessed. Factors contributing to reduced VDR expression were also explored in the LMC as well as TCGA melanoma data sets.

3.4.4.1 Whole-transcriptome correlation with tumour VDR expression in the LMC primary melanomas

The correlation of tumour *VDR* expression with 20,560 genes (each represented by a unique probe) was computed using a linear regression model. The direction of correlation (positive or negative) was denoted by the regression coefficient ('R') while the significance of correlation with a particular gene was denoted by the P-value. Given that 20,560 tests were performed (one per gene), it was necessary to account for multiple testing. This was done using the Benjamini-Hochberg FDR method which produced an adjusted P-value ('adj-P-value'). 12,158 genes with adj-P-value<0.05 were identified to be significantly correlated with *VDR*. Additionally, a regression coefficient threshold of 0.2 was used to identify the strongest correlates of *VDR* expression. In doing so, 1383 genes with R<-0.2 were identified to be significant negative correlates, whereas 2025 genes with R>0.2 were identified to be significant positive correlates with *VDR*. The aforementioned significant transcriptomic correlates of *VDR* are depicted in a volcano plot (Figure 3.12). Each of these gene lists was used in a subsequent enrichment analysis. Reactome FIViz enrichment of the 1383 negatively correlated identified pathways known to be involved in Mitotic Metaphase/Anaphase, Wnt signalling pathway, Mitochondrial translation, cadherin signalling, TCA cycle and oxidative phosphorylation, to name a few (Table 3.16), see Appendix Table T3-2 for full list). Similar Reactome FiViz analysis of the 2025 positive correlates resulted in pathways known to be involved in Extracellular Matrix organization, cytokine-cytokine receptor interaction, TNF signalling pathway, Interferon gamma signalling, Osteoclast differentiation, IL-12 mediated signalling events and NF-kappa B signalling pathway, to name a few (Table 3.17, see Appendix Table T3-3 for full list). A clear distinction between the two lists of pathways was observed: while the negatively correlating pathways seem to be predominantly cell cycle and proliferation related, the positively correlating pathways include those involved in immune response.

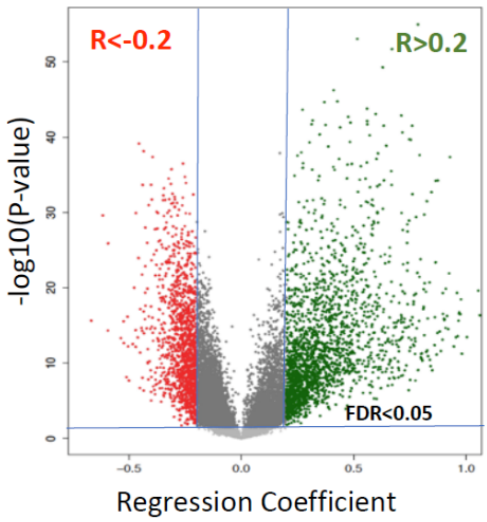


Figure 3.12: Volcano plot of genes that are significantly associated with tumour VDR expression

Genes with FDR<0.05 were considered to be significant correlates. Among these, 1383 genes with R<-0.2 were identified to be significant negative correlates (red dots), whereas 2025 genes with R>0.2 were identified to be significant positive (green dots) correlates with *VDR*.

Table 3.16: List of top pathways enriched for genes negatively correlated with tumour VDR expression

Pathways	**P-val**
Mitotic Prometaphase(R)	1.38×10^{-9}
Wnt signalling pathway(P)	3.58×10^{-8}
Mitotic Metaphase and Anaphase(R)	7.50×10^{-8}
Mitochondrial translation(R)	5.24×10^{-7}
Cadherin signalling pathway(P)	2.47×10^{-6}
SUMOylation(R)	2.47×10^{-6}
RNA Polymerase I, RNA Polymerase III, and Mitochondrial Transcription(R)	3.83×10^{-6}
The citric acid (TCA) cycle and respiratory electron transport(R)	1.06×10^{-5}
Cell cycle(K)	1.81×10^{-5}
Mitotic G2-G2/M phases(R)	2.16×10^{-5}

Table 3.17: List of top pathways enriched for genes positively correlated with tumour *VDR* expression

Pathways	P-val
Extracellular matrix organization(R)	$1.11x10^{-16}$
Cytokine-cytokine receptor interaction(K)	$1.11x10^{-16}$
Pathways in cancer(K)	$1.11x10^{-16}$
Cell adhesion molecules (CAMs)(K)	$3x10^{-15}$
TNF signalling pathway(K)	$3.28x10^{-13}$
Interferon gamma signalling(R)	$1.64x10^{-12}$
Osteoclast differentiation(K)	$8.05x10^{-12}$
Chemokine signalling pathway(K)	$1.09x10^{-11}$
IL12-mediated signalling events(N)	$2.16x10^{-11}$
NF-kappa B signalling pathway(K)	$2.38x10^{-11}$

Since VDR is known to be expressed by keratinocytes (371), I examined the possibility that putative epidermal contamination (during tumour core sampling) could confound the findings from the whole-transcriptome correlation analysis. To exclude this possibility, the analysis was adjusted for expression of the gene coding for filagrin (*FLG2*-adjusted) in a whole-transcriptome correlation with *VDR*, filagrin being a marker of keratinocyte differentiation (380). This analysis identified 11,471 genes to be significantly correlated (adj-P-value<0.05) with *VDR,* independent of the expression of *FLG2*. Of these 11,471 genes, 95.15% (10,951 genes) were also identified in the whole-transcriptome correlation analysis (unadjusted for *FLG2*) as significant correlates of *VDR*. Thus, the transcriptomic correlates of *VDR* remained largely unchanged after adjusting for *FLG2* expression. Next, the biological pathways enriched for these 11,471 genes which correlated with *VDR* after adjusting for *FLG2* were assessed. The genes negatively correlated with *VDR* (*FLG2*-adjused) were enriched for pathways such as Mitotic Metaphase/Anaphase, Wnt signalling pathway, Mitochondrial translation, cadherin signalling, TCA cycle and oxidative phosphorylation, to name a few (see Appendix Table T3-4 for full list). Similar Reactome FiViz analysis of the positive correlates (*FLG2*-adjusted) identified pathways known to be involved in Extracellular Matrix organization, cytokine-cytokine receptor interaction, TNF signalling pathway, Interferon gamma signalling, Osteoclast differentiation, IL-12 mediated signalling events and the NF-kappa B signalling pathway, to name a few (see Appendix Table T3-5 for full list). Taken together, the biological pathways enriched for genes correlating with *VDR* remained largely unchanged, suggesting that the original pathways

associated with *VDR* expression were not significantly confounded by keratinocyte signals.

Though the presence of keratinocyte-derived genes and pathways have been dealt with as 'contamination' when assessing the transcriptomic correlates of VDR, in reality presence of keratinocytes within melanoma tumour mass is not uncommon. Sampling of keratinocytes while sampling for melanomas is a common issue encountered by other groups studying primary melanomas and by our own group. However, the resolution of this issue is not trivial. In that, a subpopulation of tumours cannot be denounced as epidermal contamination, since an epidermal-like phenotype of undifferentiated melanomas have been previously described (381). However, the justification for my '*FLG2*-adjusted' whole transcriptome correlation with *VDR*, is to ensure that the correlated pathways are not simply a reflection of high epidermal content.

3.4.4.1.1 Do the pathways that correlate significantly with *VDR* also correlate with serum vitamin D in the intermediate-VDR participants?

No genes were found to correlate agnostically with serum vitamin D in the intermediate-VDR group (546 participants, excluding the outliers), based on an agnostic analysis (section 3.4.3.3). So, I adopted an alternative 'candidate approach', wherein I tested if the most significantly correlated pathways with *VDR*, also correlate with serum vitamin D in the intermediate-VDR group. To this effect, pathway scores were computed for the negatively (Wnt signalling, mitochondrial translation, cell cycle, mitotic metaphase and anaphase) and positively (extracellular matrix organisation, Interferon gamma signalling, IL12 signalling, TCR signalling on naïve CD4 and CD8 cells, TNF signalling and NK-mediated cell killing) correlated pathways. The pathway scores were computed as the average expression of all nodal genes pertaining to that pathway (see Appendix table T3-2 and T3-3 for list of nodal genes pertaining to each pathway). The following pathways which correlate positively with *VDR*, also correlate inversely with serum vitamin D in the intermediate-VDR group: IL12 signalling, TCR signalling on naïve CD4 and CD8 cells and NK-mediated cell killing (Table 3.18).

In summary: In addition to the findings in section 3.4.3.2 (where serum vitamin D was found to correlate inversely with neutrophil and monocyte imputed immune cell scores), the current 'candidate approach' revealed that serum vitamin D is inversely associated with specific immune signalling pathways in the intermediate-VDR group.

Table 3.18: Correlation of VDR-correlated pathways with serum vitamin D in the intermediate-VDR group

The regression coefficient and P-value indicate the strength of association of pathway scores with serum vitamin D in the intermediate-VDR group of tumours.

Pathway scores for pathways correlated with *VDR* expression	Regression coefficient	P-value
Pathways negatively-correlated with VDR		
Wnt signalling	0.0009	0.39
Mitochondrial translation	-0.00015	0.88
Cell cycle	-0.0002	0.78
Mitotic metaphase and anaphase	-0.0004	0.72
Pathways positively-correlated with VDR		
Extracellular matrix organisation	0.00003	0.97
Interferon gamma signalling	-0.002	0.11
IL12 signalling	-0.005	0.03
TCR signalling on naïve CD4 and CD8 cells	-0.003	0.05
TNF signalling	-0.001	0.25
NK-mediated cell killing	-0.004	0.02

3.4.4.2 Whole-transcriptome correlation with tumour VDR expression in the TCGA metastatic melanomas

High *VDR* expression was protective for melanoma death in the TCGA metastatic melanoma cohort (3.4.2.2). So, a whole-transcriptome correlation with *VDR* was carried out to identify the most significantly correlated genes and pathways in the TCGA metastatic melanoma cohort. Similar to the whole-transcriptome analysis of the LMC melanomas, a linear regression model was used. The direction of correlation (positive or negative) was denoted by the regression coefficient ('R') while the significance of correlation with a particular gene was denoted by the P-value. Upon applying the same multiple testing and regression coefficient thresholds as the LMC whole-transcriptome analysis, 8756 genes correlated significantly with *VDR* expression in TCGA data after correcting for multiple testing using Benjamini-Hochberg FDR (adj-P-value<0.05). Of these, 1500 genes correlated positively with *VDR* expression (R>0.2) and were enriched for pathways such as NFκB, TNF, IFNα/β, IFNγ, IL12-mediated, TCR and chemokine signalling in naïve CD4 T cells (see Appendix Table T3-6 for full list). 912 genes correlated negatively with *VDR* expression (R<-0.2) and were enriched for pathways such as Wnt signalling, extracellular matrix organization, cadherin signalling,

eukaryotic translation initiation, TGFβ and VEGFR1 signalling (see Appendix Table T3-7 for full list).

3.4.4.3 VDR- transcription factor binding

Upon vitamin D binding and RXR dimerization, VDR acts as a transcription factor that binds to specific VDRE-containing genomic regions, which are under direct transcriptional control of VDR. Based on my analyses, the whole-transcriptome correlation with *VDR* in the LMC and TCGA datasets identified the significant transcriptomic correlates of *VDR* expression. This put forth the question: what proportion of genes that are under the direct transcriptional control of VDR also correlate with *VDR* expression in the LMC melanomas? This was addressed using two approaches:

i) Using previously reported ChIP-Seq datasets which identify VDR-binding genomic sites
ii) By identifying genomic regions that contain the VDR-binding motif

Both analyses identified genomic regions which are likely to be under direct transcriptional control of VDR: either VDR binding targets identified by ChIP-Seq or genomic regions which contain the motif for VDR-binding. Once these genomic regions were identified by either approach, they were mapped to the coding genes which are likely to fall within the regions. The correlation of the genes thus identified, with tumour *VDR* expression in the LMC was assessed. Thus, the 'overlapping' set of genes which correlate with *VDR* in the LMC and are also likely to be transcriptionally controlled by VDR, were identified.

The description of each approach is detailed below.

3.4.4.3.1 Comparison with VDR ChIP-Seq data

Currently there are 6 tissue types for which VDR ChIP-Seq data was generated and jointly analysed in a meta-analysis by Tuoresmaki et al (374). The tissue types analysed in this study and the original studies in which the VDR ChIP-Seq was first described are were: THP-1(human monocytic cell line)- LPS stimulated and unstimulated (382), GM10855 (immortalised lymphoblastoid cell line) (383), GM10861 (immortalised lymphoblastoid cell line) (383), LX2 (hepatic stellate cells) (384) and LS180 (colorectal cancer cells) (385). The findings from the meta-analysis by Tuoresmaki et al are crucial because they used a harmonised analysis pipeline (MACS, version 2) to re-analyse results from the individual ChIP-Seq datasets, which were derived using different peak-calling software and alternative threshold settings. In total, the meta-analysis identified 21,776 non-overlapping genomic sites across all 6 datasets. Interestingly, they reported

that 67% of these sites are unique to a single cell type, while only 54 non-overlapping genomic *VDR*-binding regions were common to all 6 tissue types. I obtained data pertaining to these 54 regions: chromosomal location and genomic start and end sites, which were downloaded from the Tuoresmaki et al supplementary data. In mapping the 54 genomic binding sites to genes, 73 genes were identified (using GREAT, see Methods). Of the 73 genes, 43 genes (58%) correlated significantly (at FDR<0.05) with tumour *VDR* expression in the Leeds data, indicating that a proportion of genes which are transcriptionally controlled by *VDR* are also correlated with *VDR* expression in our data set.

In addition, I also extracted the tissue-specific VDR-binding regions i.e. the number of VDR-binding regions in each of the 6 tissues analysed by Tuoresmaki et al. The number of VDR-binding regions in each cell line are: THP-1 LPS stimulated (1318 regions) and unstimulated (1146 regions), GM10855 (7887 regions), GM10861 (13924 regions), LX2 (2291 regions) and LS180 (3770 regions). These regions were mapped to genes (using GREAT, see methods). The number of genes mapped to genomic regions in each cell line is detailed in Table 3.19. Of these identified genes. the proportion of genes which also correlated with *VDR* in the LMC dataset was estimated. This percentage was found to be consistent across all 6 cell lines (approx. 57%, Table 3.19). This analysis is therefore indicative that for a significant proportion of genes that correlate with VDR in the LMC primary melanomas are also likely to be under the direct transcriptional control by the VDR transcription factor.

Table 3.19: Comparison of the LMC VDR correlates and VDR ChIP-Seq data from 6 cell lines

First two columns denote number of VDR-binding peaks identified in 6 cell lines (derived from Tuoresmaki et al). Genes mapped to these peaks (using GREAT) and the overlap (percentage) of genes which correlated with *VDR* expression in the LMC, are on columns 3 and 4 respectively.

Tissue type	Source tissue	Number of VDR-binding regions	Number of genes mapped to the VDR-binding regions (Identified by GREAT)	% genes in peak regions that also correlate with *VDR* in the LMC
THP-1: LPS stimulated	Human monocytic cell line	1318	1728	57.29
THP-1	Human monocytic cell line	1146	1385	57.68
GM10855	Lymphoblastoid cell line	7887	6029	57.58
GM10861	Lymphoblastoid cell line	13924	8784	57.38
LX2	Hepatic stellate cells	2291	2803	57.11
LS180	Colorectal cancer cells	3770	3799	56.62

3.4.4.3.2 Comparison using VDR-binding motifs

Being a transcription factor, the protein VDR has a high affinity to bind to DNA sequences composed of a specific motif. Similarly, the VDR-RXR dimer has a high affinity to bind to DNA sequences composed of a specific motif. Since these motifs are transcription factor-specific, it can be inferred that DNA sequences (genomic regions) containing the VDR-specific motif, are likely to bind with high affinity to VDR and hence under its transcriptional control. To identify such VDR-specific motifs and the genomic regions containing the motif, I used the resource Motifmap (386, 387), which identified the binding motifs for VDR (3 regions), VDR:RXR (dimer-binding motif, 12 regions), RXR:VDR (dimer-binding motif, 23 rgions) and the motif common to other nuclear hormone receptors such as PXR and CAR (1 regions). The binding motif, the corresponding sequence logo and the number of genomic regions containing the motif (as catalogued by Motifmap) are described in Figure 3.13. This identified a total of 39 genomic regions which contain the motif for either VDR or VDR as a dimer with RXR. I

mapped these 39 regions to genes (using GREAT, see methods) and identified 49 genes which were associated with these 39 genomic regions. Among the 49 genes, 29 genes (60%) also correlated significantly with *VDR* in the LMC (at FDR<0.05). These genes are listed in Table 3.20.

In order to estimate if the probability of overlap between genes correlating with VDR in LMC and mapping to VDR motif containing regions is greater than that expected by chance, a hypergeometric test was performed, wherein:

The number of genes represented in the LMC transcriptome (N) = 20,560 genes

The number of genes that correlate significantly with *VDR* expression in the LMC (k) = 3408 genes

The number of genes associated with the 39 VDR motif-containing regions (obtained from Motifmap) (n) = 49 genes

Overlapping genes which correlate with *VDR* in LMC and map to VDR motif-containing regions (k) = 29 genes

Based on these parameters, the hypergeometric probability: P(x=49) = 0.043

In other words, the probability of a VDR motif-containing gene to be correlated with *VDR* in the LMC by random chance is 4.3%.

Much like the results from the analysis of ChIP-Seq data described in the previous section, this analysis indicates that a significant proportion of genes which are associated with genomic regions containing the VDR binding-motif are also correlated with VDR in the LMC primary melanomas.

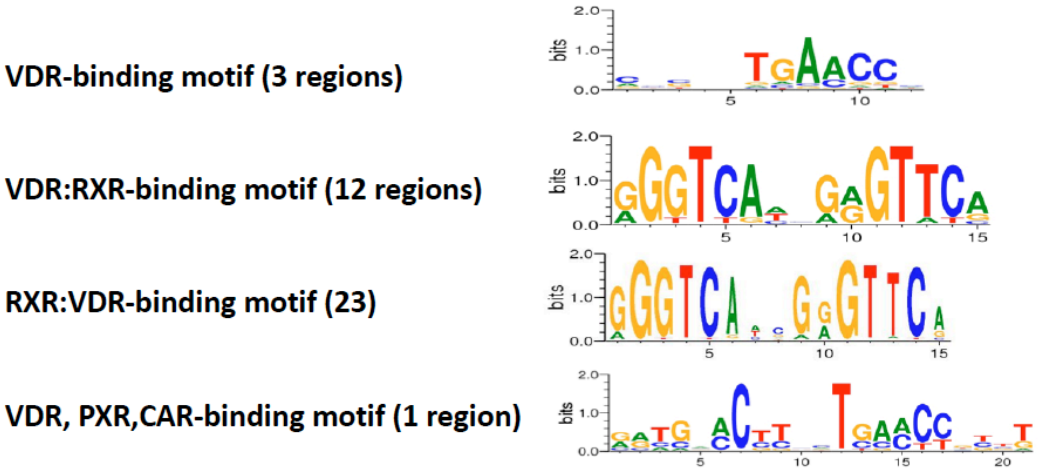


Figure 3.13: VDR-binding motifs: identified by MotifMap.

Table 3. 1: List of genes that correlate with VDR in the LMC primary melanomas and are also mapped to genomic regions containing the VDR-binding motif.

Regression coefficient and p-value from linear regression model.

Gene	Regression Coefficient	p-value
GSDMC	0.70	1E-15
ASTN2	-0.32	1E-15
CHD6	-0.16	1E-12
WDR41	-0.18	1E-11
EFNB2	0.29	1E-11
FGFR1	0.21	5E-10
MSX2	-0.17	8E-10
KCTD3	-0.12	4E-09
KLHL13	-0.34	9E-09
OLFML2B	0.19	5E-08
ADAT2	-0.16	2E-07
RUNX1T1	0.29	2E-07
LMOD3	-0.09	1E-06
ZNF688	-0.14	2E-06
GPC5	-0.15	2E-05
ATF6	-0.11	4E-05
CTLA4	0.23	5E-05
GPR116	0.16	1E-04
LETM2	-0.10	3E-04
SNTB1	-0.14	8E-04
TAF2	-0.09	4E-03
ERCC4	0.09	5E-03
CD28	0.13	6E-03
RNF5	-0.09	1E-02
ENPP2	0.10	1E-02
HOXC11	-0.06	2E-02
HAS2	-0.12	2E-02
SIX3	0.13	3E-02
TRIM32	-0.08	5E-02

3.4.4.4 Factors controlling VDR expression in primary and metastatic melanomas

In order to identify factors which could explain the gradient *VDR* expression, it was hypothesised that Copy Number Alterations (CNAs) and methylation could control expression of *VDR.* These hypotheses were tested using both the LMC primary and TCGA metastatic melanoma datasets.

3.4.4.4.1 VDR expression and Copy Number Alterations in the LMC primaries

Copy Number Alteration (CNA) data were available for 276 primary melanoma samples in LMC dataset. These samples are a subset of the 703 LMC samples used thus far. The generation and normalisation of these data is described in the Methods section. Briefly, DNA samples (from the 276 LMC tumour cores) were used to generate log2-normalised ratios of window read counts, from which gene level median copy number estimates were generated using Gistic 2.0. The availability of these CNA data enabled the comparison of *VDR* expression and *VDR* copy number in the 276 samples. The log2-normalised *VDR* copy number values were found to be centred around 0 i.e. most of the tumour samples had a log2-normalised *VDR* copy number value of 0, suggesting that the *VDR* gene remains diploid. Concordant with this observation, the GISTIC-derived q-value (a confidence measure of copy number alterations across a genomic regions) for the *VDR*-containing region 12q13.11 was >0.25, indicating that *VDR* copy number remains unaltered in most of the 276 samples. The correlation between *VDR* expression and *VDR* CNA (log2-normalised) was assessed using both variables on a continuous as well as categorical scale. When *VDR* expression and *VDR* CNA were represented on a continuous scale, a linear regression model revealed the lack of significant correlation between the two variables (R=0.005, P=0.62, Figure 3.14A). Given that a majority of the samples had log2-normalised *VDR* copy number value of 0, the scale of the VDR CNA values was revised (changed to exponential scale) and the correlation with VDR expression was assessed. The correlation between VDR CNA and expression remained statistically insignificant (R=0.46, P=0.19), suggesting that the lack of correlation is irrespective of the distribution of the CNAs.

Alternatively, *VDR* CNA was compared across the 3 *VDR*-groups (pairwise T-tests, Figure 3.14B) which have been described in section 3.4.1.4. Median *VDR* CNA did not vary significantly between the Low-VDR group compared to the Intermediate-VDR group (P=0.5) and High-VDR group (P=0.35). There was also no significant difference in *VDR* CNA between the Intermediate-VDR group and High-VDR group (P=0.33). Even though *VDR* CNA on a continuous scale reliably represents the distribution of *VDR* CNAs in the 276 samples, the extremities of the distribution

(samples with very high or low *VDR* copy number) were also compared with *VDR* expression. This categorised analysis affects the sample size of the compared groups, since stringent cut-off values will result in reduced sample size per group. However, comparison of categorised *VDR* CNA data can validate observations derived from *VDR* CNA data on a continuous scale. To this effect, *VDR* expression did not vary significantly between the samples belonging to the following *VDR* CNA categories: *VDR* CNA $\leq$-0.3 or $\geq$0.3 (P=0.34, Figure 3.14C), *VDR* CNA $\leq$-0.2 or $\geq$0.2 (P=0.39, Figure 3.14D), *VDR* CNA $\leq$-0.1 or $\geq$0.1 (P=0.17, Figure 3.14E). This is consistent with the lack of significant correlation between *VDR* expression and *VDR* CAN, when represented on a continuous scale. Thus, there is no evidence for significant correlation between *VDR* expression and *VDR* CNA in the 276 LMC primary melanomas.

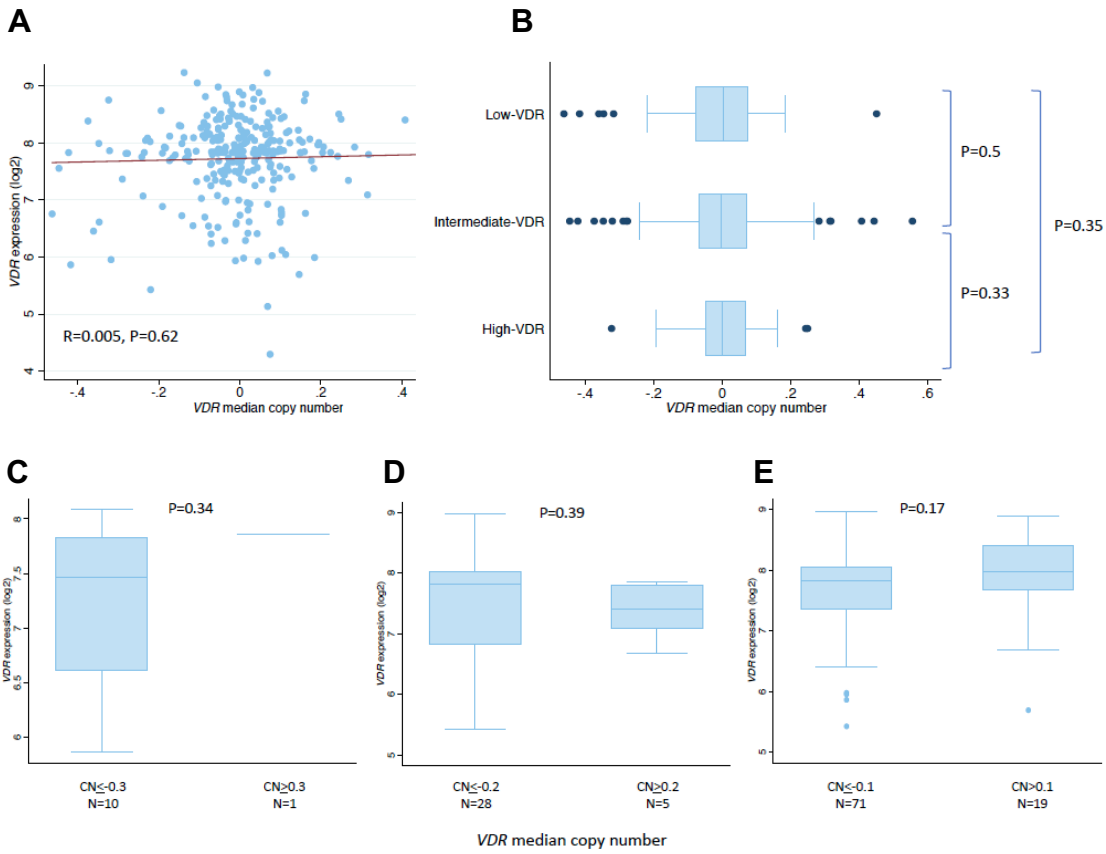


Figure 3.14: Association of VDR expression with VDR CNAs in LMC primary melanomas

A: Correlation between *VDR* expression and *VDR* CNAs: both being represented on a continuous scale, R: Regression Coefficient, P: P-value from linear regression model;

B: Comparison of *VDR* CNAs across the 3 VDR-groups. P-values from pairwise T-tests;

C: Comparison of *VDR* expression between tumours with *VDR* CNA ≤-0.3 or ≥0.3. P-values from pairwise T-tests;

D: Comparison of *VDR* expression between tumours with *VDR* CNA ≤-0.2 or ≥0.2. P-values from pairwise T-tests;

E: Comparison of *VDR* expression between tumours with *VDR* CNA ≤-0.1 or ≥0.1. P-values from pairwise T-tests

3.4.4.4.2 VDR expression and Copy Number Alterations in the TCGA metastatic melanomas

The TCGA melanoma dataset includes matching copy number data and gene expression data for the metastatic melanoma samples. Matching *VDR* CNA and *VDR* expression data were available for distant metastases (n=68) and regional lymph node metastases (n=222), but not for primary tumours. The lack of CNA data in the TCGA data and the relative lack in the LMC data reflect the difficulties in sampling very small melanoma samples. In the TCGA data, this permitted the comparison of *VDR* expression with *VDR* CNA, as well as the variation of *VDR* CNA across metastatic melanomas, but not primary melanomas. To this effect, *VDR* expression correlated significantly and positively with *VDR* CNA in the TCGA metastatic melanomas (R=0.76, P=3.1x10^{-6}, Figure 3.15A). Lower VDR copy number were more common in distant metastases than in regional lymph node metastases (P=0.015, Figure 3.15B). It is to be noted that distant metastases are considered to be indicators of more aggressive disease with worse prognosis (compared to regional metastasis). Thus, the association of *VDR* expression with *VDR* CNA in the TCGA, with increase in copy number with advanced disease, suggests that expression is controlled at least in part by CNA in a progression-dependent manner.

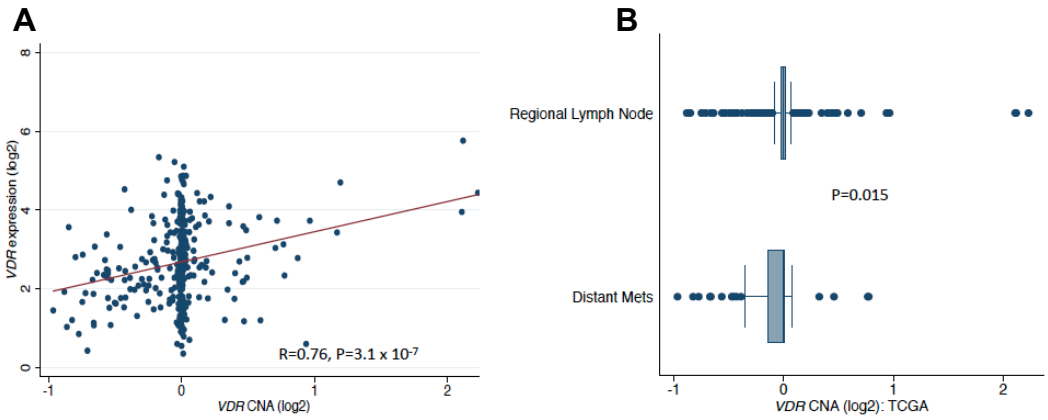


Figure 3.15: Association of VDR expression with VDR CNAs in TCGA melanomas

A: Correlation between *VDR* expression and *VDR* CNAs: both being represented on a continuous scale. R: Regression Coefficient, P: P-value from linear regression model;

B: Comparison of VDR CNAs between distant metastases and regional lymph nodes in the TCGA data. P-value from T-test

3.4.4.4.3 *VDR* expression and methylation in TCGA data

Matching *VDR* expression and methylation data were available for 469 samples in the TCGA melanoma dataset. The VDR methylation status was represented by Beta values, which were downloaded from cBioportal. These data were generated on the HM450 methylation array in which the probe corresponding to a gene covers sites in the promoter region, 5' UTR, first exon, gene body and 3' UTR. *VDR* expression was significantly and inversely correlated (R=-0.06, P= 5.06 x 10^{-11}, Figure 3.16) with methylation Beta values pertaining to the *VDR* genomic region. This observation, combined with the analysis using VDR CNAs suggest that both methylation and copy number control VDR expression.

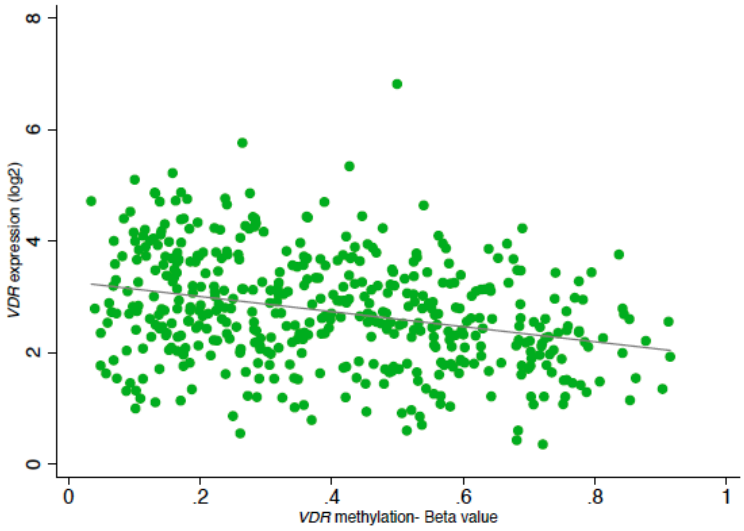


Figure 3.16: Association of VDR expression with methylation in TCGA melanomas

Correlation between *VDR* expression and *VDR* methylation (beta values): both being represented on a continuous scale. P-value from linear regression model

3.4.5 Validation of transcriptomic correlates of tumour *VDR* expression

3.4.5.1 In silico validation

3.4.5.1.1 Reported molecular phenotypes

The whole-transcriptome correlation analysis of the LMC and TCGA melanoma datasets indicated that tumour VDR was strongly correlated with immune and proliferation-associated signalling pathways. In an effort to validate this observation,

VDR expression was compared across molecular melanoma phenotypes. These molecular phenotypes were derived from classification approaches applied to independent melanoma data sets, to identify melanoma subtypes which are both biologically and prognostically significant. The rationale behind this analysis is that: if *VDR* is indeed strongly associated with immune and proliferative pathways, *VDR* expression should vary significantly between the independently-derived molecular phenotypes, which are defined by these characteristics. The melanoma molecular phenotypes used for this analysis are those described in the TCGA melanoma data set (388) and in a Swedish melanoma data set (389). Both studies identified pro-proliferative and immune phenotypes as key melanoma subtypes predicting survival. In 2015, Nsengimana et al described a centroid-based approach to replicate the Swedish subtypes in 300 LMC samples (372), which are a subset of the 703 LMC melanomas analyzed in the current report. In brief, a tumour was assigned to the particular melanoma subtype with which its expression profile has the highest correlation (see Methods 3.3). Using this approach, each of the 700 LMC tumours were classified using the TCGA and Swedish classifications. Tumours which didn't pass the correlation threshold of 0.10 were deemed 'unclassifiable'.

Classification by TCGA subtypes split the 700 LMC tumours into the following groups: Immune (n=192), Keratin (n=247) and *MITF*-low (n=150) with 111 samples unclassifiable. The Leeds melanoma tumours which were classified according to TCGA classification will henceforth be referred to as 'TCGA-subtypes'. Among the TCGA-subtypes in the LMC melanomas, the *MITF*-low subtype had the worst prognosis, being significantly worse compared to the Immune subtype (HR=1.98, P=0.0004) (Figure 3.17A). On the other hand, prognosis of the Keratin subtype did not vary significantly compared to the Immune subtype (P=0.86) (Figure 3.17A). In comparing *VDR* expression among the three TCGA-subtypes, *VDR* expression was significantly higher in the Immune subtype compared to the *MITF*-low subtype ($P=1.07 \times 10^{-6}$) but was not significantly different from the Keratin subtype (P=0.07) (Figure 3.17B).

Classification of the LMC dataset using the Lund classification produced the following subtypes: High-immune (n=173), Normal-like (n=198), Pigmentation (n=222) and Proliferative (n=83) groups, with 24 samples unclassifiable. The Leeds melanoma tumours which were classified according to the Lund classification will henceforth be referred to as 'Lund-subtypes'. Among the Lund-subtypes, the Pigmentation (HR=2.5, $P=3.1 \times 10^{-6}$) and Proliferative (HR=2.62, P=0.0001) subtypes had significantly worse prognosis compared to the High-immune subtype. On the other hand, prognosis of the Normal-like subtype did not vary significantly compared to the High-immune group (P=0.98) (Figure 3.17C). In comparing *VDR* expression across the 4 Lund-subtypes

(Figure 3.17D), *VDR* was significantly higher in the High-immune compared to the Proliferative (P= $7.4x10^{-8}$) and Pigmentation (P=$6x10^{-13}$) subtypes. However, *VDR* expression was also significantly higher in the Normal-like subtype (P=0.0004) compared to the High-Immune subtype. Overall, these data support the view that *VDR* expression is higher in primary melanoma subtypes identified by analysis of TCGA and Lund molecular signatures, which have a better prognosis and have less proliferative, more immune active phenotypes. Notably, these observations are consistent with findings from the whole-transcriptome correlation with *VDR* expression.

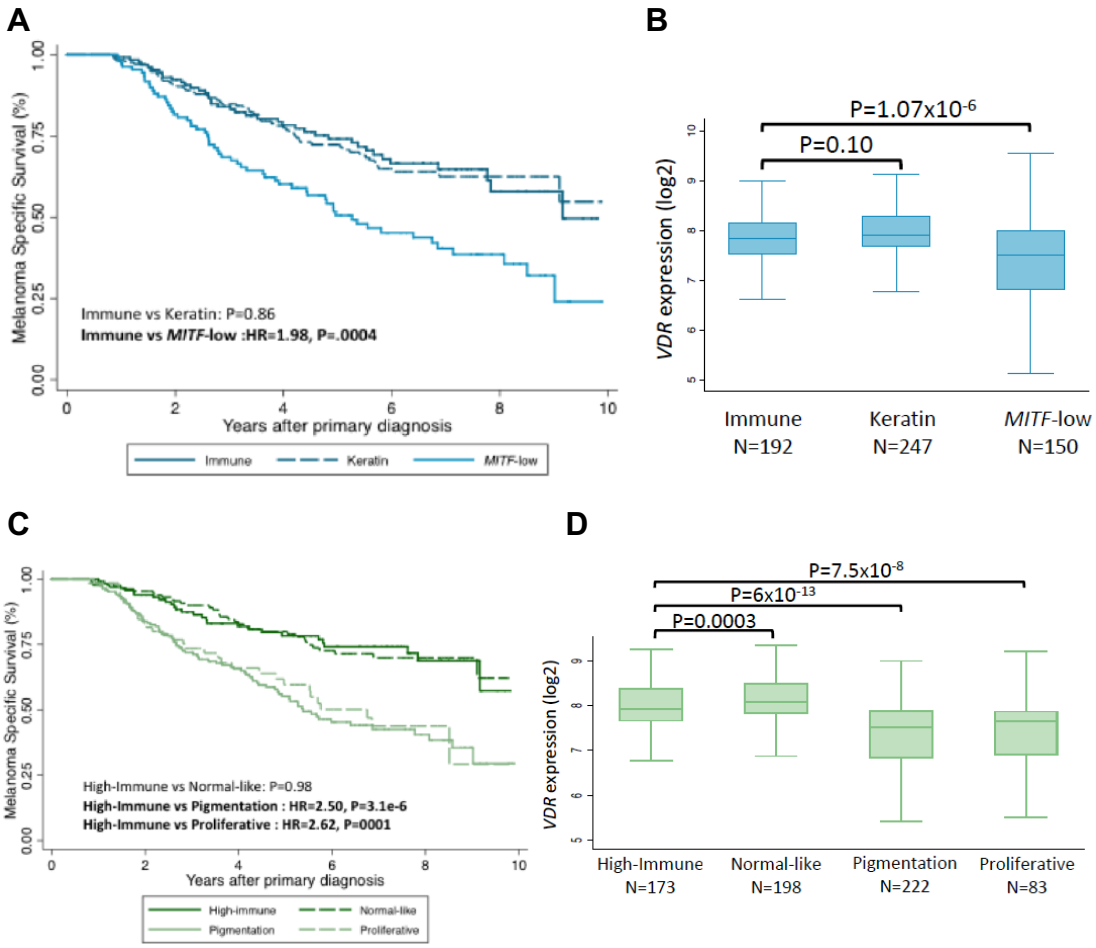


Figure 3.17: Comparison of VDR expression across reported melanoma signatures

A: Melanoma specific survival of participants classified according to the three TCGA-subtypes: Immune (n=192), Keratin (n=247) and *MITF*-low (n=150);

B: Comparison of VDR expression across the three TCGA-subtypes. P-values from T-test;

Figure 3.17 description cont.

C: Melanoma specific survival of participants classified according to the three Lund-subtypes: High-immune (n=173), normal-like (n=198) pigmentation (n=222) and proliferative (n=83);

D: Comparison of VDR expression across the three Lund-subtypes. P-values from T-test

3.4.5.1.2 Imputed immune cells scores

The whole-transcriptome correlation (section 3.4.4.1) and the validation using reported melanoma molecular phenotypes (section 3.4.6.1.1) indicated that *VDR* expression is positively associated with immune-associated pathways and immune-active subtypes respectively. These findings prompted the assessment of the correlation of *VDR* with imputed immune cells scores, with the view of estimating the immune cell types with which *VDR* is most significantly associated. To this effect, the correlation of each of the 26 Angelova imputed immune cells scores (described in section 3.4.3.2) with *VDR* expression was assessed. A linear regression model was used to estimate the significance (denoted by P-value) and direction (denoted by regression coefficient R) of correlation. *VDR* expression correlated positively with 25 of the 26 immune cell scores (at $P<0.05$, Table 3.20) of which the strongest correlation ($R>0.20$, $P<0.05$) was with NK, cytotoxic, dendritic, Th1, activated B, Th2, central memory CD4, T, effector memory CD8 cell and neutrophil scores. Memory B cell scores were the only cell type which had a negative correlation, although the correlation was not strong ($R=-0.06$).

Table 3.20: Correlation of tumour VDR expression with imputed immune cell scores

Immune cell score	Regression Coefficient (R)	P-value
Natural Killer cells	0.27	4.63E-21
Cytotoxic T cells	0.27	5.28E-16
Dendritic Cells	0.25	4.65E-33
Th1	0.24	1.99E-20
Myeloid Derived Suppressor Cells	0.23	8.86E-18
T Follicular Helper cells	0.23	9.15E-13
Activated B cells	0.23	7.56E-15
Th2	0.22	3.71E-17
Central memory CD4	0.22	6.40E-23
T cells	0.22	1.82E-18
Neutrophils	0.20	7.78E-18
Effector memory CD8	0.20	1.48E-14
T Gamma Delta cells	0.19	8.97E-13
T regulatory cells	0.19	1.76E-13
Macrophages	0.18	5.96E-15
Immature Dendritic Cells	0.17	3.22E-12
Central memory CD8	0.17	2.38E-12
NK56 dim	0.17	1.39E-10
NK56 bright	0.16	1.12E-06
Immature B cells	0.15	1.55E-07
Monocytes	0.15	4.29E-08
Mast cells	0.14	1.01E-11
Plasmacytoid Dendritic Cells	0.13	0.0028
Natural Killer T cells	0.12	0.00007323
Activated CD4 cells	0.10	0.0006
Memory B cells	-0.06	0.01170618

3.4.5.2 Histopathological and immunohistochemical validation

The transcriptomic analyses described in the above sections indicate that *VDR* is significantly and positively associated with immune pathways, signatures and imputed immune cell scores. However, it remained to be assessed if this positive association could be validated by independent histopathological measures of immune infiltrate. To this effect, *VDR* expression was compared across pathologist-graded measure of tumour immune infiltration. Of the 703 tumours whose cores were used to generate the transcriptome, 665 tumours were subjected to detailed histological assessment by the pathologist working with our group (Dr. Sally O'Shea) who was blinded to *VDR* status.

Among the histopathological characteristics that were measured, the ones pertaining to immune infiltrate were: whole section immune infiltrate (absent, brisk or non-brisk immune infiltrate) and core immune infiltrate (none, some, moderate or lots of immune infiltrate). Direct comparison of *VDR* expression and immune/lymphocytic infiltrate was thus done for the 665 tumours.

As described in section 3.4.1.3 (Figure 3.5) *VDR* expression was significantly lower in tumours whose whole-section immune infiltrate estimate was 'absent' compared to those with non-brisk (P=0.021) and brisk (P=0.005) immune infiltrate. While the whole-section immune infiltrate indicated the immune infiltration status of the tumour as a whole, it did not necessarily represent the tumour core that was used to generate the transcriptome. Which is why *VDR* expression was compared across measures of immune infiltrate within the tumour core. To this effect, *VDR* expression was significantly lower in tumours with none/barely perceptible immune infiltrate compared to those with 'lots' of immune infiltrate (P=0.035, Figure 3.18, T-test). However, VDR expression did not vary significantly in tumours with 'some' or 'moderate' immune infiltrate compared to those with no immune infiltrate (Figure 3.18). Thus, *VDR* expression was significantly associated with histopathological measures of tumour immune infiltrate, which is concordant with the transcriptome-derived findings described in previous sections.

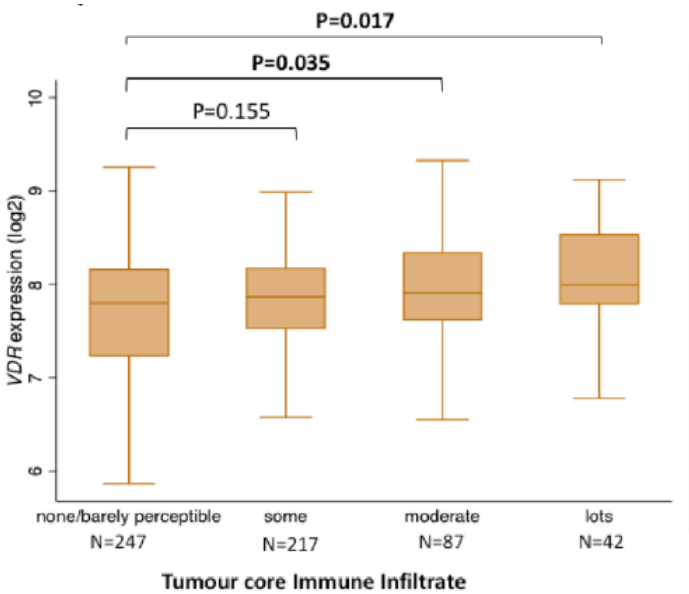


Figure 3.18: Comparison of VDR expression across histopathological measures of immune infiltrate in the tumour core

VDR expression was compared between tumours with 'non/barely perceptible' (n=247), 'some' (n=217) or 'moderate' (n=87) or 'lots' (n=42) immune infiltrate in the tumour core. P-values from linear regression model.

VDR is known to be expressed by immune cells. In the LMC analyses described thus far, *VDR* expression was strongly associated with transcriptomic and histopathological indicators of immune-activity. Increased tumour immune infiltrate is associated with improved melanoma prognosis, as is *VDR* expression in the current LMC dataset. This posed the possibility that the prognostic significance of *VDR* could simply be an attribute of increased immune infiltrate, rather than being a feature of *VDR* expression itself. This issue was addressed using two approaches. First, in the multivariable survival analysis described in section 3.4.2.2, tumour immune infiltrate was one of the variables that was adjusted for. In other words, the prognostic significance of *VDR* expression was persistent even after adjusting for the degree of histopathologically measured tumour immune infiltration. Secondly, I asked if the expression of VDR protein in the LMC primary melanoma tumour sections was localised to the tumour cells or the infiltrating immune cells. To this effect, a subset of the LMC tumours (n=30) were sectioned and stained for VDR expression using Immunohistochemistry (IHC). The stained sections were then evaluated for nuclear and cytoplasmic expression of VDR in both the tumour and the infiltrating immune cells. The evaluation was restricted to the region surrounding the site from which the tumour core was taken, since this was indicative of the tumour from which the transcriptome was generated. Of the 30 tumour sections that were evaluated, 2 had ‘lots’ (>5 cells), whereas 4 had ‘some’ (1-5 cells) TILs that were positive for VDR expression. The other 24 sections had no TILs that stained positive for VDR expression, though the tumour cells themselves stained positive for cytoplasmic and/or nuclear VDR expression (Appendix Table T3-8). Figure 3.19 is comprised of representative images of tumour sections which show nuclear and cytoplasmic positivity for VDR expression in the tumour cells (with corresponding VDR negative sections for reference), with no VDR positivity on TILs in any of the sections. These data suggest therefore that the correlation between immune signals and *VDR* is not simply a reflection of numbers of TILs infiltrating the tumour.

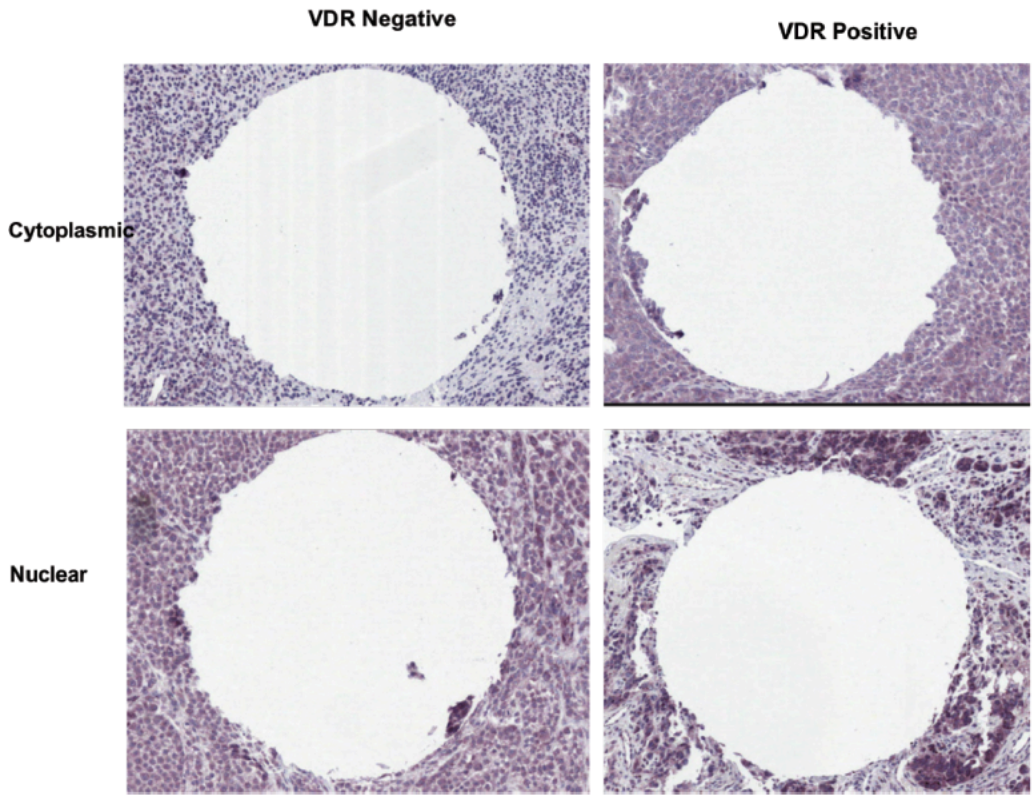


Figure 3.19: Representative images of tumour sections showing nuclear and cytoplasmic positivity for VDR expression in the tumour cells (with corresponding VDR negative sections for reference)

3.4.6 *VDR* expression and Wnt/β-catenin signalling in the LMC primary melanomas

The agnostic whole-transcriptome correlation with *VDR* expression identified the Wnt/β-catenin signalling pathway as one of the most significant negatively correlated pathways in both the LMC primary (section 3.4.4.1) and the TCGA metastatic melanomas (section 3.4.4.2). This finding is of particular interest because VDR has been shown to inhibit Wnt/β-catenin signalling in colon carcinoma cells, with consequent increase in anti-tumour immune infiltrate (356). It is to be noted that colon cancer is the other cancer type (in addition to melanoma) for which most evidence exists for a protective role of vitamin D-VDR signalling on survival. However, the inhibitory effect of *VDR* on Wnt/β-catenin in melanomas has not previously been demonstrated. This highlighted the significance of Wnt/β-catenin being agnostically identified to correlate inversely with *VDR*, thus prompting further investigation to gain evidence for this mechanism.

A previous study from our group by Nsengimana et al reported a classification of the 703 LMC primary melanomas into 6 melanoma subtypes. This classification was based on the transcriptomic immune profile of the melanomas, characterised by immune cell scores. These melanoma subtypes were hence referred to as Consensus Immunome Clusters (CICs). One of the CICs: CIC4 was characterised by high Wnt/β-catenin signalling, reduced immune infiltrate and worse prognosis, compared to 5 other subgroups (316). On comparing *VDR* expression across these 6 previously reported CICs, *VDR* expression was lowest in CIC4: a high-Wnt/β-catenin tumour group compared to the other CICs (Figure 3.20A) which complemented the results from the agnostic correlations with *VDR.* This is also represented in a heatmap of the 6 CICs, depicting the expression of key Wnt//β-catenin signalling along with distribution of the 3 *VDR*-groups (Low-, Intermediate- and High-VDR) (Figure 3.20B). The majority of the low-VDR tumours belonged to CIC4, which also had high expression of genes in the β-catenin signalling pathway *CTNNB1, CMYC, SOX11, SOX2, VEGFA, TCF12, TCF1, APC2* and *EFNB3*.

The observation that low *VDR*-expressing tumours have high expression of Wnt/β-catenin signalling genes in the LMC dataset put forth the possibility that these tumours could have high *CTNNB1* copy number. Since *CTNNB1* codes for one of the key transcription factors controlling the Wnt//β-catenin pathway, increased copies of *CTNNB1* could lead to increased expression of genes in this pathway. So, it was necessary to ascertain that high Wnt//β-catenin signalling in low *VDR*-expressing tumours was attributable to *VDR*-inhibitory effects, rather than a consequence of increased *CTNNB1* copy number. To this effect, it was assessed if *CTNNB1* copy number varied significantly across the three *VDR*-groups. *CTNNB1* copy number (on a continuous scale) did not vary significantly across 3 VDR-groups (Figure 3.20C). On categorising *CTNNB1* copy number, the frequency of tumours with *CTNNB1* deletions (median copy number$\leq$-0.3) and amplifications (median copy number>0.3) did not vary significantly across the 3 *VDR* groups (Fisher's exact P=0.83, Table 3.21). Similarly, the frequency of tumours with *CTNNB1* deletions (median copy number<-0.2) and amplifications (median copy number>0.2) did not vary significantly across the 3 *VDR* groups (Fisher's exact P=0.44, Table 3.22).

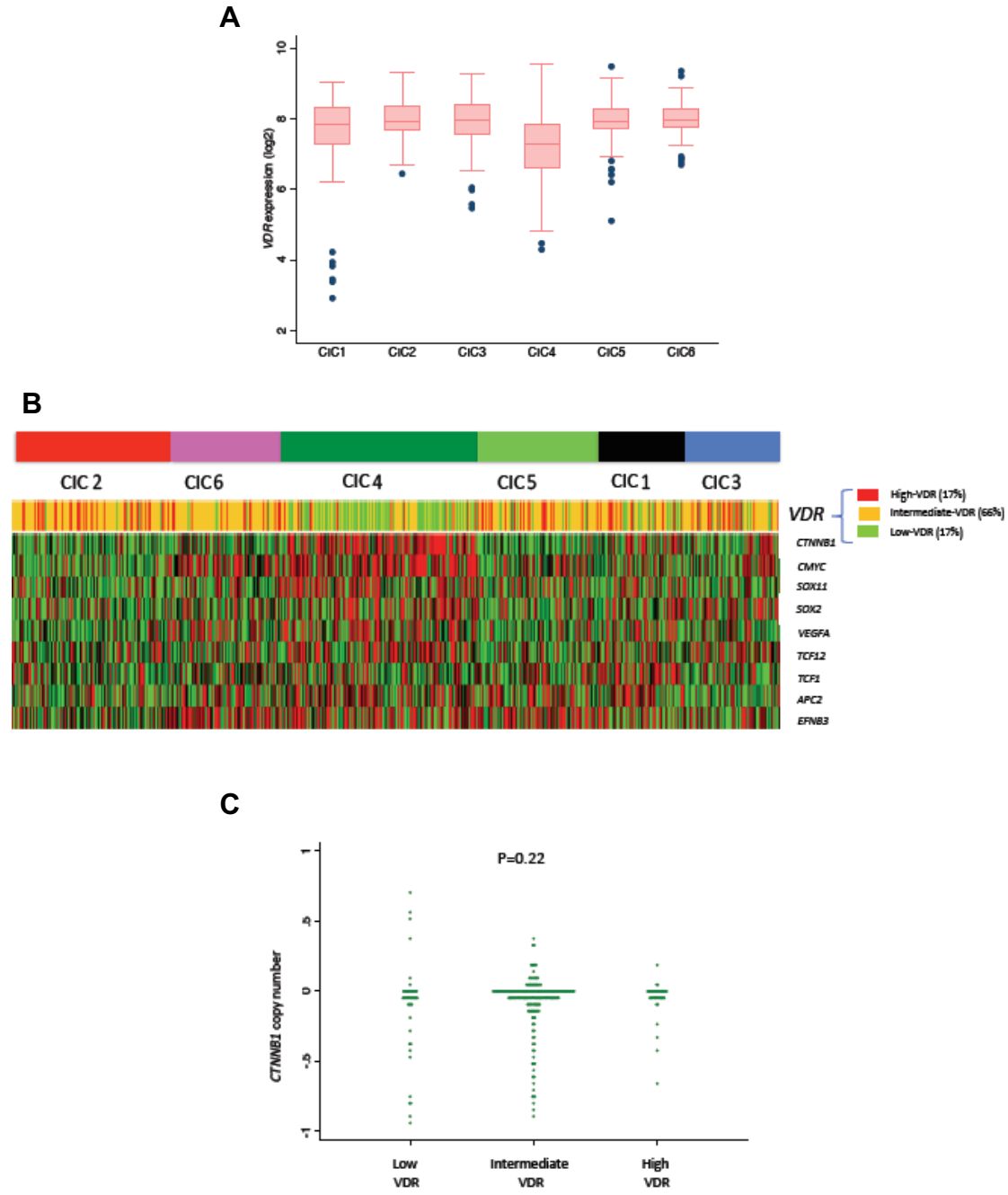


Figure 3.20: VDR expression and Wnt/β-catenin signalling in the LMC primary melanomas

A: Comparison of *VDR* expression across the 6 Consensus Immunome Clusters (CICs) described by Nsengimana et al;

B: Heatmap depicting distribution of the 3 VDR groups along with other key Wnt/β-catenin signalling genes in the 700 LMC primary melanomas;

C: comparison of CTNNB1 CAN across the 3 VDR-groups

Table 3.21: Frequency of tumours with CTNNB1 copy number <-0.3 and >0.3 across the 3 VDR-groups

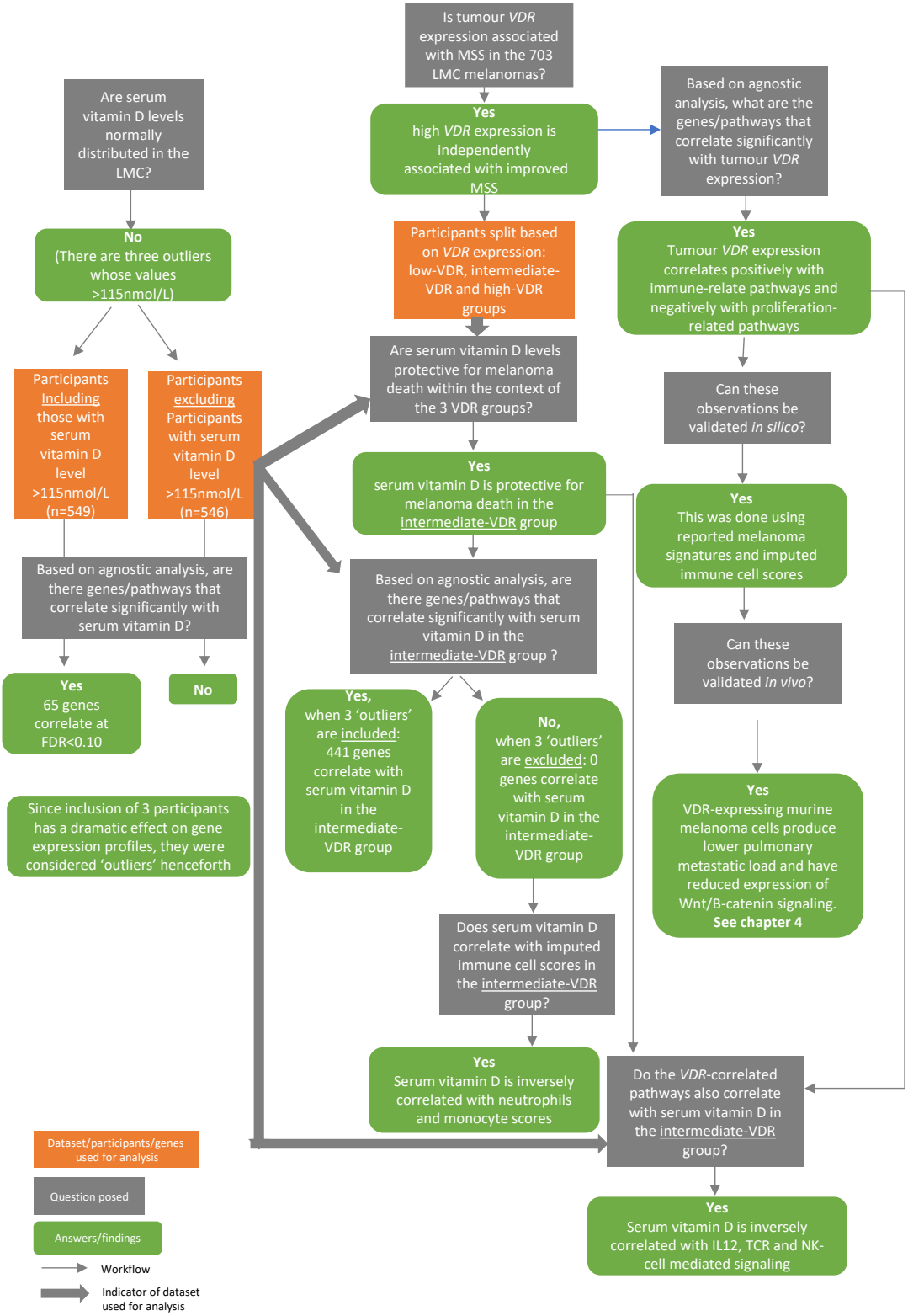
	Low *VDR*	Intermediate *VDR*	High *VDR*
CTNNB1 Del (≤-0.3)	6	13	1
CTNNB1 Amp (≥0.3)	2	5	1

Table 3.22: Frequency of tumours with CTNNB1 copy number <-0.2 and >0.2 across the 3 VDR-groups

	Low *VDR*	Intermediate *VDR*	High *VDR*
CTNNB1 Del (≤-0.2)	9	22	2
CTNNB1 Amp (≥0.2)	5	9	3

3.5 **Graphical summary**

A graphical summary of the key questions and relevant findings addressed in this chapter is presented below



3.6 Discussion

Vitamin D_3 is an essential nutrient important for human health. It is derived from dietary sources and also synthesised in the skin in the presence of sunlight. The role of vitamin D_3 in human health has been the focus of multiple research efforts aiming to understand the effect of this steroidal hormone on various diseases. Since the synthesis of vitamin D_3 is strongly linked to incident UV radiation and the melanin content of the skin, these factors contribute partially to the heterogeneity in global trends of vitamin D levels. Additionally, even though there is agreement on the levels that indicate deficiency (for instance levels that cause rickets), there is lack of consensus about vitamin D insufficiency i.e. as to what levels are causally associated with any number of diseases. This issue is further complicated by factors such as variability in assay differences, seasonal variations, use of single measure vitamin D levels and association of vitamin D with 'active' lifestyle patterns (going out more, leaner, fitter individuals free from illnesses causing systemic inflammation).

The issue of causality is important for melanoma. As melanoma is caused by sunburn in the pale skinned, the usual advice given to patients treated for primary melanoma is to avoid the sun after diagnosis: advice that potentially could worsen outcomes if the result was vitamin D deficiency and vitamin D deficiency is causally related to tumour progression.

While previous studies have assessed the individual effects of either vitamin D treatment or VDR expression, this research study used both tumour *VDR* expression and corresponding serum vitamin D levels at recruitment (on average some 5 months after diagnosis of primary melanoma) to assess their synergistic as well as individual contributions to melanoma survival and their genome-wide effects. This study aimed to systematically identify the clinical and histopathological factors that are significantly associated with both serum vitamin D and tumour *VDR* expression. This was followed by identification of transcriptomic correlates of serum vitamin D and tumour *VDR* expression. An agnostic approach was chosen for these analyses with the view of identifying signalling pathways which are most significantly associated with low serum vitamin D levels or evidence for *VDR* signalling. The pathways/processes which were identified to be the most strongly associated with either, were subject to validation. This included both *in-silico* and *in vivo* validation approaches. While the former reinforced transcriptome-derived correlative evidence, the latter added causal evidence to the findings. Taken together, the aforementioned 'pipeline' enabled a better understanding of the effect of vitamin D-VDR signalling in primary melanomas.

In the LMC primary melanomas, serum Vitamin D itself was not significantly associated with variables such as sex, AJCC stage, mitotic rate and tumour site. Though low serum vitamin D was shown to be associated with poor prognosis tumours in the larger cohort of 2184 melanomas (of which the 703 tumours are a subset) (288), this effect was not significant in the current cohort of 703 tumours. The observation that serum vitamin D tended to be lower in Stage II and III tumours compared to Stage I tumours, albeit not significantly, could be attributed to reduced sample size compared to the original cohort. Serum vitamin D levels did not correlate significantly with anabolic and catabolic enzymes of the Cytochrome P450 superfamily. This could be owing to the fact that the serum vitamin D pertains to a single measure at diagnosis does not necessarily reflect the expression of vitamin D-regulating enzymes in the tumour. The issues surrounding the use of single measure of serum vitamin D are: i) possibility of the participant making lifestyle changes after diagnosis and excision of the melanoma. Examples of lifestyle changes which could lead to variable serum vitamin D levels before and after diagnosis include staying indoors and commence/increase dietary supplement intake and ii) concerns around the reproducibility of the LC-MS-MS-based assay for measuring vitamin D. Taken together, it is to be noted that while the single measure of serum vitamin D has provided useful insights, it remains a relatively 'blunt tool' and requires further optimisation.

Vitamin D signals via its canonical receptor VDR which has been shown to be associated with tumour progression in colon cancer and in melanoma. It was therefore hypothesised that exploration of tumour VDR expression would provide insight into vitamin D-VDR signalling in the LMC melanomas. The LMC tumours were thus stratified based on their VDR expression, to produce 3 *VDR*-groups. This was a survival-based stratification i.e. the cut-off points to establish low-, intermediate- and high-VDR groups were based on the most divergent melanoma survival in these groups. This identified groups that not only varied in *VDR* expression but also had significantly different melanoma survival. The high-VDR group had the best survival (and highest VDR expression) while the intermediate- and low-VDR groups had progressively worse survival. These subgroups proved to be a useful categorisation of *VDR* expression in the LMC and were used throughout the study. When these proportions were applied to the TCGA metastatic melanoma dataset, the resultant *VDR*-groups also had similar survival patterns: high-VDR being best, low and intermediate-VDR significantly worse survival. This suggests that these *VDR* categories are relevant not only in primary but also in metastatic melanomas. The rationale behind applying LMC-derived *VDR* proportions (lowest 17%, middle 66% and highest 17%) to stratify the TCGA metastatic tumours was: since mean *VDR* expression is lower in metastatic tumours compared to primary melanomas, the expression level-based cut-offs derived from primary

melanomas would not be directly applicable to metastatic melanomas. Moreover, the LMC and TCGA transcriptomic datasets were generated on different platforms, adding a further justification to apply the proportions rather than the expression level-based cut-offs.

To address the hypothesis of a possible context-specific effect of vitamin D, a vitamin D-*VDR* subgroup analysis was done. This analysis (which categorised the 700 LMC tumours into 6 subgroups) revealed that serum vitamin D indeed was protective for melanoma survival, but only in the intermediate-VDR group of tumours. In patients with low-VDR expression, the lack of a protective effect of higher vitamin was not unexpected, since low receptor expression could preclude effective signalling despite ligand sufficiency. On the other hand, a lack of benefit in the high-VDR tumours could be a reflection of receptor saturation as reported in other NHR family receptors (390) or potentially a ligand-independent effect of VDR, which has been described in other cancers (348, 349). The context-specific effect observed in the intermediate-VDR group naturally prompted the identification of factors which could contribute to this effect. Intermediate-VDR tumours from patients with low serum were more likely to have higher frequency of AJCC stage II tumours (compared to AJCC stage I), increased frequency of vascular invasion and higher Breslow thickness, all of which could explain why these patients had a worse prognosis. Patients with high serum vitamin D in this group had reduced imputed immune cell scores for neutrophils and monocytes, as well as reduced pathways scores for IL12, TCR and NK cell-mediated signalling (identified by 'candidate approach' of checking *VDR*-correlated pathways). The role of vitamin D in modulating innate immune responses has been previously reported, with particular focus on the vitamin D_3-mediated anti-bacterial innate immune response (391). There is also extensive literature which indicate a protective role for vitamin D on respiratory tract infections such as tuberculosis (392), which are characterised by innate immune responses. While a vast portion of the literature indicate that vitamin D is associated with a pro-innate immune response, the consensus remains equivocal. There are studies which indicate that vitamin D can modulate both pro and anti-innate immune responses, depending on the assaulting antigen and responding immune cell (393). Individual studies have shown Vitamin D deficiency to be inversely associated with Neutrophil-to-Lymphocyte ratio (NLR) (394) and eosinophil counts (395). These studies used NLR and eosinophil counts as indicators of systemic inflammation, suggesting that vitamin D has an anti-systemic inflammation role. Collectively, the observation in the intermediate-VDR group in the LMC of serum vitamin D being inversely correlated with imputed neutrophils and monocytes cell scores could be a reflection of systemic inflammatory differences in this group of tumours. But this hypothesis cannot be tested

owing to lack of data pertaining to indicators of systemic inflammation. Overall, the relationship between vitamin D and immune cell function remains unclear.

However, it was of interest to assess if transcriptomic variations could also explain the difference in prognosis in this group. An agnostic whole-transcriptome correlation analysis did not identify genes that correlate with serum vitamin D in the intermediate-VDR group, after applying multiple correction thresholds. While the self-reported supplement intake data indicated that the patients with high serum vitamin D were more likely to have taken vitamin supplements, this still did not explain the difference in survival. One possible explanation for this context-specific protective effect is systemic factors i.e. the effects of serum vitamin D in this group could be via systemic effects. This cannot be tested in the present study but present a possible question for future studies. The other possible reason which could explain the improved survival in participants with high serum vitamin D in the intermediate-VDR group (compared to low serum vitamin D) could be owing to more effective signalling via VDR in this subset of tumours. Though there is no statistically significant difference in *VDR* expression between these subgroups, it is worth positing that the transcriptomic data are insensitive to subtle changes in expression. Moreover, it is to be noted that the 'season-adjusted' level of serum vitamin D, while accounting for the drop in circulating vitamin D in winter, could perhaps have masked and could explain the lack of correlation between one measure of vitamin D and *VDR* in the intermediate group.

The intermediate-VDR group is also of interest as the 3 patients with serum vitamin D> 115nmol/L all belonged to this group. The whole-transcriptome correlation analysis (discussed above) was performed excluding these 3 samples. The reason being: these levels are beyond the normal distribution of serum vitamin D in this cohort of patients and hence were considered statistical outliers rendering the data set difficult to interpret with confidence. The levels were not in themselves very high when considering world-wide data. However, when the whole-transcriptome correlation analysis (in the intermediate-VDR group) was repeated after including these 3 patients, a significant set of genes crossed the multiple-testing threshold. In analysing these genes, serum vitamin D was positively associated with retinol metabolism, Cytochrome P450-drug metabolism, fatty acid degradation and glycolysis/gluconeogenesis, to name a few. The pathways are composed of the recurrent nodal genes such as *ADH1AC*, *TUBA3*, *UGT1A10*, *DYNC1L1*: genes involved in cellular metabolism. This could imply a reflection of a subset of tumours that have increased metabolic activity. On the other hand, the pathways negatively correlated with serum vitamin D in the intermediate-VDR group include T cell activation, mitochondrial translation, MHC class II antigen presentation, HIF-1 pathway and Renal cell carcinoma. This could indicate that tumours

from high serum vitamin D patients have overall reduced T cell activity and transcriptional activity. However, it is to be noted that these pathways were identified after including 3 patients with >115nmol/L, making these finding inconclusive but worthy of further exploring the effects of 'very high' vitamin D levels on melanoma primaries.

Since high tumour *VDR* expression by itself was strongly associated with improved melanoma survival, the effects of the receptor were explored in further detail. Though serum vitamin D did not correlate significantly with tumour VDR expression, this could be a reflection of the dynamic interactions between circulating vitamin D and tumour expression, being regulated by feedback loops. Moreover, the measures of vitamin D and *VDR* are from a single time point (at diagnosis), which could explain the lack of correlation in the 703 LMC primary melanomas. However, tumour *VDR* expression correlated significantly with clinicopathological features (age at diagnosis, sex, AJCC stage, mitotic rate and tumour site) that by themselves are predictors of melanoma prognosis. So, to ensure that the protective effect of *VDR* is not confounded by these features, a multivariable analysis was done. This revealed that the association of *VDR* with melanoma prognosis was independent of these factors, indicating that the protective effect of *VDR* expression was not simply a reflection of staging and degree of immune infiltrate. Additionally, this protective effect of *VDR* was independent of the expression of other NR1L family genes, despite reports of integrated activity among other nuclear receptors (396).

The finding that *VDR* expression is lower in tumours from habitually covered skin is novel. It is of interest in terms of understanding progression as tumour origin on the back, thorax, upper arm, neck, and scalp has long been recognized as a marker of poor prognosis (377) in melanoma. The underlying biology is however unknown. *VDR* expression has been reported to increase as a result of experimental exposure to ultraviolet light (UVB) in man (397), and the *in vitro* data reported here showed upregulation of *VDR* after melanoma cell lines were treated with vitamin D. These data suggest the possibility that regular sun exposure resulting in increased VDR signalling in early melanoma cells may moderate cancer progression. This is consistent with previous epidemiological data reported by the Leeds group in which the only behavioural measure associated with melanoma risk was actually a protective effect of increased regular sun exposure at home in the north of England during summer months (398). That is, that it is possible that there is a very complex relationship between sun exposure and melanoma risk: that intense intermittent sun exposure associated with sunburn (67) is causal for melanoma but that regular non-burning sun exposure may be protective and that effect may be mediated by VDR signalling. This has considerable implication for public health advice given.

Given the prognostic significance of *VDR* expression in the LMC primaries and TCGA metastases, I sought to explore its mechanistic basis by adopting an agnostic approach using whole genome correlations and stringent false-discovery filters to identify the most significant genomic correlates of *VDR* expression. Broadly speaking, *VDR* expression was strongly positively correlated with immune-related pathways whereas it was negatively correlated with proliferative pathways. Since VDR is also expressed in keratinocytes, it was ensured that the *VDR*-correlated genes were not an artefact of keratinocyte contamination and to this effect observed no significant change in the correlated pathways after adjusting for *FLG2* expression.

Being the primary transcription factor for vitamin D-VDR signalling, VDR has been the focus of ChIP-seq-based approaches to identify transcriptional targets. Tuoresmaki et al performed a combined analysis of VDR ChIP-seq data from 6 cellular models and identified VDR-binding regions, a large proportion of which are tissue type specific. In an effort to gauge the overlap between genes that are transcriptionally controlled by VDR in other tissue types and those that correlate with *VDR* in melanoma primaries. Irrespective of tissue type, about 57% of genes with VDR-binding sites also correlated significantly with *VDR* in our analysis, indicating that a significant proportion of the *VDR*-correlating genes are under direct transcriptional control of the VDR transcription factor.

Given the strength of correlation of *VDR* expression with immune pathways, this was validated by pathologist-graded TILs and indeed found that *VDR* expression was higher in tumours with a brisk immune infiltrate. Melanoma molecular phenotypes have been described which identify biologically relevant tumour subtypes based on their gene expression profiles. By applying these signatures to the LMC tumour transcriptomes, it was observed that *VDR* expression was significantly higher in high-immune compared to proliferative subtypes, even using independently-derived molecular subtypes. The high *VDR* expression in TCGA Keratin subtype and Lund Normal-like subtype could be a reflection of high *VDR* expression in epidermis and keratinocytes. Alternatively, given that these two subtypes have the best prognosis among their counterparts, it could represent high VDR expression in thinner, early stage tumours, which have remnants of the skin/keratinocyte molecular signature.

In assessing the correlation of *VDR* with individual immune cell scores, the aim was to identify specific immune cell scores associated with *VDR* expression. However, almost all immune cell scores correlated significantly, with no indication of any trend towards innate or adaptive immune cells. This observation could be a reflection of the methods of imputed immune cell scoring. In that, all the imputed immune cell scores are highly correlated with each other (Pozniak et al, in press). This could be a possible explanation for the strong positive correlation of VDR with all the immune scores.

Alternatively, it could indicate that *VDR* is part of a broader immune response, but this would require detailed IHC experiments staining for various immune cell types. The collective transcriptome-based evidence for a pro-immune effect of *VDR* is supported by a previous IHC-based study for the association of immune cell infiltration with VDR expression, albeit for a smaller number of tumours (368). This study has furthered the understanding of this pro-immune effect by identifying specific immune pathways and genes associated with *VDR* in melanoma primaries. The reported strong positive correlation of *VDR* with genes involved in pathways such as extracellular matrix organization, TNFα signalling, NFκB activation, IFNγ and IL-12 mediated signalling are novel findings in primary melanoma.

Among the proliferation-related pathways that correlated inversely with *VDR* in both the LMC and TCGA, Wnt/β-catenin signalling was of particular interest since it has previously been reported to be inhibited by vitamin D-VDR signalling in colon cancer cells with consequent increase in anti-tumour immune infiltrate (356). On comparing *VDR* expression across the 6 previously reported CICs (Nsengimana et al), it was observed that *VDR* expression was lowest in the 'high-Wnt/β-catenin' tumours which complements the results from the agnostic correlations with *VDR*. This suggests that some of the effects of VDR signalling in melanoma cells are mediated by inhibition of Wnt/β-catenin signalling as reported for colon cancer cells. It was also demonstrated that low expression of Wnt/β-catenin signalling genes in high-VDR tumours was not simply a consequence of decreased copies of *CTNNB1* which codes for the key transcription factor for Wnt/β-catenin signalling. VDR has been shown to inhibit Wnt/β-catenin signalling by directly binding to the transcription factor β-catenin and by increasing expression of E-cadherin: an inhibitor of Wnt/β-catenin signalling (304). This could also be the case in melanoma, which requires further functional validation.

In an effort to identify factors that could affect *VDR* expression, it was observed that metastatic tumours with high *VDR* expression were more likely to be hypomethylated, suggesting that epigenetic control of VDR expression is active in melanoma as previously reported (399) (400). Though *VDR* copy number was not associated with expression in LMC primaries, distant metastases (which have worse prognosis) had lower median copy number compared to regional lymph node metastases in the TCGA, suggesting a progression-associated genomic loss of *VDR*.

From among the findings reported in this chapter, I performed functional validation to add causal evidence to two particular findings:

i) The strong independent protective effect of VDR expression which is also strongly inversely correlated with proliferation-associated pathways.
ii) The inverse association of VDR with Wnt/β-catenin signalling

These functional validation studies are detailed in Chapter 4

Chapter 4

In-vitro and *in-vivo* validation of transcriptomic evidence

4.1 Introduction

The transcriptomic and histopathological findings described in chapter 3 indicated that *VDR* expression was positively associated with an immune-active environment, while being negatively associated with factors indicative of melanoma progression. These significant correlative evidences provided the basis for functional validation experiments. The hypothesis which was functionally tested was that VDR-expressing melanoma cells would have reduced proliferative potential and lower expression of Wnt/β-catenin signalling genes, compared to cells which don't express VDR. To this effect, an *in vivo* murine metastatic assay was used to assess the effect of VDR-expressing murine melanoma cells on pulmonary metastasis formation. The murine melanoma B16-BL6 cell line was chosen for this experiment owing to its endogenously very low expression of *VDR*, compared to other B16 strains such as B16-F0 and B16-F10 (unpublished data, personal correspondence with and courtesy of Dr. Martin del Castillo and Dr. David Adams, Wellcome Trust Sanger Institute). Briefly, early passage B16-BL6 cells were subjected to transposon-mediated transfection with a vector carrying the murine *VDR* cDNA (referred to as the 'VDR construct'), to generate a cell line with increased VDR expression. These cells will be referred to henceforth as 'VDR: B16-BL6 cells'. As a control, B16-BL6 cells were stably-transfected with only the backbone vector to generate cells expressing only the endogenously low levels of VDR: referred henceforth as 'control: B16-BL6' cells. The resultant clones, from both *VDR* cDNA and control transfections were screened for VDR expression using Western Blot and qRT-PCR. Two clones per cell-type, V1 and V2 (VDR: B16-BL6 cells), C1 and C2 (control: B16BL6 cells) were used in the in the *in vivo* tail-vein metastasis assay. The details of this experiment and the relevant results are discussed below.

4.2 Chapter aims and overview

Chapter aims	Section
To describe the procedure and outcome of steps involved in cloning the VDR construct	4.2.1
To describe the steps involved and outcome of transposon-mediated transfection of B16-BL6 cells	4.2.2
To describe the steps and outcome of the *In-vivo* tail-vein metastasis assay	4.2.3
To validate the transcriptome-derived observation that VDR is inversely correlated with Wnt/B-catenin signalling	4.2.4

Contributions to this chapter

Sathya Muralidhar performed the following work under supervision of Dr David J. Adams (Wellcome Trust Sanger Institute) Dr Louise van der Weyden (Wellcome Trust Sanger Institute) and Prof. Julia Newton-Bishop.

- In vitro propagation and maintenance of B16-BL6 cells
- Cloning of VDR construct, with help from Dr James Hewinson (WTSI)
- Transposon transfection of B16-BL6 cells to generate VDR: B16-BL6 and B16-BL6: control cells
- Western blot screening of transfection clones
- qRT-PCR screening of transfection clones, with help from Dr Mark Harland (UL*)
- Immunohistochemistry of the FFPE blocks of murine lungs generated from the *in-vivo* metastatic tail-vein assay for CD3
- Digital estimation of metastatic area and CD3 positive immune infiltration
- Statistical analysis of raw data from metastatic assay (metastatic load and CD3 positive immune infiltrate) and graphical representation
- Array-based comparison of expression of Wnt/B-catenin genes between VDR: B16-BL6 and control: B16-BL6: cells, with help from Dr. Mark Harland (UL)

Dr Jon Laye performed:

- Sectioning of FFPE blocks of murine lungs lung blocks generated from the *in-vivo* metastatic tail-vein assay
- H&E staining of lung sections.

Dr Louise van der Weyden performed:

- The *in-vivo* metastatic tail-vein assay which included tail-vein injection of early passage transfected cells, sacrifice of mice upon reaching experimental end point and macroscopic counting of pulmonary metastases.

4.3 Methods

4.3.1 Cloning the murine pB-VDR construct

The visualisation of plasmids, primer alignment and analysis of results from Sanger sequencing were done using SnapGene viewer® (401).

The Insert: A pMS-VDRiresPuroKATGx construct (4986bp) with flanking PmeI and NotI restriction sites construct was synthesized by GeneArt™. The construct consisted of 1266bp mouse VDR sequence, an IRES (Internal Ribosomal Entry Site), the Puromycin Resistence gene (Puro) and a Kozak sequence (gccAccatgg). The synthesized vector was digested using PmeI, NotI and EcorV (in NEB buf 2.1 and BSA, 2 hours at 37°C) to produce three fragments of which the 2.4kb fragment (containing the VDR-IRES-Puro) was gel-purified (Wizard SV Gel and PCR cleanup system, Promega) and used as the Insert for subsequent ligation step.

The Backbone: The backbone vector used was: PB-BirA_P2A_rtTA_P2A_PURO-TRE-cherry (11,849bp) which includes the inverted terminal repeat sequences that are flanked once again by PmeI and NotI digestion sites. A PmeI and NotI double digest (in NEB buf 2.1 and BSA, 2h at 37C) and the resulting 6.7kb fragment was gel purified and used as the backbone for subsequent ligation. This backbone vector was also used to generate the control cells which do not express VDR (for comparison with the VDR-expressing cells).

The Ligation: the Insert and the Backbone were ligated at 5:1 and 3:1 using T4 DNA Ligase (70) for 2h at room temperature and transformed into OneShot Top10 chemically competent bacteria (Life Technologies) as per manufacturers' protocol. The Transformed bacteria were cultured in SOC (Super Optimal Broth with Catabolite repression, ThermoFisher) medium and were plated on Ampicillin-Agar plates and incubated at 37C overnight. Two colonies were spotted, both of which were inoculated in LB broth+ Ampicillin (1ul/ml). Both colonies (Colony 1 and Colony 2) were used for plasmid isolation using miniprep (PureYield Plasmid Miniprep system, Promega) per manufacturers' protocol. Resulting plasmids were digestion-checked with SpeI and PmlI (CutSmart buf 10x, BSA, 2hrs at 37C) to produce 3.2kb and 5.9kb fragments. Both

Colony 1 and Colony 2 were positive for digestion checks, but Colony 1 had low concentration so Colony 2 was used for sequencing check. Primers were designed at approximately 700bp intervals, to span the entire 9186bp Cloned plasmid. Colony 2 was sent for sequencing using the 13 primers ('Round 1' primers) and the sequencing results were verified. In addition, 6 primers were designed to span only the VDR-IRES-Puro region ('Round 2' primers) to be absolutely sure about the sequence similarities. Sanger sequencing was done using the Eurofins service at the Wellcome Trust Sanger Institute. See Appendix Table T4-1 for primer sequences and related information. The ligated PB-BirA_P2A_rtTA_P2A_VDR_IRES_Puro construct will be referred to as 'pB-VDR' henceforth.

4.3.2 Transfection of B16-BL6 cells

B16-BL6 stock vials were revived in complete DMEM (DMEM, 10% FBS, 1% PenStrepGlutamine) and early passage cells were plated in 6-well plates to reach 90% confluence at which time they were used for transfection. Lipofectamine^{TM} 2000 transfection reagent (11668027, Thermo Fisher) was used as per manufacturer's protocol (10μl Lipofectamine^{TM} 2000 + 240μl OptiMem per 6-well). Briefly, the 'transposon construct' (either pB-VDR or the backbone vector) and Transposase plasmid (which codes for expression of the enzyme transposase) were mixed in 1:2 ratio and made up to 1500μl using Opti-MEM (Thermo Fisher). This was combined with Lipofectamine 2000 mixture (10μl Lipo2000 + 240μl Opti-MEM per 6-well), incubated for 20 minutes at room temperature and added to the 90% confluent cells. 24 hours post-transfection, the cells were treated with 5μg/ml puromycin. 48hours post-transfection, medium was replaced with 5μg/ml puromycin in high serum DMEM (20% FBS). 13 days post-antibiotic treatment, all wells were observed to have cell death but with surviving colonies which were sizeable and conducive for colony-picking. Individual colonies were picked (after trypsinisation) using pipette, seeded into 12-well plates and supplemented with puromycin-DMEM. Transfected colonies were always maintained in 5μg/ml Puromycin-DMEM medium (with 10% FBS and 1% Pen-Strep-Glutamine). Each colony was expanded onto 6-well plates and then 10cm dishes. Upon reaching confluence in 10cm dish, 1/3 of cells were frozen down whereas 2/3 of cells were used for western blot screening.

4.3.3 Western blot screening

Cells were trypsinised, PBS-washed, lysed (RIPA buffer for 20 mins in cold room, with agitation) and sonicated (30 second cycles at max and centrifuged at 13000 rpm for 30 minutes to spin down the debris. The clear lysate was collected, combined with loading

dye (1X NuPAGE LDS Sample buffer NP0007) and denatured at 95C for 5 minutes. Denatured samples were loaded into protein gel (NuPAGE 4-12% Bis-Tris Protein gels NP0321) which were used for electrophoresis using the X-cell SureLock Mini-Cell system with 1X NuPAGE MOPS SDS Running buffer (ThermoFisher), 0.05% NuPAGE antioxidant (ThermoFisher) and run at 50V 1hr and then 100V 15mins. Protein transfer was done using the X-cell SureLock system using 1X NuPAGE transfer buffer (ThermoFisher), 0.05% antioxidant (ThermoFisher), 10% MeOH onto to a nylon membrane (Amersham Hybond XL) for 2hrs at 30V (room temperature). The membrane was then washed with Tris-buffered saline, 0.1% Tween 20 (TBST) followed by blocking with 5% skim milk solution (in TBST) for 1hour at room temperature. This was followed by washing (3x wash in TBST on rocker) and incubation with primary antibody (antibody info in table below) in 5% milk solution (in TBST) at 4^{o}C on a rocker overnight. This was followed by washing (3x wash in TBST on rocker) and then secondary antibody incubation for 1 hour at room temperature. Proteins on blot were visualised using the Western Bright ECL HRP-conjugate Spray (Advansta, K-12049-D50) on the Image Quant LAS4000 (exposure time: 10sec).

4.3.4 Quantification of VDR expression using qRT-PCR

Of the transfected clones screened by Western blot for VDR expression, only 4 transfected clones were assessed for murine *VDR* expression and used for subsequent assays. These were: the VDR-expressing clones V1 and V2, and the control clones C1 and C2.

Murine *VDR* expression was measured using the TaqMan Assay for murine *VDR*: Mm00437297_m1 (ThermoFisher, catalogue number 4331182). The assay spans the boundaries of exon 3 and 4, with an amplicon length of 95bp. As a housekeeping control, murine *GAPDH*: Mm99999915_g1 (ThermoFisher, catalogue number 4453320) was used. The TaqMan™ Gene Expression Master Mix (Thermo fisher) was used for both reactions. The qRT-PCR cycles were programmed and executed in the QS5 system (Thermo Fisher, University of Leeds core facility) with the following parameters:

Block type: 96 -well.0.2ml block

Passive reference dye: ROX

Experiment type: comparative C_t (Δ-ΔC_t) method

Amplification cycle: Step 1: 50^{o}C, 2 minutes

Step 2: 95^{o}C, 10 minutes

Step 3: 50°C, 1 minute → Record C_t

Step 4: 95°C, 15 seconds

Step 5: 95°C, 15 seconds

Step 6: 60°C, 1 minute

The results at the end of the PCR run were exported as a Microsoft excel (.xls) file which listed the C_t values from each well (one sample per well). Relative expression was calculated using the Δ-ΔC_t method, normalized to average C_t of the housekeeping gene (GAPDH). Fold change (FC) of the *VDR: B16-BL6* clones (V1 and V2) and the control: B16-BL6 clone C2 were calculated relative to the clone C1, which was chosen as the 'baseline'. The Fold Change was calculated as follows: FC = 2^(-$\Delta\Delta$Ct) where $\Delta\Delta Ct = \Delta Ct - \Delta Ct_{C1}$.

4.3.5 In-vivo tail-vein metastasis assay

The care and use of all mice in this study were in accordance with the UK Animals in Science Regulation Unit's Code of Practice for the Housing and Care of Animals Bred, Supplied or Used for Scientific Purposes, the Animals (Scientific Procedures) Act 1986 Amendment Regulations 2012, and all procedures were performed under a UK Home Office Project license, which was reviewed and approved by the Sanger Institute's Animal Welfare and Ethical Review Body. Housing and husbandry conditions were as described previously [van der Weyden *et al*., 2017] and mice were maintained on Mouse Breeders Diet (Laboratory Diets, 5021-3) throughout the study. V1, V2, C1 and C2 cells (detailed above) were tail vein administered to 6-10 weeks old sex-matched wildtype (C57BL/6NTac) mice (10^4 cells in 0.1mL PBS). After 21 days, mice were humanely sacrificed and their lungs macroscopically examined to determine the number of metastatic deposits in all 5 lobes ('met count'). Lungs were formalin fixed, processed and embedded in paraffin wax blocks, from which consecutive 5μm sections were cut and used for H&E and CD3 staining.

4.3.5.1 Estimation of metastatic area

H&E (Chapter 2: methods for H&E protocol) sections were digitally scanned (Leica Aperio AT2) and metastatic area (in μm^2) was digitally estimated using Aperio Imagescope (Leica Biosystems) software. Total area ('met area') was calculated as the sum of area of all metastatic deposits, across all 5 lung lobes. Statistical significance in metastatic load from mice injected with VDR: B16-BL6 cells (V1 and V2 combined into a group) and control: B16-BL6 cells (C1 and C2 combined into a group) was determined using a Mann-Whitney U-test (Stata command: 'ranksum').

4.1.1.1 CD3 Staining and counting strategy

5μm sections were stained for mouse CD3 (see below for protocol). The CD3-stained sections were digitally scanned as described above. The number of peri-tumoural (within two-cell distance form tumour margin) and intra-tumoural (within the tumour margin) CD3 positive infiltrating lymphocytes were digitally estimated (at 20x magnification) in all 5 pulmonary lobes using the Aperio Imagescope (Leica Biosystems) software. The total number of CD3 positive cells thus estimated was divided by the total tumour area (described above) per lung, to calculate the number of CD3 positive cells per 100 mm^2 tumour. Statistical significance in tumour-infiltrating CD3 positive lymphocytes from mice injected with VDR: B16-BL6 cells (V1 and V2 combined into a group) and control: B16-BL6 cells (C1 and C2 combined into a group) was determined using a Mann-Whitney U-test (Stata command: 'ranksum').

4.3.5.1.1 Protocol for immunohistochemistry of FFPE lung sections for CD3 expression:

1) The FFPE blocks were sectioned using a microtome to produce 5μm sections
2) The tissue sections were deparaffinised on a hotplate at 70°C followed by antigen retrieval in a pressure cooker, with the slides submerged in 1x antigen retrieval solution (Menapath technologies). After antigen retrieval, the slides were washed in wash buffer (1x, Menapath technologies) followed by rinsing in running tap water.
3) The slides were then marked around the tissue region of interest, with a wax marker. Slides were kept in a humidity chamber after this step to ensure tissue hydration.
4) The tissue was then treated with 100μl peroxide blocking solution (Menapath technologies) for 11 minutes. The solution was then washed away with wash buffer (1x, Menapath technologies)
5) The tissue was then treated with 100μl Casein blocking solution (Menapath technologies) for 10 minutes
6) The tissues were then treated with the primary antibody: anti CD3 (ab5690, Abcam) diluted in antibody-diluent solution at 1:100 dilution, as per the antibody manufacturer's recommendation (Menapath technologies). The primary antibody treatment was allowed to incubate at room temperature for 10 minutes.
7) This was followed by secondary antibody treatment using ImmPRESS HRP reagents (MP-7452, Vector Laboratories) and visualised using purple Vector VIP substrate (SK-4600, Vector Laboratories). Slides were counterstained using H&E (Chapter 2: methods for H&E protocol).

8) The sections were left to air dry overnight, after which they were digitally scanned.

4.3.6 Difference in β-catenin signalling genes between VDR: B16-BL6 and control:B16BL6 cells

RNA was extracted from V1, V2, C1 and C2 cells (RNeasy Mini kit, Qiagen). 2.5μg was used for cDNA synthesis (SuperScript™ VILO™ cDNA Synthesis Kit, Thermo Fisher). The 4 clones were analysed in triplicate on a pathway-specific RT-PCR array of 84 mouse Wnt/β-catenin pathway genes (RT² Profiler™ PCR Array #330231 and RT² SYBR Green ROX qPCR Mastermix, Qiagen) as per manufacturer's protocol. The qRT-PCR cycles were programmed and executed in the QS5 system (Thermo Fisher, University of Leeds core facility) with the following parameters:

Block type: 96 -well.0.2ml block

Passive reference dye: ROX

Experiment type: comparative C_t (Δ-ΔC_t) method

Amplification cycle: Step 1: 50°C, 2 minutes
Step 2: 95°C, 10 minutes
Step 3: 50°C, 1 minute → Record C_t
Step 4: 95°C, 15 seconds
Step 5: 95°C, 15 seconds
Step 6: 60°C, 1 minute

Relative expression was calculated using the Delta-Delta CT method, normalized to average Ct of the 5 housekeeping genes provided in the array. Fold change (FC) of the VDR: B16-BL6 clones relative to control:B16BL6 clones was calculated as follows:

$$FC_{V1(or)V2} = 2^{(-\Delta\Delta Ct)_{V1(or)V2}} \text{ where } \Delta\Delta Ct_{V1(or)V2} = \Delta Ct_{V1(or)V2} - \Delta Ct_{avg(C1\ \&\ C2)}.$$

4.4 Results

4.4.1 Cloning of the VDR construct

A transposon-mediated transfection approach was used for to generate the VDR: B16-BL6 and Control: B16-BL6 cells. This approach is based on the concept of transposable repeat elements (also known as Inverted Terminal Repeat sequences- ITRs), which enable genes of interest (situated between two ITRs) to be stably integrated into target genomes. This 'cut and paste' mechanism requires transposase: an enzyme which targets and excises the transposon elements (the ITRs) along with the genes of interest

in-between the repeats. This system allows for insertion of multiple copies of the gene(s) of interest into random TTAA sites within the target genome (402). The transposon-transposase system used in this study is the *piggyBac* (PB) transposon and corresponding PB-transposase. The steps involved in this process are depicted in Figure 4.1. A vector containing the transposon element (PB-BirA-P2A-rtTA-P2A-Puro-Tre-Cherry) was used as the backbone vector, within which the gene of interest had to be inserted. The gene of interest in this case was the murine *VDR* sequence, which was synthesised commercially (GeneArt, see methods) in an 'insert' plasmid (pMS-*VDR*-Ires-PuroRKATGx). Both the backbone vector and insert plasmids were double digested at the same restriction digestion sites (PmeI and NotI) to excise the requisite sequences: the transposon repeats (6.7kb) from the backbone vector and the *VDR*-IRES-PuroR (2.9kb) sequences from the insert plasmid. The excised sequences were ligated to produce the 'VDR-construct' (9.2kb) (2.9+6.7=9.2kb). This construct was Sanger sequenced using 13 primers that were designed to span the length of the construct. Sanger sequencing produced short reads which upon alignment revealed that no mismatches were present in any of the important regions i.e *VDR*, PuroR, CAAGS promoter and IRES. The sequence reads from the 13 primers aligned to the Cloned construct is represented in Figure 4.2. This meant that the VDR-construct was suitable for subsequent co-transfection into the B16-BL6 cells. In addition, the transposase-expressing plasmid was revived from glycerol stock and plasmid isolation was done to be used for subsequent co-transfection.

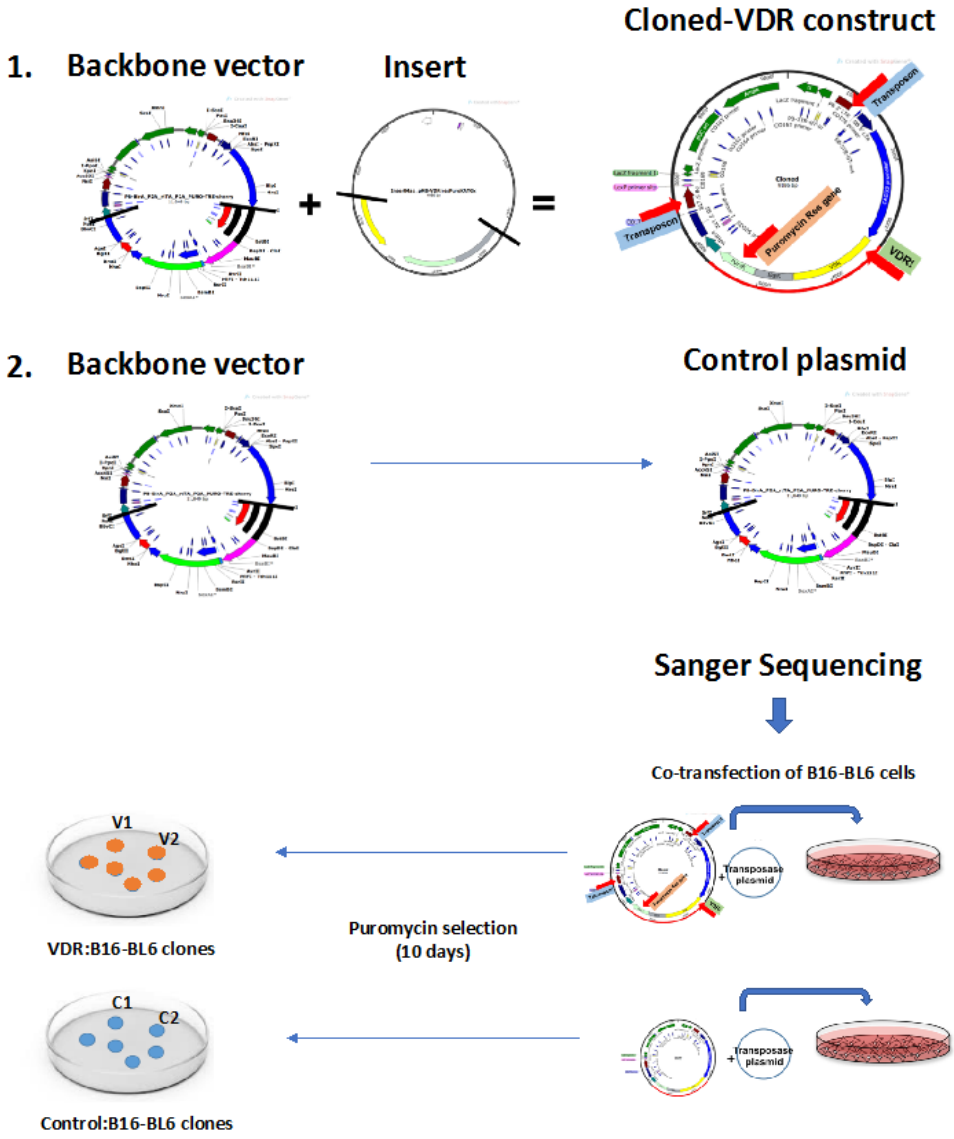


Figure 4.1: Schematic of cloning and transfection strategy used to generate VDR: B16-BL6 and control: B16-BL6 cells

The backbone vector and the insert or only the backbone vector was used to generate the VDR-construct and the control plasmid respectively. While the former includes the VDR cDNA insert, the latter does not contain this, hence making it a null VDR-expressing control. The plasmids were then checked for mutations/aberrations using Sanger sequencing. The VDR-construct and the control plasmid were then co-transfected with the vector which expresses the transposase enzyme ('transposase' plasmid) into B16-BL6 cells. The former transfection produced VDR: B16-BL6 cells while the latter produced control: B16-BL6 cells after puromycin selection. Of the clones that survived puromycin selection and screened for VDR expression using Western blot, only two clones per group (V1 and V2, C1 and C2) were used for subsequent assay.

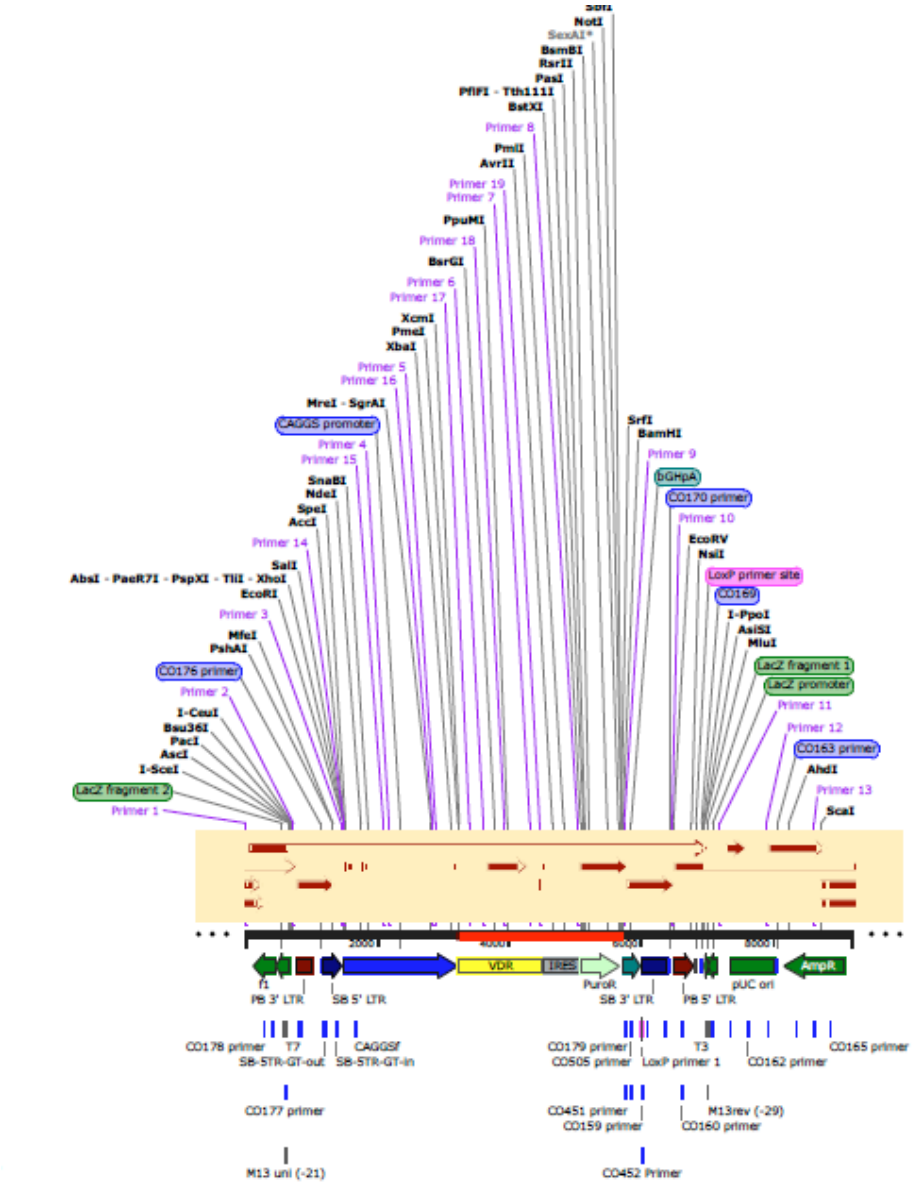


Figure 4.2: The thirteen primers used for Sanger sequencing of the VDR-construct

The VDR-construct was verified by Sanger sequencing using 13 primers which were designed to periodically span the entire length of the VDR-construct. Graphical representation was generated using SnapGene viewer®

4.4.2 Co-transfection of B16-BL6 cells with VDR construct and transposase plasmid

B16-BL6 cells were grown to 60% confluency in a 24-well plate and were co-transfected with the transposase construct along with either the VDR-construct or the backbone vector (the steps involved in this process are depicted in Figure 4.1. Two separate

transfection experiments were done. All transfections in an experiment were set up at the same instance (day and time).

Experiment 1: The aim was to identify the 'transposon to transposase' ratio which would result in optimum integration. Two ratios were tested: i) 0.5ug pB-VDR or backbone vector + 0.2ug transposase vector and ii) 1ug pB-VDR or backbone vector + 0.5ug transposase vector. B16-BL6 cells were seeded at density of 10^6 cells per well and grown to confluency, prior to transfection. This was performed in two 6-well plates, the plating schema is described in Figure 4.3.

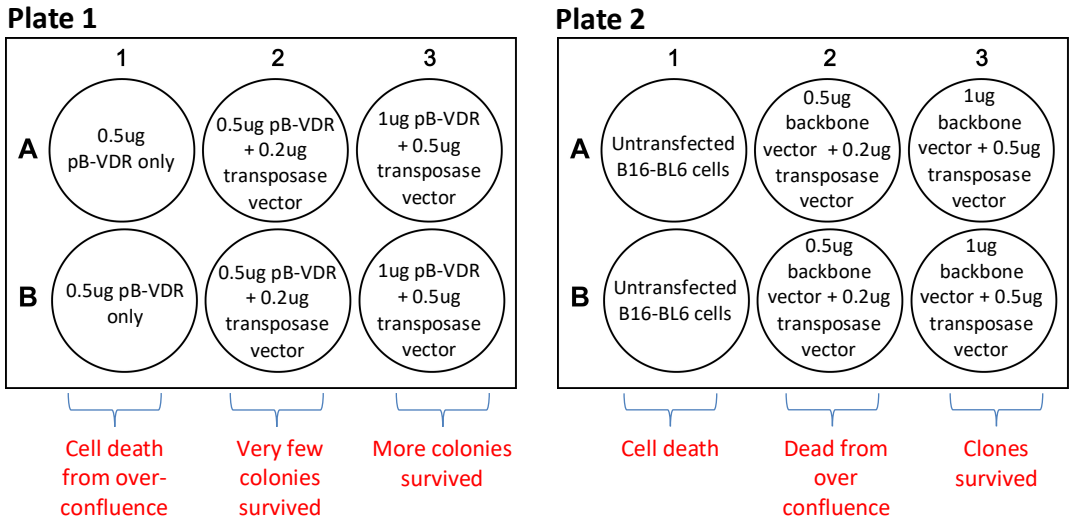


Figure 4.3: Plating schema of Experiment 1: to identify the optimal ratio of 'transposon to transposase' plasmids

Plate 1 and plate 2 are the plating schemas for two separate 6-well plates which were plated on the same instance. Labels in red indicate the state of the cells, 10 days after puromycin selection. Duplicates were included for every transfection, i.e. two wells per transfection.

After 10 days of puromycin selection, the following observations were made. The observations are representative of duplicate wells.

- The untransfected cells were B16-BL6 cells taken from the pool of B16-BL6 cells plated, to ensure they were a 'healthy' pool. The cells looked viable and attained confluence in 2-3 days, as was expected. However, after puromycin treatment, the cells underwent cell death.
- The cells transfected with only the pB-VDR cells were over-confluent and underwent eventual cell death after 10 days of puromycin treatment. The reason being: without the transposase vector (which is required for successful integration

of the VDR cDNA into the host genome), the B16-BL6 cells can only uptake the pB-VDR plasmid and thus express 'baseline' levels of VDR. Since this level is not sufficient to inhibit cell growth, the cells became over confluent and underwent cell death eventually

- Of the two ratios used, the 2:1 ratio (1ug pB-VDR or backbone vector + 0.5ug transposase vector) was the one that produced a viable number of colonies in the case of both pB-VDR and the backbone vector transfections. Hence, this ratio was used for the subsequent Experiment 2.

Experiment 2: The aim was to produce viable colonies of clones which survive the puromycin selection and hence should have successfully integrated the pB-VDR or backbone vector. The transfection schema for this experiment is described in the Figure 4.4.

Experiment 2

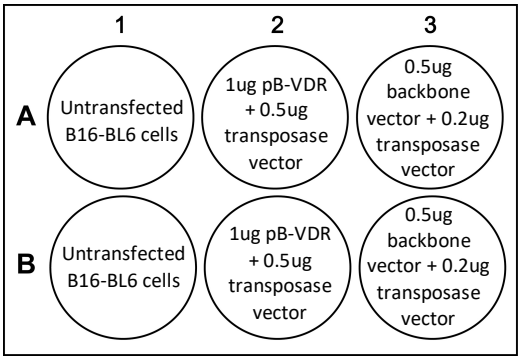


Figure 4.4: Plating schema of Experiment 2: to produce viable colonies of clones which survive puromycin selection and hence should have successfully integrated the pB-VDR or backbone vector

After 10 days of puromycin selection, the following observations were made. The observations are representative of duplicate wells.

- The untransfected cells were B16-BL6 cells taken from the pool of B16-BL6 cells plated, to ensure they were a 'healthy' pool. The cells looked viable and attained confluence in 2-3 days, as was expected. However, after puromycin treatment, the cells underwent cell death.
- B16-BL6 cells which had assimilated plasmids containing the puromycin resistance gene survived and formed colonies. In other words, the cells which survived and formed colonies were the ones which had been transfected with

either the pB-VDR plasmid (in wells A2, B2) or the backbone vector (in wells A3, B3).

Thus, the cell colonies which survived puromycin treatment were picked and each colony was individually seeded into single wells in a 24-well plate. Total of 10 colonies were picked from the wells which were transfected with the VDR-construct plasmid, while 8 colonies were picked from the wells which were transfected with the backbone vector. Upon reaching confluence, the cells were expanded successively in 12-well, 6-well, 10cm plates and finally in 15cm plates. When the cells reached 90% confluency in 15cm plates, the cells were screened for VDR expression using Western blot. All 10 colonies transfected with the VDR-construct plasmid produced a 48.5 kDa band when exposed to anti-VDR antibody, whereas none of the 8 colonies transfected with the backbone vector produced a band (Figure 4.5A). Of these screened colonies, 2 per group were selected for subsequent *in vivo* experiments. Of the colonies transfected with the VDR-construct plasmid, V1 and V2 (VDR: B16-BL6 henceforth) were chosen for subsequent *in vivo* experiments. Of the colonies transfected with the backbone vector, C1 and C2 (control: B16-BL6 henceforth) were chosen for subsequent in vivo experiments. Western Blot was repeated for these 4 clones (to replicate observations from the first Western Blot screen) and confirmed that the V1 and V2 clones expressed VDR, whereas C1 and C2 had no detectable VDR expression (Figure 4.5B). Additionally, qRT-PCR to assess differential expression of *VDR* revealed that V1 and V2 cells expressed significantly higher levels of *VDR*, relative to C1 and C2 cells (Figure 4.5C), with V1 expressing more *VDR* compared to V2. Thus, both Western blot and qRT-PCR screening indicated that the V1 and V2 cells expressed VDR whereas the C1 and C2 cells had little to no VDR expression.

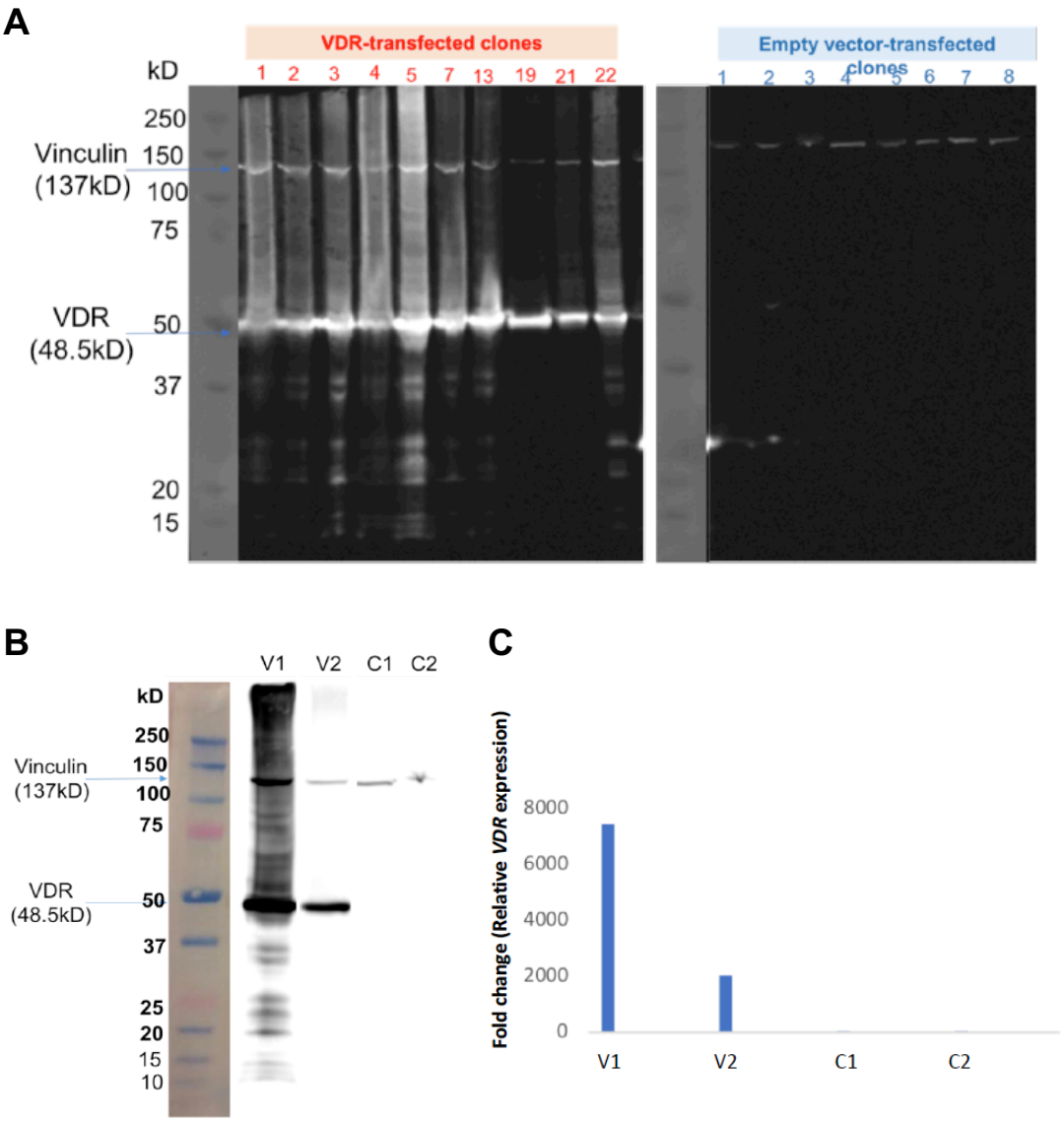


Figure 4.5: Screening of transfected clones using Western blot and qRT-PCR

A: Western Blot of 10 VDR: B16-BL6 clones and 8 control: B16-BL6 clones for VDR (48.5 kDa) relative to housekeeping protein Vinculin (137 kDa);

B: Western Blot of the 4 selected clones: V1 and V2 (VDR: B16-BL6) and C1 and C2 (control: B16-BL6) for VDR (48.5 kDa) relative to housekeeping protein Vinculin (137 kDa);

C: qRT-PCR of the 4 selected clones: V1 and V2 (VDR: B16-BL6) and C1 and C2 (control: B16-BL6) relative to the housekeeping gene GAPDH. Relative expression of VDR among the 4 clones was estimated using the Δ-ΔCt method to calculate Fold Change.

4.4.3 In vivo tail-vein metastasis assay

4.4.3.1 Estimation of metastatic load

4.4.3.1.1 Choice of optimal dosage and duration of tail-vein assay

The VDR: B16-BL6 (V1 and V2) and control: B16-BL6 (C1 and C2) cells were used in an *in vivo* experimental metastasis assay (403, 404). The optimal choice of cell dosage for tail-vein injection and experimental duration (number of days between tail-vein injection and culling of the mouse) was made based on an experiment where multiple cell dosages and time points were evaluated in order to assess the ideal choice. Both VDR: B16-BL6 (clone V1) and control: B16-BL6 (clone C1) were used for this experiment, whose outcome is described in Table 4.1 and Table 4.2 respectively. The mice injected with a specific cell dosage (10^4, $2.5x10^4$, $5x10^4$ or $7.5x10^4$ cells) were periodically assessed for signs of ill-health and humanely sacrificed if found to be so. In cases where the mice reached the specified end-point without any signs of ill-health, the macroscopic count of pulmonary metastases was performed by Dr Louise van der Weyden.

In the following cases, the cell dosage and duration were considered to be a non-viable option:

- If the mouse (during regular inspection), was found to suffer from ill health, the cell dosage was not considered as a viable experimental option. Such cases are indicated by 'x' in Table 4.1 and Table 4.2. This was the case in mice injected with VDR: B16-BL6 cells at 25 days after injection of 10^4, $2.5x10^4$ or $5x10^4$ cells. This was also the case in mice injected with control: B16-BL6 cells at 25 days after injection of 10^4, $2.5x10^4$, $5x10^4$ or $7.5x10^4$cells.
- If a mouse reached the specified end-point without any signs of ill-health, then the mouse was humanely sacrificed and the estimate of macroscopic count of pulmonary metastases was too 'diffuse' (too many metastases to count accurately), then that particular dosage/duration was also considered a non-viable viable experimental option, since it does not permit accurate estimation of macroscopic metastatic load. Such cases are indicated by 'too many to count' in Table 4.1 and Table 4.2. This was the case in mice injected with the VDR: B16-BL6 cells at 25 days after injection $7.5x10^4$ cells and in mice injected with the control: B16-BL6 cells at 21 days after injection with $5x10^4$ or $7.5x10^4$ cells

The mice which reached their respective end points with no signs of ill-health were humanely sacrificed and contain pulmonary metastases that could be counted by macroscopic evaluation (upon estimation of macroscopic metastases) were considered

viable options. In such cases, the number of macroscopic metastases (counted by Dr Louise van der Weyden) are indicated in Table 4.1 and Table 4.2.

The main aim of this experiment was to identify the ideal dosage and duration, which would be a viable option in both groups of mice: those injected with VDR: B16-BL6 and those injected with control: B16-BL6. To this effect, the following dosage and duration were considered viable options:

i. 14 days after injection with 10^4, $2.5x10^4$, $5x10^4$ or $7.5x10^4$cells
ii. 21 days after injection with 10^4, $2.5x10^4$, $5x10^4$ or $7.5x10^4$cells

Since a single dosage and a single duration had to be chose, I (after discussion with Dr Louise van der Weyden, Dr David Adams and Prof. Julia Newton-Bishop) decided to use the dosage of 10^4 cells and an experimental duration of 21 days for subsequent experiments.

Even though this experiment was done as an 'initial pilot' (with one mouse per dosage/duration group) to estimate ideal conditions for the subsequent 'main' experiments, it is worth noting that a mouse in a particular dosage/duration group had lower metastatic counts when injected with VDR: B16-BL6, compared to control: B16-BL6. This was one of the initial suggestions of the possible effect of VDR expression on metastatic potential, which was explored formally in a larger cohort of mice.

Table 4.1: Experiment to assess the optimal cell dosage of VDR: B16-BL6 cells and duration for the tail vein metastasis assay

Number of cell injected refers to the total number of cells used in tail-vein injection. D14, D21, D25 and D26 refer to the time elapsed (number of days) between tail-vein injection and culling of the mouse.

Number of cells injected (dosage)	D14	D21	D25	D26
10^4 cells	5	20	x	x
2.5 x 10^4 cells	34	52	x	x
5 x 10^4 cells	12	110	x	x
7.5 x 10^4 cells	274	220	Too many to count	x

Table 4.2: Experiment to assess the optimal cell dosage of control: B16-BL6 cells and duration for the tail vein metastasis assay

Number of cell injected refers to the total number of cells used in tail-vein injection. D14, D21, D25 and D26 refer to the time elapsed (number of days) between tail-vein injection and culling of the mouse.

Number of cells injected (dosage)	D14	D21	D25	D26
10^4 cells	31	50	x	x
2.5 x 10^4 cells	85	115	x	x
5 x 10^4 cells	191	Too many to count	x	x
7.5 x 10^4 cells	490	Too many to count	x	x

4.4.3.1.2 Performing the tail vein metastasis assay using the optimal dosage and experimental duration

Having identified the optimal dosage for injection and experimental duration, two independent experiments were performed using separate two mouse cohorts, wherein wild type mice were injected with 10^4 V1, V2, C1 or C2 cells and pulmonary metastatic load was estimated after 21 days. Pulmonary metastatic load was estimated both as macroscopic counts of surface lesions (referred to as met-counts henceforth) and as metastatic area estimated using digitally scanned FFPE sections (met-area henceforth, measured in μm^2).

In the first experiment (Experiment 1) consisting of a cohort of 34 mice (Table 4.3), mice injected with either the VDR: B16-BL6 clones: V1 and V2, produced significantly fewer pulmonary metastases compared to the control: B16-BL6 clones C1 and C2 (Figure 4.6A). This was the case when using both microscopic met-area (P=0.04) or macroscopic met-count (P=0.0006). Since Experiment 1 revealed a significant difference in metastatic load between the VDR-expressing and control cells, the experiment was repeated in order to replicate the observations. To this effect, in Experiment 2 consisting of a cohort of 39 mice (Table 4.3), mice injected with the VDR: B16-BL6 clones V1 and V2, produced significantly fewer pulmonary metastases compared to the control: B16-BL6 clones C1 and C2 (Figure 4.6B). This was the case when using both microscopic met-area (P=0.0002) or macroscopic met-count (P=0.00002). Thus, observations from both experiments indicated that VDR: B16-BL6 produced significantly lower metastatic load compared to the control cells, as measured by both macroscopic and microscopic measures of metastases.

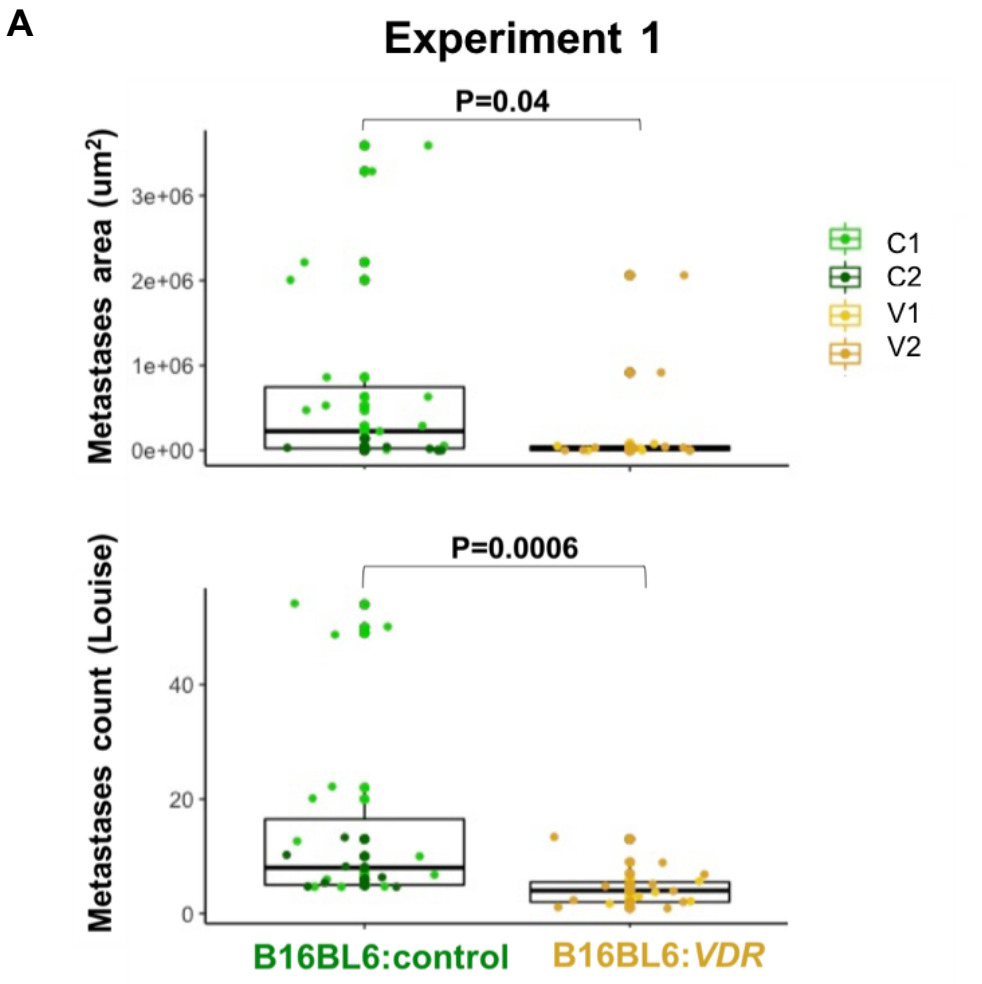
Table 4.3: In-vivo tail-vein metastasis assay: experimental layout

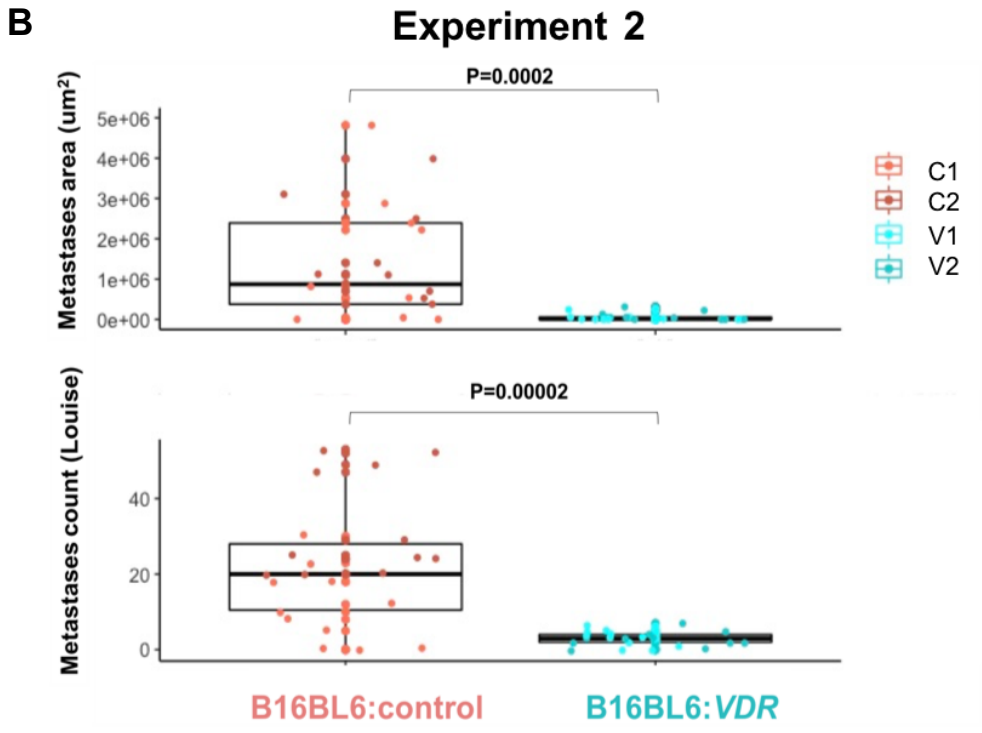
Experiment	Number of cells injected	Duration	Number of mice per group
Experiment 1	10^4 cells per mouse	21 days	C1 (n=12), C2 (n=7), V1 (n=10), V2 (n=5)
Experiment 2	10^4 cells per mouse	21 days	C1 (n=6), C2 (n=11), V1 (n=11), V2 (n=11)

Figure 4.6 (following page): *In vivo* tail-vein metastasis assay: estimation of metastatic load

A: Comparison of metastatic load from Experiment 1, estimated by microscopic metastatic area (in μm^2, determined digitally) and macroscopic metastatic count (number of pulmonary metastases) between VDR: B16-BL6 (V1 and V2) and control: B16-BL6 cells (C1 and C2). The data points from each clone is denoted in a different colour to visualise the effects of individual clones. P value from Mann-Whitney U-test

B: Comparison of metastatic load from Experiment 2, estimated by microscopic metastatic area (in μm^2, determined digitally) and macroscopic metastatic count (number of pulmonary metastases) between VDR: B16-BL6 (V1 and V2) and control: B16-BL6 cells (C1 and C2). The data points from each clone is denoted in a different colour to visualise the effects of individual clones. P value from Mann-Whitney U-test


A
Experiment 1
P=0.04
Metastases area (um²)
3e+06
2e+06
1e+06
0e+00
C1
C2
V1
V2
P=0.0006
Metastases count (Louise)
40
20
0
B16BL6:control
B16BL6:VDR


B
Experiment 2
P=0.0002
Metastases area (um²)
5e+06
4e+06
3e+06
2e+06
1e+06
0e+00
C1
C2
V1
V2
P=0.00002
Metastases count (Louise)
40
20
0
B16BL6:control
B16BL6:VDR

4.4.3.2 Estimation of tumour-infiltrating CD3 positive lymphocytes

To assess if the reduced metastatic load in mice injected with VDR: B16-BL6 cells was accompanied by increased tumour immune infiltrate, the number of CD3 positive tumour infiltrating lymphocytes was computed. For this, lung sections were used, which were sectioned consecutive to the sections used for Met-area estimation. The sections were stained with anti-CD3 antibody, after which the sections were scanned and the number of CD3 positive cells were digitally counted. Only the lymphocytes which stained positive for CD3 (membranous) in the intra-tumour region or peri-tumour region (2 cell distance) were counted. An illustrative example of the counting strategy used is depicted in Figure 4.7A. Since the estimation of CD3 positive lymphocytes was relative to the area of the tumour, using absolute count of CD3 positive lymphocytes could be potentially biased: greater the area, higher the likelihood of counting a CD3 positive lymphocyte. To circumvent this issue, the number of CD3 positive lymphocytes per 10^5 μm^2 of pulmonary met-area was estimated. In other words, estimation of the CD3 positive tumour immune infiltrate was done relative to the total tumour area (see methods). In some mice, especially in some of the mice injected with B16BL6-VDR cells, the pulmonary metastases were too few or in some cases absent (when sectioned) despite having a valid met-area and met-count. Owing to this reason, CD3 positive lymphocyte counting was not possible in all the samples, thus impacting the sample size of this comparison. Nevertheless, in Experiment 1 (Figure 4.7B), the number of tumour-infiltrating CD3 positive lymphocytes (per 10^5 μm^2) was significantly higher in metastases produced by clone V1 compared to control clone C1 ($P=0.02$). However, this difference was not apparent between the other VDR-expressing clone V2 and the control clones C1 or C2. In Experiment 2 (Figure 4.7C), the number of tumour-infiltrating CD3 positive lymphocytes (per 10^5 μm^2) was not significantly higher in metastases produced by clones V1 compared to control clone C1 and C2, with borderline significant increase in clone V2 compared to C1 ($P=0.06$).

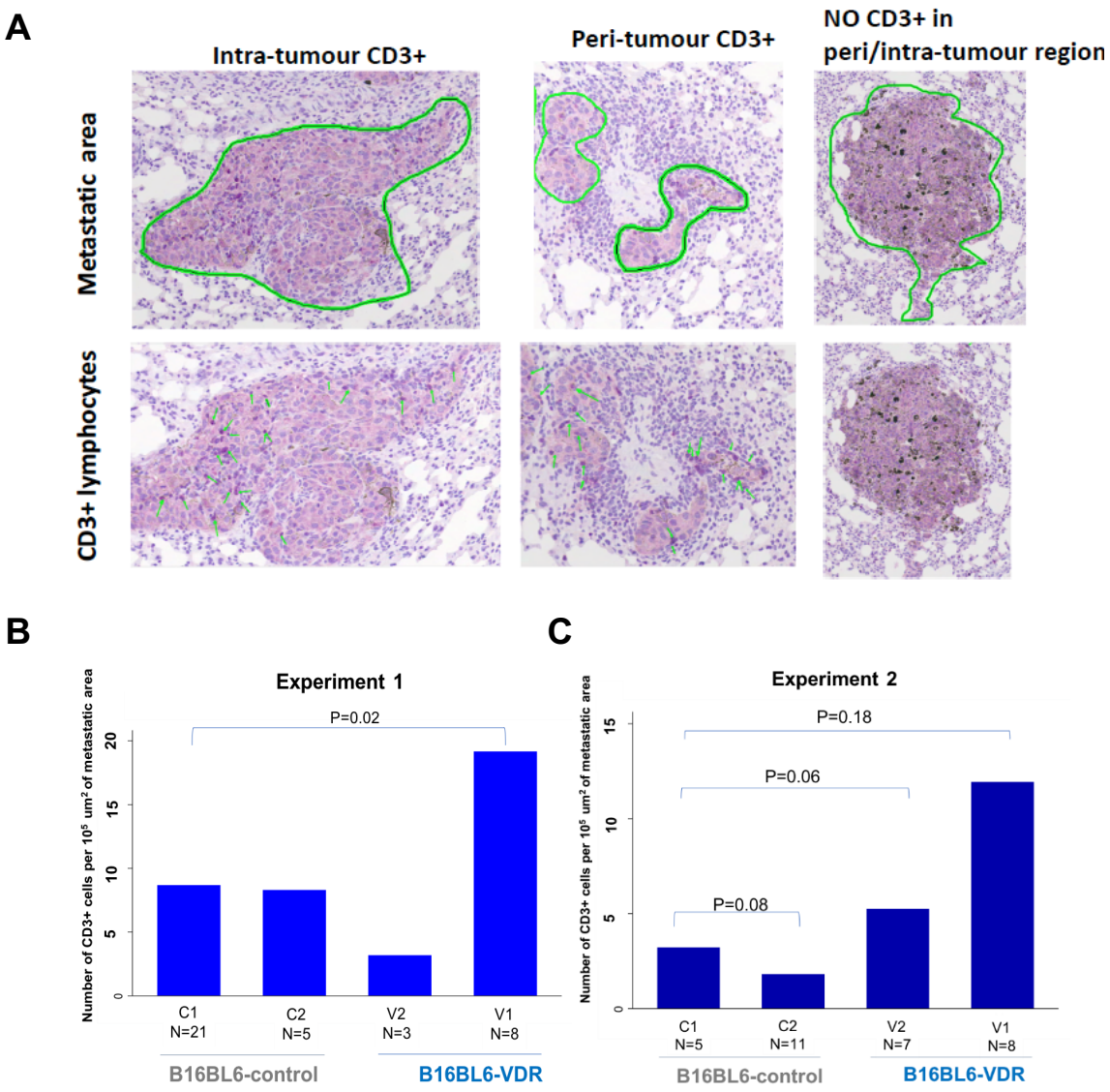


Figure 4.7: In-vivo tail-vein metastasis assay: estimation of tumour infiltrating CD3 positive lymphocytes

A: Illustrative examples of the counting strategy used to determine tumour infiltrating CD3 positive lymphocytes. Consecutive sections from FFPE lung blocks were stained for H&E and murine anti-CD3. This enabled estimation of CD3 positive tumour lymphocytes, within the context of the defined metastatic area. The top panels indicate metastatic area from H&E sections drawn digitally (as described in section 4.4.3.1) around a pulmonary metastasis. The bottom panels indicate: i) Intra-tumour CD3 positive lymphocytes are those which fall within the defined metastatic area ii) peri-tumour CD3 positive lymphocytes are those which fall just on the border of the defined metastatic area, within a 2 cell distance and iii) the lower corner right panel is an example of a pulmonary metastasis with no intra- or peri-tumour CD3 positive lymphocytes. All images were captured at 20x magnification.

Figure 4.7 description cont…

B: Comparison of number of CD3 positive tumour infiltrating lymphocytes from Experiment 1, between VDR: B16-BL6 (V1 and V2) and control: B16-BL6 cells (C1 and C2). P value from Mann-Whitney U-test. N denotes the number of lung sections which were used in each group.

C: Comparison of number of CD3 positive tumour infiltrating lymphocytes from Experiment 2, between VDR: B16-BL6 (V1 and V2) and control: B16-BL6 cells (C1 and C2). P value from Mann-Whitney U-test. N denotes the number of lung sections which were used in each group.

4.4.4 In vitro validation of the transcriptome-derived inverse correlation between VDR and Wnt/B-catenin signalling

Transcriptome-based evidence in the LMC indicates that *VDR* expression was inversely associate with proliferation associated pathways including the Wnt//β-catenin signalling pathway, which was explored further using previously-reported melanoma subtypes. In addition, the *in-vivo* experiment described in section **Error! Reference source not found.** revealed that VDR-expressing cells produced fewer pulmonary metastasis, indicating causal evidence for the anti-proliferative effect of VDR. Thus, it remained to be tested if the VDR-expressing murine melanoma cells (VDR: B16BL6) expressed significantly lower Wnt/β-catenin signalling genes compared to the control: B16BL6 cells. Two clones per cell-type, V1 and V2 (VDR: B16-BL6 cells), C1 and C2 (control: B16BL6 cells), were compared for expression of Wnt/β-catenin genes using a qRT-PCR-based array. Of the 84 Wnt/β-catenin genes tested (pre-formatted qRT-PCR-based array), 62 genes had lower expression (Fold Change <1) with 25 genes having Fold Change $\leq$0.5 in both VDR clones compared to control clones. Twelve genes had increased expression (Fold Change>1) with none having Fold Change$\geq$2 expression in both VDR clones. Concomitantly, *VDR*-expressing cells had significantly reduced expression of Wnt/β-catenin genes, including *Fgf4*, *Dkk1*, *Fzd8*, *Nkd1*, and multiple *Wnt* genes. (Figure 4.8).

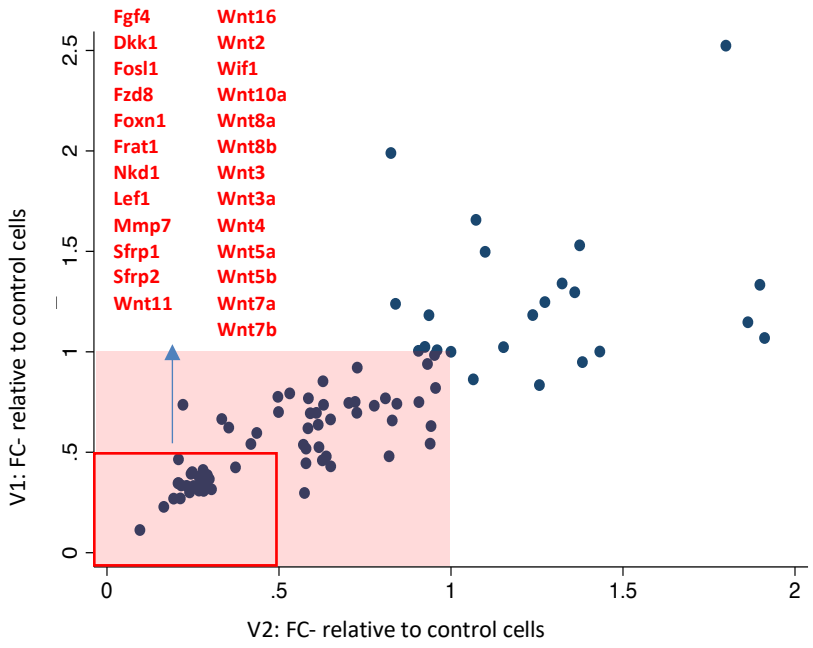


Figure 4.8: *In-vitro* validation of the transcriptome-derived inverse correlation between VDR and Wnt//β-catenin pathway genes

qRT-PCR-based comparing expression of 84 Wnt/ β-catenin genes in the two VDR-transfected clones (V1 and V2) compared to control clones. Relative expression of each gene was estimated using the Δ-ΔC_t method to calculate Fold Change. Genes with Fold Change $\leq$0.5 (n=25) in both VDR clones compared to control clones are listed in red (solid red box in graph). The genes with Fold Change $\leq$1 (n=62) in both VDR clones compared to control clones are listed in red (red region in graph).

4.5 Discussion

Even though *VDR* has been shown to have reduced expression in advanced melanomas and *in vitro* evidence for its anti-proliferative effect exists for other cancers, causal evidence in melanoma *in vivo* models is lacking. To address this, we used B16BL6 murine melanoma cells to create stably-transfected B16BL6-*VDR* cells which were used in an experimental metastasis assay. In line with the results from the primary melanoma transcriptomes, the *VDR*-transfected cells produced a significantly lower pulmonary metastatic load after tail vein administration, indicating that *VDR* expression had an anti-tumourigenic effect, which has not been previously proven in melanomas.

The estimation of metastatic area from FFPE-sections offered us the potential to estimate tumour infiltrating lymphocytes, which we did by counting the intra-tumoural and peri-tumoural CD3 positive cells as proxy for TILs. Since lungs from B16BL6-*VDR*-injected mice had no/low metastasis counts they could not be used for intra-tumoural and peri-tumoural CD3 positive TILs estimation, thus reducing sample size for this analysis. This caveat combined with the possibility that CD3 positivity might not necessarily represent the immune milieu amenable to *VDR*-specific effects, could contribute to non-significant results from this analysis. Moreover, I have only estimated the intra- and peri-tumoural CD3 positive TILs, this does not preclude the possibility of a 'pan-pulmonary' immune response. This can be quantified if by use of flow-cytometry. Taken together, the trend to greater numbers of CD3 positive cells in *VDR*-transfected metastases warrants further experimental validation in the form of additional assays (e.g. flow cytometry) for multiple immune cell lineages.

Causal evidence for VDR-mediated inhibition of Wnt/β-catenin signalling was sought, owing to the strong inverse correlation between *VDR* and Wnt/β-catenin signalling in both the LMC and TCGA data. The expression of Wnt/β-catenin signalling genes was significantly reduced in VDR-expressing murine melanoma cells. VDR-expressing cells had significantly reduced expression of the vast majority of the genes on the array i.e. 65 of the 84 genes were had fold change <1 in the VDR-expressing cells. Among these, 25 genes had fold change <0.5 in both VDR-expressing clones, including *Fgf4*, *Dkk1*, *Fzd8*, *Nkd1*, and multiple *Wnt* genes. The Wnt/β-catenin signalling pathway plays a significant role in the control of skeletal development and homeostasis (405), as does vitamin D-VDR signalling. This could explain why expression of Wnt/β-catenin signalling genes could also be controlled by VDR, as has been shown to be the case in colon cancer cells by Larriba et al (304). Similarly, in my analysis of the LMC transcriptome, *FGFR1* and *EFNB2*, which are targets of Wnt/β-catenin signalling axis, were among the genes that correlate with VDR and also have a VDR-binding site (section 3.4.4.3.2). This indicates a 'cross-talk' between the two signalling axis, which could have perhaps evolved as a means to maintain bone homeostasis. In the case of the 25 Wnt/β-catenin signalling genes downregulated in the VDR-expressing cells, some have been previously shown to be inhibited by VDR. For instance, *Dkk1* and *Sfrp2* have previously been shown to be inhibited by VDR during adipogenic differentiation (406), but not in melanomas. Interestingly, the 'classic' non-canonical Wnt ligands *Wnt5a, Wnt5b, Wnt10a, Wnt7* and *Wnt11* were among those downregulated in the VDR-expressing cells, which has not previously been reported in melanomas. This finding is of significance because Wnt5a (and some other non-canonical Wnt ligands) affect cell motility and invasion and is implicated in worse

melanoma prognosis (407, 408). Thus, the qRT-PCR findings are complementary and provide functional validation of the transcriptome-based findings for the inverse association between VDR and Wnt/β-catenin signalling.

Since the above qRT-PCR based results are derived from murine melanoma cells which are only 'controlled for' VDR expression, I was curious to assess the effect of vitamin D treatment on Wnt/β-catenin signalling. An ideal case would have been to treat the VDR: B16-BL6 and control: B16-BL6 cells with 1,25-hydroxy vitamin D_3 and compare the expression of Wnt/β-catenin signalling genes among treatment groups. However, this was not possible owing to logistical and time limitations. Alternatively, I performed an analysis of microarray data from human melanoma cells (MeWo and SkMel28) treated with (and without) 1,25-hydroxy vitamin D_3, which was generated by a past PhD student in the group- Dr. Anastasia Filia (unpublished data). I queried the microarray data[4], which were generated from two treatment time points: 24 and 48 hours after treatment with 1,25 hydroxy vitamin D_3. My analyses revealed the following:

i. VDR expression was upregulated at 24 and 48 hours after treatment with 1,25 hydroxy vitamin D_3
ii. Among the most significantly downregulated genes (in both cell lines), were those enriched for Wnt/β-catenin signalling. This was identified using an agnostic analysis i.e. an enrichment analysis (using Reactome FIViz) of the downregulated genes in response to 1,25 hydroxy vitamin D_3 treatment (after 24 and 48 hours) identified Wnt/β-catenin signalling as one of the top downregulated pathways

These findings, despite being from a different *in vitro* model (human rather than murine melanoma cells) agnostically identified Wnt/β-catenin signalling to be downregulated in response to 1,25 hydroxy vitamin D_3 treatment, along with the upregulation of *VDR* expression. These findings are complementary to those derived from the VDR-expressing murine melanoma cells.

Taken together, findings from both *in vitro* models, along with the transcriptomic findings suggest a significant role for the vitamin D-VDR signalling in inhibiting Wnt/β-catenin signalling in melanomas.

[4] I have not presented data pertaining to this analysis, since it is based on data-derived by Dr. Anastasia Filia's work.

Chapter 5

Transcriptomic interrogation of microscopic ulceration of primary melanomas

5.1 Chapter aims and overview

Chapter aims	Section
To assess the frequency of ulcerated and non-ulcerated tumours in the LMC dataset	5.4.1
To assess the clinicopathological features associated with melanoma ulceration	5.4.2
To assess the differentially expressed genes and signalling pathways that vary between ulcerated and non-ulcerated melanomas	5.4.3

Contributions to this chapter:

- Sathya Muralidhar performed statistical and bioinformatics analyses described in this chapter, under the supervision of Dr Jeremie Nsengimana (senior statistician in the group) and guided by Prof. Julia Newton-Bishop and Prof. Tim Bishop.
- Histopathological measures of immune infiltrate (described in 3.4.6.2) in the LMC primary melanomas was derived from extensive work done by Dr Sally O'Shea who reviewed all the histological slides according to protocol

5.2 Introduction

5.2.1 Melanoma ulceration

Ulceration status of primary melanomas is an independent predictor of adverse prognosis (312, 409) and has been an integral part of the AJCC melanoma classification system (410). The most recent AJCC melanoma staging (8th edition) defines ulceration as '*full thickness absence of an intact epidermis above any portion of the primary tumour with an associated host reaction (characterized by a fibrinous and acute inflammatory exudate) above the primary tumour based on histopathological examination'* (Figure 5.1). The AJCC staging system is widely clinically implemented, however there are studies that demonstrate the complexity of the ulceration phenomenon. For instance, ulceration has been reported to be associated with

alterations to the surrounding epidermis (411) and with histopathological features such as Consumption of Epidermis (COE) (412) and Sub-epidermal Cleft Formation (SCF) (413). Interestingly, the extent of ulceration (as a percentage of tumour) and type of ulceration (attenuative vs infiltrative) have also been shown to be independent predictors of adverse melanoma prognosis (414, 415). Given the prognostic significance of ulceration in predicting adverse melanoma prognosis, efforts to gain a better understanding of the ulceration 'phenomenon' have revealed valuable insights.

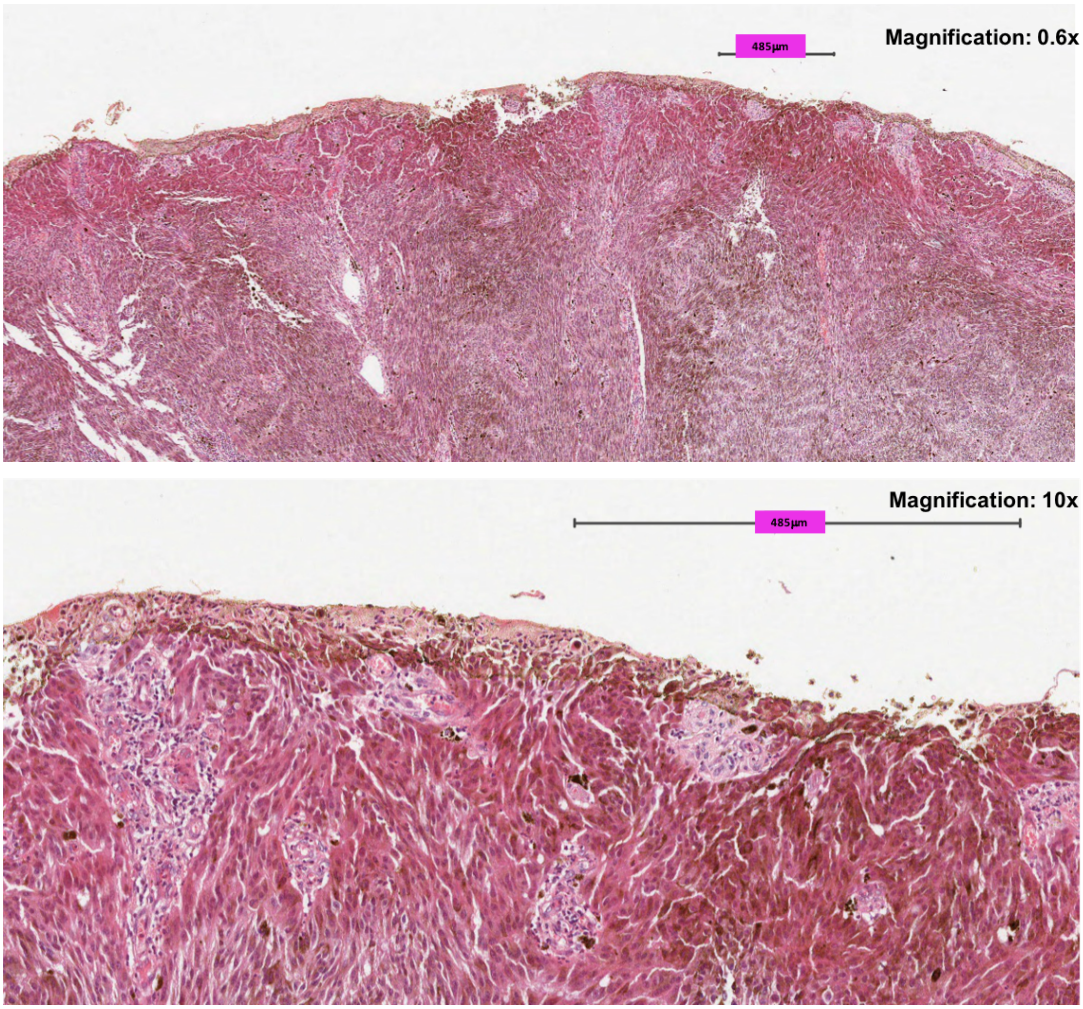


Figure 5.1: Representative images of melanomas with evidence for microscopic ulceration

Top image depicts loss of epidermal integrity at 0.6x magnification and the bottom image at 10x magnification of a selected section of the tumour.

Ulceration of primary cutaneous melanomas has been associated with histopathological features, indicators of systemic inflammation and therapy response factors. Histopathological evidence indicates that ulcerated primary melanomas are

associated with increased macrophages and lymphatic vessel invasion. This was shown by a previous study by our group, where 202 cutaneous primary melanoma sections were assessed for IHC-based expression of CD34 (endothelial marker) and CD68 (macrophage marker) to estimate blood vessel density and macrophage counts respectively. This study showed that vessel invasion density (both blood and lymphatic vessels) and macrophage counts were significantly higher in ulcerated tumours compared to non-ulcerated tumours (416), suggesting ulceration to be a marker of tumour-associated inflammatory microenvironment. The association of ulceration with an inflammatory microenvironment has also been shown in a study by Jewell et al (from our own group) and was based on a subset of the 702 LMC primary melanomas (417). These findings combined with the view that obesity, diabetes, hypertension and cardiovascular disease are known to cause pro-inflammatory tumour states, a subsequent study by von Schuckmann et al aimed to address the association of melanoma ulceration with diabetes and statin usage. This study demonstrated in a cohort of 787 melanomas (194 ulcerated, 593 non-ulcerated) that regular statin users had a lower likelihood of being diagnosed with ulcerated melanomas. In the same cohort of patients, those with tumours $\leq$2 mm thick and diagnosed with diabetes had a higher likelihood of having an ulcerated melanoma. The findings from this study collectively supported the hypothesis that statin use is inversely and diabetes is positively associated with ulcerated melanomas (418). Taken together, there is evidence for the association of ulcerated melanomas with indicators of both systemic and tumour-associated inflammation.

Ulceration status has also been shown to be associated with response to melanoma therapy. Patients with ulcerated tumours benefit significantly from IFN/PEG-IFN adjuvant therapy i.e. ulcerated tumours treated with IFN/PEG-IFN adjuvant therapy had improved Relapse Free Survival (RFS), Overall Survival (OS) and Distant Metastases Free Survival (DMFS). This observation suggested that ulceration might be a marker of melanoma response to IFN/PEG-IFN therapy (419). In a recent study by Koelblinger et al, ulcerated melanomas had a significantly higher proportion of PDL1-expressing tumour cells compared to non-ulcerated tumours, suggesting that the ulceration phenomenon could be involved in immune evasion with a consequent effect on response to immune therapy (420).

Efforts to understand the genomic basis of ulceration have been addressed by two studies to date: by Rakosy et al and Jewell et al. Rakosy et al used a dataset of 32 samples, comparing 16 ulcerated to 16 non-ulcerated primary melanomas. This study identified genes/pathways, CNAs and methylation patterns associated with ulceration status in these tumours (421). The study by Jewell et al (417) was from our own group

and was based on a subset of the 702 LMC primary melanomas: comparing 50 ulcerated to 145 non-ulcerated primary melanomas. This study interrogated the differential expression of a cancer gene panel of 502 genes between ulcerated and non-ulcerated tumours. Both studies offer valuable insight into the factors that underpin ulceration. They indicate that ulcerated tumours are associated with reduced expression of genes pertaining to cell-cell-adhesion pathways and increased expression of proliferation-associated pathways. Moreover, the study by Rakosy et al indicated that ulceration is associated with changes in both transcriptomic and copy number level. To this effect, they identified the following pathways to be significantly downregulated in ulcerated melanomas: p53, NFκB and Wnt/β-catenin signalling. In addition, they also reported loss of regions in 6q and one region in 10q, which had significant loss of copy number in the ulcerated compared to non-ulcerated tumours.

The current study uses 671 tumours (a subset of the 702 primary melanomas described in previous sections of this thesis) and their corresponding clinical, histopathological, copy number and transcriptomic data to gain a deeper insight into the phenomenon of melanoma ulceration.

5.3 **Methods**

5.3.1 Correlations with clinicopathological variables

In the case of the clinicopathological variables in a continuous scale: age at diagnosis (in years) and Breslow thickness, a univariable linear regression model was used to assess significant differences between ulcerated and non-ulcerated tumours. In the case of clinicopathological variables which were categorical: sex, tumour site, vascular invasion and histopathological tumour immune infiltration, a Pearson chi-squared test was performed to assess if the proportion of ulcerated and non-ulcerated tumours was significantly different between respective categories.

5.3.2 Whole-transcriptome differences between ulcerated and non-ulcerated tumours

The classification of ulceration status by the clinical pathologist's report ('Reported' ulceration) was used for this analysis, since this information was available for most tumours. Each gene (from a total of 20,560 genes) was checked for differential expression between the ulcerated (n=234) and non-ulcerated (n=468) tumours, using a Mann-Whitney U-test (Stata command: 'ranksum'). Since a total of 20,560 tests were performed (one per gene), multiple correction had to be applied in order to adjust for

false discovery. The Benjamini-Hochberg method of multiple correction was used to compute a False Discovery Rate (FDR). The Mann-Whitney U-test also produced a z-score, which indicated the 'direction' of differential expression i.e. if a gene was significantly higher or lower in the ulcerated tumours compared to non-ulcerated tumours. At z-score>0, genes were identified as having significantly lower expression in ulcerated tumours compared to non-ulcerated tumours. At z-score<0, genes were identified as having significantly higher expression in ulcerated tumours compared to non-ulcerated tumours.

5.3.3 Enrichment analysis

Reactome FIviz was used to perform enrichment analysis to identify the pathways enriched for a set of genes that were differentially expressed in ulcerated tumours (genes identified by the whole-transcriptome Mann-Whitney U-test described above).

5.4 Results

5.4.1 Distribution of ulcerated and non-ulcerated tumours in the LMC dataset

The histopathological classification of a melanoma tumour as 'ulcerated' or 'non-ulcerated' has been shown to be variable depending on the reviewing pathologist. Thus, it was necessary to assess the degree of concordance between two independent histopathological classifications of ulceration status. The 702 primary melanomas were classified as either ulcerated (n=234) or non-ulcerated (n=468) upon review by the clinical pathologists. A subset of the 702 tumours (n=675) were also reviewed independently by Dr Sally O'Shea, in our group. The concordance between the classification of ulceration status ('Yes' or 'No') by the clinical pathologists' ('Reported' henceforth) and Dr Sally O'Shea ('SOS' henceforth) was assessed in the 675 tumours whose ulceration status was reported by both (Figure 5.1Table 5.1). Of the 675 tumours, the ulceration status of 84% of the tumours (n=567) was in agreement between the two reports. However, the other 16% of tumours (n=108) were in disagreement: 94 tumours were classified as 'No' ulceration by SOS but as 'Yes' ulceration by Reported; 14 tumours were classified as 'Yes' ulceration by SOS but as 'No' ulceration by Reported.

Since the clinical pathologist's classification of ulceration ('Reported') was used to classify the tumours based on the AJCC staging system, this measure of ulceration was used for analyses henceforth.

Table 5.1: Concordance between ulceration status reported by clinical pathologist ('Reported' ulceration) and ulceration status reported by pathologist in our group: Dr Sally O' Shea ('SOS' ulceration)

** indicates number of overlapping samples

	Reported No	Reported Yes	Total SOS ulceration
SOS No	435	94	529
SOS Yes	14	128	142
Total Reported ulceration	451	224	675**

5.1.1 Clinicopathological features associated with ulcerated tumours in the LMC

Clinicopathological features were compared between ulcerated and non-ulcerated tumours. The age at diagnosis was significantly higher in participants whose tumours were ulcerated compared to non-ulcerated tumours (Table 5.2). There was no significant difference in the proportion of ulcerated and non-ulcerated tumours between participants who were male or female (Table 5.2). Ulcerated tumours also had a significantly higher Breslow thickness compared to non-ulcerated tumours (Table 5.2, $P=6.2x10^{-23}$). Tumours in the rare sites (sun protected sites) were more likely to be ulcerated compared to tumours arising on the head (P=0.0003, Table 5.2). However, there was no significant difference in the proportion of ulcerated and non-ulcerated tumours, in comparing tumours from the head with those from the limbs (P=0.47) or the truncal tumours (P=0.62). A significantly higher proportion of ulcerated tumours (54%) had reported vascular invasion compared to non-ulcerated tumours (46%) (P=0.0003, Table 5.2). The proportion of tumours with absent and non-brisk tumour immune infiltration did not vary significantly between ulcerated and non-ulcerated tumours (P=0.26). However, the proportion of tumours with brisk immune infiltrate was significantly higher among ulcerated tumours (75%) compared to non-ulcerated tumours (48%) (P=0.005) (Table 5.2).

Table 5.2: Association of clinicopathological features with ulceration status in the LMC primary melanomas

[B] indicates baseline

Ulceration association with	Regression Coefficient (R) or Pearson's chi^2 (χ^2)	P-val
Age at diagnosis (years)	R=3.36	0.001
Sex		
Females[B]		
Males	χ^2=1.65	0.19
Breslow thickness (mm)	R=1.77	6.2e-23
Tumour site		
Head[B]		
Limbs	χ^2=0.50	0.47
Trunk	χ^2=0.23	0.62
Rare (sun-protected sites)	χ^2=12.85	0.0003
Vascular invasion		
No[B]		
Yes	χ^2=12.54	0.0003
Tumour Immune Infiltrate		
Absent[B]		
Non-brisk	χ^2=1.22	0.26
Brisk	χ^2=7.73	0.005

As expected, ulcerated tumours also had a significantly worse prognosis compared to non-ulcerate tumours (Figure 5.2). This effect was significant even in a multivariate survival analysis i.e. the worse prognosis of ulcerated tumours was independent of age, Breslow thickness, tumour site and mitotic number (HR=1.66, P=0.001).

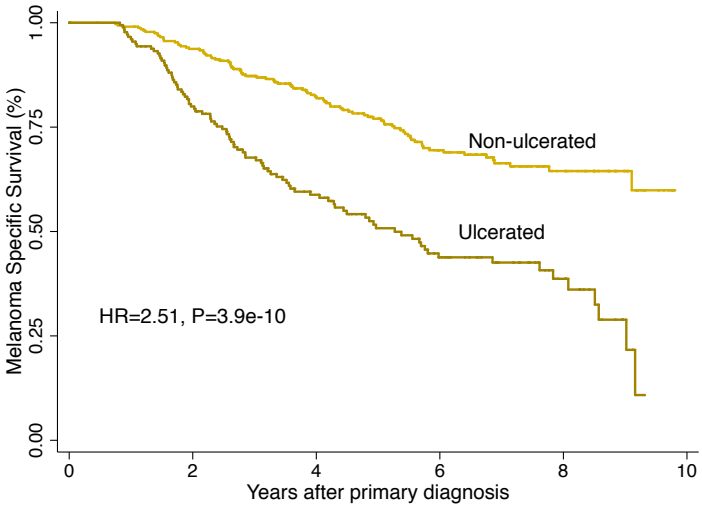


Figure 5.2: Difference in survival of participants whose melanomas were classified as ulcerated or non-ulcerated at diagnosis

Hazard Ratio (HR) and P-value (P) are from univariable Cox Proportional Hazards model

5.4.2 Transcriptomic correlates of ulcerated and non-ulcerated tumours

5.4.2.1 Genes differentially expressed between ulcerated and non-ulcerated tumours

In order to agnostically identify the genes which are significantly differentially expressed between the ulcerated and non-ulcerated tumours, a whole-transcriptome Mann-Whitney U-test was performed. The classification of ulceration status by the clinical pathologist ('Reported' ulceration) was used for this analysis. Each gene (from a total of 20,560 genes) was checked for differential expression between the ulcerated (n=234) and non-ulcerated (n=468) tumours, using a Mann-Whitney U-test. Since a total of 20,560 tests were performed (one per gene), multiple correction had to be applied in order to adjust for false discovery. Using the Benjamini-Hochberg method of multiple correction, 4660 genes were identified (at FDR<0.05) whose expression was significantly different between ulcerated and non-ulcerated tumours. The Mann-Whitney U-test also produced a z-score, which indicated the 'direction' of differential expression i.e. if a gene was significantly higher or lower in the ulcerated tumours compared to non-ulcerated tumours. At z-score>0, 1979 genes were identified as having significantly lower expression in ulcerated tumours compared to non-ulcerated tumours. At z-score<0, 2681 genes were identified as having significantly higher

expression in ulcerated tumours compared to non-ulcerated tumours. To identify the biological pathways enriched for the aforementioned genes differentially expressed in ulcerated versus non-ulcerated tumours, pathway enrichment analyses were done, described in the following section.

5.4.2.2 Pathway enrichment for genes differentially expressed in ulcerated versus non-ulcerated tumours

The 2681 genes whose expression was significantly higher in ulcerated tumours (compared to non-ulcerated tumours) were enriched for pathways such as mitotic prometaphase, signalling by Rho-GTPases, cell cycle checkpoint, mitochondrial translation, PLK-signalling and FOXM1 transcription network (top 20 pathways listed in Table 5.3, see Appendix T5-1 for full list of pathways). Conversely, the 1979 genes which had significantly lower expression in ulcerated tumours were enriched for Extracellular matrix organisation, Interferon gamma signalling, cytokine-cytokine receptor interaction, IL-12 mediated signalling, PI3K-Akt signalling, TCR signalling in naïve CD8 T cells and Focal adhesion (top 20 pathways are listed in Table 5.4, see Appendix T5-2 for full list of pathways).

Table 5.3: List of pathways enriched for genes which are expressed significantly higher in ulcerated tumours compared to non-ulcerated tumours

Pathways	P-value
Mitotic Prometaphase(R)	1.11E-16
Mitotic Metaphase and Anaphase(R)	1.11E-16
Signalling by Rho GTPases(R)	2.22E-16
Mitotic G1-G1/S phases(R)	1.97E-13
Cell Cycle Checkpoints(R)	2.35E-13
Synthesis of DNA(R)	2.51E-11
S Phase(R)	4.18E-11
Mitochondrial translation(R)	2.80E-10
Cell cycle(K)	3.76E-10
HDR through Homologous Recombination (HR) or Single Strand Annealing (SSA)(R)	4.00E-10
PLK1 signalling events(N)	1.04E-09
RNA Polymerase I, RNA Polymerase III, and Mitochondrial Transcription(R)	1.14E-09
DNA replication(K)	1.29E-09
Nucleosome assembly(R)	2.06E-09

Pathways	P-value
Mitotic G2-G2/M phases(R)	3.63E-09
Nucleotide Excision Repair(R)	6.54E-09
Aurora B signalling(N)	6.89E-09
Validated targets of C-MYC transcriptional activation(N)	8.28E-09
M/G1 Transition(R)	1.23E-08
Fanconi anemia pathway(N)	4.29E-08
ATR signalling pathway(N)	6.74E-08

Table 5.4: List of pathways enriched for genes which are expressed significantly lower in ulcerated tumours compared to non-ulcerated tumours

Pathways	P-value
Extracellular matrix organization(R)	1.45E-11
Pathways in cancer(K)	9.67E-10
ECM-receptor interaction(K)	1.38E-09
Cytokine-cytokine receptor interaction(K)	9.87E-09
Interferon gamma signalling(R)	3.48E-08
Beta1 integrin cell surface interactions(N)	1.25E-07
HTLV-I infection(K)	2.54E-07
Cell adhesion molecules (CAMs)(K)	3.73E-07
PI3K-Akt signalling pathway(K)	4.57E-07
Inflammatory bowel disease (IBD)(K)	4.61E-07
Axon guidance(K)	1.60E-06
Chemical carcinogenesis(K)	2.50E-06
Metabolism of xenobiotics by cytochrome P450(K)	2.79E-06
Drug metabolism - cytochrome P450(K)	3.67E-06
IL12-mediated signalling events(N)	3.90E-06
T cell activation(P)	9.37E-06
Amoebiasis(K)	1.28E-05
Toxoplasmosis(K)	1.52E-05
TCR signalling in naïve CD8+ T cells(N)	1.75E-05
Staphylococcus aureus infection(K)	2.13E-05

5.5 Discussion

The ulceration status of melanomas tumours is the second most powerful independent predictor of survival in melanoma patients, Breslow thickness being the first (312, 313). It is not just the presence but also the extent of melanomas which is associated with melanoma survival (415). More importantly, ulceration could be a marker of response to interferon treatment (419). Though ulceration is an important part of the AJCC staging system which has a defined 'guideline' for classifying a melanoma as ulcerated or non-ulcerated, there remains discordance among pathologists with regards to classification. One of the reasons for this is the loss of epidermis (which is a defining characteristic of ulcerated melanomas) arising from sample handling (the epidermis being prone to lacerations during sectioning) rather than ulceration itself. This means that though there is overall concordance between pathologists' classification of ulceration status, there still remains a portion of melanomas that are 'misclassified'. I have described one such instance in the LMC primary melanomas, where ulceration status from the clinical pathologist ('Reported ulceration') and Dr. Sally O'Shea in our group ('SOS ulceration') were largely concordant, with 84% of tumours being in agreement. However, the remaining 16% of tumour were discordant. This observation, in addition to ulceration being a potential marker of therapy response, necessitate the identification of genomic features that characterise the phenomenon of ulceration. This also served as motivation for this project: to identify the clinical, histopathological and transcriptomic correlates associated with ulcerated melanomas in the LMC dataset. Though this has previously been queried in other datasets using different approaches (Rakosy et al and Jewell et al, discussed below), the advantage of using the LMC primaries to interrogate ulceration are:

i. The relatively large sample size of the LMC primary melanomas, enabling comparison of ulcerated (n=234) and non-ulcerated (n=468) tumours
ii. Availability of *in silico* measures of immune compartments i.e. imputed immune cell scores, which would enable identification of specific immune components associated with ulceration status.
iii. The availability of CNA data for a subset of the LMC data enables comparison of these features between ulcerated and non-ulcerated tumours.

As a first step, I identified the clinicopathological correlates of ulcerated tumours and report that ulcerated tumours were more likely to be thicker, have more vascular invasion and more likely to be from rare (sun-protected) melanomas rather than sun-exposed melanomas arising in the head. Older participants were also more likely to be diagnosed with ulcerated melanomas. These findings indicate that ulceration status is associated with clinicopathological features that are associated with poor melanoma

prognosis (increased age, thicker tumours, tumours in rare sun-protected sites) and hence is a poor predictor of prognosis by itself.

With the view of interrogating the transcriptomic basis of melanoma ulceration, I identified the differentially expressed genes in ulcerated versus non-ulcerated melanomas in the LMC, followed by identification of signalling pathways enriched for these genes. In doing so, I have compared my findings with those from other reported studies. Previous attempts to understand the molecular basis of ulceration have revealed that ulcerated melanomas have reduced expression of genes involved in cell adhesion such as Desmoplakin, Integrins, Cadherin 1 and fibroblast growth factors 2 and 3 (417) and proliferation associated pathways such as p53 and Wnt/β-catenin signalling (421). In addition, the study by Jewell et al, which is from the Leeds group and was based on subset (n=195) of the 703 LMC primary melanomas, identified increased histopathological evidence of macrophages in ulcerated tumours. My analyses of ulceration in the LMC was concordant with these findings. In that, ulcerated tumours (n=235) had significantly higher expression of genes enriched for cell cycle, mitotic prometaphase and anaphase, mitochondrial translation and Aurora B signalling. PLK1 signalling was also identified, which is concordant with Jewell et al. Interestingly, DNA damage repair pathways such as ATR signalling and nucleotide excision repair were also identified to have higher expression in ulcerated tumours, which has not been previously reported. The pathways identified by Rakosy et al (Wnt signalling and p53 signalling) though were not part of the top most significant pathways, were still expressed higher in ulcerated melanomas in the LMC. Among the pathways that were expressed significantly lower in the ulcerated LMC melanomas were integrin-beta signalling, extracellular matrix organisation, cell adhesion molecules, IL12 signalling, PI3K signalling, IFNG signalling and TCR signalling by naïve CD8 T-cells. Once again, these pathways were concordant with the findings from Rakosy et al and Jewell et al, who identified cell adhesion processes to be reduced in ulcerated tumours. However, the reduced expression of TCR signalling, IL12 and PI3K signalling in ulcerated tumours have not been previously reported. Taken together, my analyses suggest a concordance with previously studies, with regards to the signalling pathways/processes that are associated with the microscopic ulceration of primary melanomas.

The analyses I have done thus far have provided useful insights with regards to identifying the transcriptomic correlates of ulceration in the LMC primary melanomas. However, these are preliminary analyses and are meant to complement findings from additional approaches, which include:

i. Comparison of imputed immune cell scores between ulcerated and non-ulcerated tumours

ii. Using a machine learning-based approach to perform a combined analysis of transcriptomic gene expression and CNAs, with the view of identifying the nodal genes/CNAs which most significantly characterise melanoma ulceration. This approach will use an Artificial Neural Network (ANN) to identify the nodal genes and/or CNAs that are best able to distinguish between the profiles of ulcerated and non-ulcerated tumours. This would be followed by a Network Inference (NI) algorithm, which would identify the relationship (interactions) between the aforementioned nodal genes. This is currently work in progress and is being performed in collaboration between Prof. Graham Ball (Nottingham Trent University) and myself, under the supervision of Prof. Julia Newton-Bishop and Prof. Tim Bishop.

The rationale for electing to use a machine learning-based approach is to account for the non-linearity of the transcriptomic and CNA data, which I have previously not accounted for in using linear regression models. In my analyses of vitamin D-VDR signalling, the linear regression model was used to identify correlates which I subsequently was able to validate *in vitro* and *in vivo*. However, in the case of case of ulceration, there are no known murine/cellular models of ulcerated melanomas. This necessitates the use of an approach which would reduce error and increase the likelihood of identifying the nodal genes/pathways. Nevertheless, as a validation of the findings from the machine learning-based approach, I plan to perform IHC-based validation. The findings from this approach, in addition to providing biological insight, could serve to identify a biomarker to enable classification of ulcerated melanomas with increased reliability.

Taken together, the aim of this chapter was to describe the rationale and preliminary analyses pertaining to identifying the transcriptomic basis of melanoma ulceration in the LMC. However, the additional analyses using a machine learning-based approach and IHC validation are work in progress and will be pursued by me imminently.

Chapter 6
Collaborative initiatives

Being the largest treatment-naïve cohort of primary melanomas for which transcriptomic and clinical data are available, the LMC dataset is a useful source to validate in-vitro or in vivo-derived hypotheses. To this effect, I have worked with other melanoma research groups to interrogate specific questions using the LMC dataset. Though these projects were largely related to my own PhD projects, they have been useful learning experiences. In that, these projects presented me the opportunity to design and optimise a workflow in the LMC dataset, which were best suited to address *in vitro* or *in vivo* derived hypotheses.

In this chapter I describe three such projects which I had undertaken on a collaborative basis, during the course of my PhD.

6.1 Association of IFNG/IL6 signatures with sun exposure in the LMC

This project was undertaken in collaboration with Dr Amaya Viros (PI) and Ms. Katharina Roeck (PhD student) from the CRUK Manchester Institute. Based on the *in vitro* and *in vivo* findings from Dr Viros's group, I tested the following hypothesis using the LMC dataset: IL6 and Interferon gamma signalling is significantly higher in chronic sun-exposed tumours from patients >60 years of age, compared to tumours with relatively lower sun exposure. I tested this hypothesis in the LMC using two variables which indicate sun exposure: the direct measure of sun exposure (described in Chapter 3: section 3.3.1.7) and the anatomical tumour site (described in Chapter 3: section 3.3.1.3). I also tested two different approaches to assess if IL6 and Interferon gamma signalling was associated with sun-exposure: a candidate approach and an agnostic approach.

Candidate approach: 200 genes involved in IL6 and Interferon Gamma signalling were chosen (from MSigDB: HALLMARK_INTERFERON_GAMMA_RESPONSE) and their differential expression between head and trunk tumours in patients $\geq$60 years was checked. Of the 200 genes, only 11 genes were differentially expressed between head and trunk tumours in patients $\geq$60 years, none of which were identified in their *in vitro* screens. So alternatively, I created pathway scores for IL6 and Interferon gamma: each pathway score was the average expression of the composite genes, across the 703

LMC tumours. However, the 'IL6 signalling score' and 'Interferon gamma signalling score' did not vary significantly between head and trunk tumours in patients ≥60 years. Similarly, neither the IL6 signalling score nor the Interferon gamma signalling score correlated significantly with the measure of self-reported sun exposure.

Agnostic approach: Whole genome differences between tumours on the head and those on the trunk was assessed (Mann Whitney U-test). The head and truncal tumours were chosen as comparison groups because these were considered to be the most sun-exposed and least sun-exposed respectively. Upon using a linear regression model and a multiple correction threshold of FDR<0.05, 802 genes were found to be differentially expressed between head and trunk tumours. The signalling pathways enriched for these genes were assessed using Reactome FIViz and are summarized as follows: i) pathways enriched for genes whose expression is higher in tumours on the head (compared to truncal tumours) include mitotic prophase and metaphase, cell cycle checkpoint, Wnt signalling pathways and ATM pathway. ii) pathways enriched for genes whose expression is higher in tumours on the trunk (compared to those on the head) include PI3K-Akt signalling pathway, focal adhesion, Eukaryotic translation initiation and Natural killer cell-mediated cytotoxicity. Though this particular analysis provides significant insight into the transcriptomic differences between tumours on the head ('sun-exposed') and trunk ('non sun-exposed'), the IL6/Interferon gamma pathway was not amongst those identified.

Based in the analyses described above, it was concluded that there was no substantial evidence for the hypothesis that IL6 or Interferon gamma signalling vary with respect to sun exposure in the LMC primary melanomas.

6.2 Midkine signatures in the LMC

This project was undertaken in collaboration with Dr Marisol Soengas (PI) and Dr David Olmeda (Post-doctoral fellow) from CNIO, Spain. The work done by Dr Soengas's group was focussed on the secretory protein Midkine (coded for by the gene *MDK*), which they had shown to be a systemic inducer of neo-lymphangiogenesis that defines patient prognosis (422). More recently, the group had generated Gain of function (GoF) and Loss of function (LoF) MDK "signatures" from melanoma cell lines which were genetically altered to be either *MDK*-null or overexpress *MDK*. The signatures were composed of genes which were significantly upregulated or downregulated in response to *MDK* loss.

The MDK signature derived by the Soengas group was applied to the 703 Leeds primary tumours transcriptome. This classified each tumour into either 'High-MDK' or

'Low-MDK' cluster. This classification was performed by Dr. Jeremie Nsengimana in the Leeds group. All subsequent analyses of transcriptomic and clinical features in MDK-high versus MDK-low tumours were performed by me.

The High-MDK tumours had significantly better prognosis than Low-MDK tumours. To better understand why this protective effect exists, I used two approaches:

6.2.1 Check if the immune contexture is different in the Low vs High MDK tumours

I used two immune signatures to assess this:

a) Angelova/Pozniak immune scores- derived by our group (Joanna Pozniak) and is based on unique gene expression of 27 immune cell types. The High-MDK cluster appears to have a higher proportion of 'high-immune' tumours compared to Low-MDK cluster. Because of this significant difference, I checked which of the 27 immune cell scores best define the MDK clusters. 26 immune cell scores were significantly higher in High-MDK cluster (compared to Low-MDK cluster).
b) The TCGA classification- These are published molecular phenotypes which have been applied to the 703 primaries, thus classifying each tumour into either Immune, Keratin or MITF-low subtypes. The High-MDK cluster has a higher proportion of Immune subtype tumours compared to Low-MDK cluster.

6.2.2 Check for whole-genome differences to agnostically identify genes/pathways that vary between Low and High MDK tumours

The genes that vary significantly between Low and High MDK clusters were identified using a whole genome Mann-Whitney test. The pathways corresponding to these genes were identified using Reactome FIViz enrichment. The High-MDK cluster appears to have significantly higher expression of genes corresponding to pathways such as NFKB, TLR, IL12-mediated, TRAIL, MAPK and TNF signalling. The nodal genes involved in these pathways were *STAT3, HLA-DRA, REL, JUNB, MAPK134* (slide8).

Conversely, pathways such as Mitotic Metaphase and Anaphase, mitochondrial translation and mitochondrial translation were higher in low-MDK tumours. The nodal genes involved in these pathways were *PSMD8, NDUFA7, UBA52, MRPL14 and RPS27A*

Similarly, The MDK signature was applied to the TCGA metastatic tumours. This classified each tumour into either 'High-MDK' or 'Low-MDK' cluster. Converse to the findings from the LMC primary melanomas, the High-MDK tumours had significantly

worse prognosis than Low-MDK tumours. To better understand why this effect exists, I used the exact same approaches as described for the Leeds primaries

- Angelova/Pozniak immune scores: contrary to the Leeds data, the proportion of high/low/intermediate immune tumours does not vary between High-MDK and Low-MDK clusters. However, High-MDK cluster has significantly higher NK56dim, pDC and Th17 cells scores. Also, High-MDK cluster has significantly lower central memory and effector memory CD8 T-cells.
- The High-MDK cluster appears to have significantly higher expression of genes corresponding to pathways such as mitochondrial translation, oxidative phosphorylation, TNF and VEGF signalling. The nodal genes involved in these pathways are *RELA, AKT1, UBA52*. Conversely, pathways such as Mitotic Metaphase/Anaphase and TLR signalling were higher in low-MDK tumours. The nodal genes involved in these pathways are *RPS27A and HDAC2.*

In summary, High-MDK metastatic tumours have worse prognosis in the LMC, which is contrary to the TCGA metastatic melanomas. The LMC primary melanomas with high-MDK also had increased immune cell scores for NK56dim, pDC and Th17 cells, while having lower central and effector memory CD8 cell scores. Pathways such as TNF, VEGF signalling and oxidative phosphorylation are upregulated in these tumours.

6.3 G9A in the LMC

This project was undertaken in collaboration with Dr David Fisher (PI) and Dr. Shinichiro Kato (Post-doctoral fellow) from Massachusetts General Hospital, Harvard Medical School, USA. The design and execution of analyses related to this project were performed jointly by myself and Ms. Joanna Pozniak (PhD student in the group). The focus of the research done in Dr Fisher's group was the methyltransferase G9A (also known as EHMT2), which was shown to play a role in silencing tumour suppressor genes and increasing the expression of genes involved in tumour survival. However, the role of G9A as a direct oncogenic driver has not previously been described, which was the primary hypothesis tested in this project. *In vitro* data from Dr Fisher's group indicated that loss of G9A expression in melanoma cells resulted in reduced expression of Wnt/β-catenin target genes (such as *Myc* and *Lef1*) and increased expression of Wnt antagonists (such as *Dkk1, Sfrp1* and *Wnt5a*). These *in vitro* observations prompted the assessment of the potential role of G9A in the LMC primary melanomas. To this effect, I assessed if G9A CNAs varied significantly across the 6 CICs (consensus

immunome clusters), which have been previously described by our group to identify an immunologically “cold” tumour subtype in the LMC (Consensus Immunome Cluster 4-CIC4) of which 73% was associated with increased Wnt/β-catenin signalling (316). My analysis revealed that G9A copy number was significantly higher in CIC4 i.e. G9A copy number was positively associated with increased Wnt/β-catenin signalling in the LMC primary melanomas. This observation was concordant with the *in vitro* findings by the Fisher group. G9A was also prognostically significant: participants whose tumours had low G9A copy number had a significantly improved survival compared to those with higher G9A copy number. Moreover, an agnostic analysis revealed that Wnt/β-catenin signalling was the most significantly upregulated pathway in tumours with high G9A copy number compared to tumours with low G9A copy number. Taken together, my analyses indicated a concordance with the *in vitro* findings for a positive relationship between G9A and Wnt/β-catenin signalling.

These findings were written-up, to be part of a manuscript, which is currently under review.

6.4 **Discussion**

The approach I have used to address questions pertaining to my own PhD projects (described in Chapter 3, 4 and 5) involves the interrogation of primary melanoma-derived transcriptomic data, which was followed by *in vitro* and *in vivo* functional validation studies. On the contrary, the approach which I adopted to address the collaborative projects involved the testing of *in vitro*-derived hypotheses using the primary melanoma transcriptome. The outcome of adopting this approach was variable, depending on the question which was addressed. While some findings were concordant with the *in vitro* findings, some others showed lack of evidence to support the *in vitro* evidence.

For instance, in the case of the project assessing the effect of Midkine in the LMC primary melanomas, the association of MDK with prognosis was found to be reversed in primary melanomas (my analyses) compared to metastatic melanomas (findings from Dr Soengas’s group). Similarly, my analyses identified specific signalling pathways and imputed immune cell scores associated with *MDK* expression, which were different from those identified in metastatic melanomas. The potential explanation for these differences could be that *MDK* could have context-specific effects on the tumour microenvironment, which is amenable to the state of progression of the tumour: the metastatic melanoma TME is different from primary melanoma TME. However, this

hypothesis of context-specific role for MDK in primary versus metastatic melanomas, needs further testing.

In the case of the project testing the association of IL6/Interferon gamma signalling with sun exposure and age, the lack of evidence for an association could be owing to differences in experimental settings. However, an interesting observation which arose from this analysis was the list of genes and pathways which are differentially expressed in sun-exposed versus non sun-exposed tumours. Though this list did not include IL6/ Interferon gamma signalling (which was the expectation), among the identified pathways was the ATM pathway, which plays a role in DNA damage response and was found to be highly expressed in tumours on the head (sun exposed tumours). Thus, the identification and validation of anatomical site-specific differences in tumour gene expression could lead to potentially novel findings and hence requires further interrogation.

The G9A project revealed perhaps the strongest evidence for an *in vitro*-derived hypothesis being validated using the LMC transcriptome. This suggests the role for G9A as an oncogenic driver, whose expression is associated with melanoma prognosis as well as with the Wnt/β-catenin signalling pathway.

Taken together, the opportunity to adopt, optimise and derive conclusions from both these approaches has been a valuable learning experience in the analysis of omic data, with the view of interpretation biological questions.

Chapter 7
Discussion

Biological systems are complex. At a cellular level, the complexity of biological systems is characterised in part by dynamic changes in the 'coding components' of any given cell, i.e. changes in the structure, level and function of DNA and mRNA. The dynamics of these changes are spatiotemporally controlled. In that, the localisation and temporal sequence of events are crucial in dictating cellular function and homeostasis. In other words, changes in the genetic code ('genotype') have a consequent effect on structure and function ('phenotype'). Understanding this genotype-phenotype relationship is one of the key concepts underlying the use of gene expression technologies and pipelines to understand cellular function.

This is particularly the case in cancer, which is typically characterised by aberrations in the DNA. Perturbations in homeostatic cellular function is one of the defining characteristics of cancer. Thus, efforts to identify the genomic basis of various types of cancer is being actively explored with the aim of understanding the cause, the progression of the cancer and eventually being able to improve the treatment of the disease. One of the high throughput technologies used in cancer research, is transcriptomics. Among the earliest transcriptomic technologies developed to quantify the transcriptional programme of a cell was the microarray technology. To explain very briefly, the expression of a gene (composed of a specific target sequence) is quantified by the measure of fluorescence emitted from the hybridisation of the target sequence to a 'probe', which is a sequence of nucleic acids complementary to the target gene's sequence. Each probe is 'spotted' on a surface, producing a chip which contains multiple probes and is hence capable of quantifying an array of genes. The workflow of a typical microarray experiment is depicted in Figure 7.1. In reality, this process involves multiple steps, each of which is subject to various sources of variability and error. Thus, the use of microarrays to generate gene expression profiles from a population of cells involves multiple considerations which need to be identified and addressed in order to produce a reliable readout of the gene expression profile.

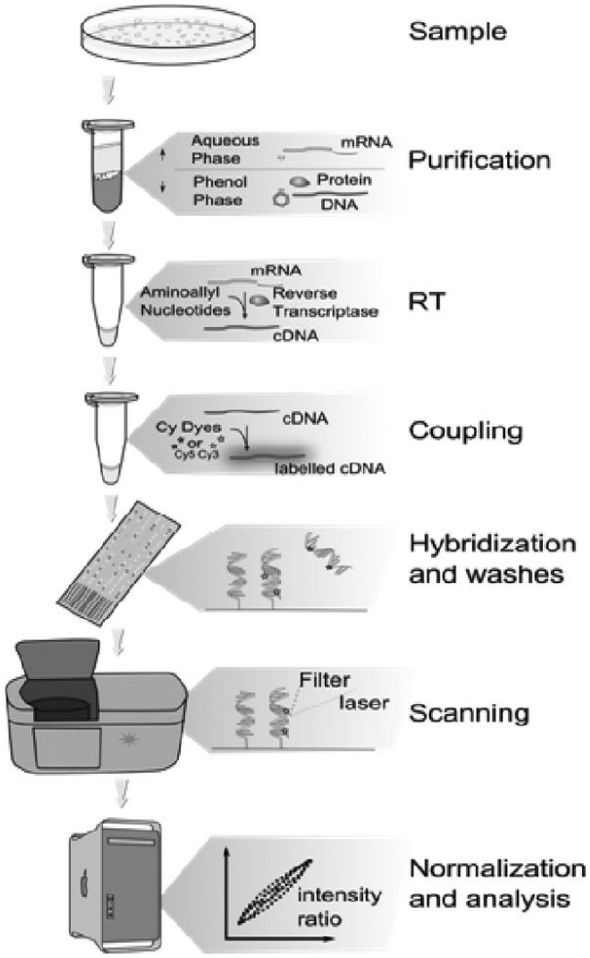


Figure 7.1: Overview of the workflow of a typical microarray experiment.

Adapted from Science behind Non-Specific Science; For Molecular Biologist & Bio-technologist, (423).

The use of microarrays in cancer research has been prevalent since the technique was first described in 1995 (424, 425). Oligonucleotide and cDNA microarrays are amongst the commonly used microarrays used to generate ‘gene expression profiles’ of tumour samples. The range of information gleaned from the gene expression profiles is dependent on the range of the probes spotted on the microarray. For instance, a ‘cancer gene panel’ could include a set of genes which are known to be significant contributors to the cancer phenotype. Alternatively, a ‘whole-genome’ array would include probes pertaining to all genes known to be expressed in the genome of a specified organism. This type of array is subject to updates, depending on the improvements made to the reference genome. Some of the common applications of these microarrays in cancer research can be broadly classified into the following categories (adopted from (426-428)), where ‘classes’ refer to groups of samples that are being interrogated.

- **Class comparison:** the comparison of gene expression profiles between classes, to identify the genes which are differentially expressed between the predefined classes. For instance, comparison of gene expression profiles from tumours that carry a particular mutation with those that don't. In the case of my thesis, identifying genes that vary significantly between tumours with a) high *VDR* compared to low *VDR* expression and b) ulceration compared to those with no ulceration
- **Class prediction:** the prediction of sample sub-groups based on their gene expression profiles, into biologically or clinically relevant groups i.e. predicting if a sample belongs to a particular group depending on its gene expression profile**.** For instance, using a gene expression signature which best predicts response to therapy in phase II studies (gene signature varies between responders and non-responders) to predict response in phase III studies.
- **Class discovery:** identification of discrete subsets (within a set of samples) based on the gene expression profiles. This is often used to estimate disease heterogeneity on the genomic level. For instance, identification of subgroups of patients whose tumours share a similar gene expression profile.

These approaches have been proven to be powerful tools in both basic and translational research. However, like every experimental technique it comes with its own set of drawbacks, which need to be taken into consideration when using microarray data to address biologically-relevant questions. The LMC transcriptome was generated using the Illumina WG-DASL-HTv12.4, from FFPE-derived RNA samples. The following table (Table 7.1) outlines the general issues associated with microarray analysis, along with the steps implemented to circumvent the different potential issues while processing and analysing the LMC transcriptome.

Table 7.1: Common issues associated with microarray experiments and the actions implemented to resolve issues

Common issues associated with microarray experiments	Action implemented to resolve this issue during the generation of the LMC transcriptome
"Noise" in the data generated requiring normalisation whilst retaining biological differences. "Noise" might occur due to variation between samples: • As a result of variation in age of the block (RNA degradation) • Differences between different laboratory personnel's techniques in extracting RNA • Experimental variability due to batch effect • Other tissue-specific technical issues, as discussed below	• The associations between top principal components and technical variables such as batch, age of FFPE block and RNA concentration were assessed. • The technical variables found to be associated with the top components were 'adjusted-out'. • Outliers were detected using normalised full intensity plots
Mapping probes to genes accurately, while accounting for probes which hybridise with low specificity among samples.	I have implemented a 'probe to gene' mapping strategy which ensures that probes which do not detect multiple splice forms (if they exist for a given gene) and are not reliably detected across the majority of the LMC samples, are not used for analysis.
Identifying biologically significant differences between classes rather than 'background' differences	• Use of stringent multiple correction thresholds to reduce the probability of the finding being by chance. • *In silico* validation strategies as well as in IHC and/or *in vitro* validation of findings • Validation in different data sets to reduce overfitting or bias of selection of tumours

In the case of the LMC transcriptome, the primary reason for electing to use the Illumina WG-DASL-HTv12.4 microarray was because the LMC tissue samples were formalin fixed. The main advantage of using FFPE melanoma samples is the possibility of sampling melanoma samples which would otherwise be too small to be sampled from cryopreservation. The disadvantage of FFPE-based sampling is that the formalin-fixing process is known to degrade and produce low yields of RNA from source tissue (429). The Illumina DASL array is designed to interrogate partially degraded RNA as well as intact RNA (430) and hence was the chosen to generate the LMC transcriptome.

In addition to the above tabulated 'common considerations' which were addressed during the generation of the LMC transcriptome, there were specific sources of error, as summarized below:

- Participants were recruited at different time periods, so the corresponding FFPE tumour samples were of varying age. The age of the FFPE tumour block could dictate the degree of RNA degradation
- Melanin, which is abundant in melanomas, is known to inhibit the polymerase activity, thus affecting the assay readout.
- Differences in technique between technicians extracting RNA, although standard operating procedure was used.

The LMC transcriptome was generated from tumour cores (TMA-derived), meaning that the transcriptomic data represented a 'core' region of the tumours. Though this approach was performed to reduce bias, by selecting to core the least inflamed part of the tumour, it involves the following issues:

- Bias towards sampling thicker tumours: if the tumours were not thick enough to 'punch a hole' using the TMA needle, then these tumours were not selected.
- In the 'thick' tumours which permitted sampling, it was not always possibly to sample from the thickest part of the tumour, since the tissue was not homogenously stable throughout the block
- Necrotic tumours were not sampled and hence are not represented in the LMC transcriptome

Having taken into account the potential issues associated with microarray data analysis, I have primarily used microarray-based gene expression profiles from the LMC primary melanomas along with clinical data, with the view of addressing specific aspects of melanoma biology. The general 'pipeline' I have used throughout my thesis is pictorially depicted in Figure 7.2. Depending on the research question to be addressed, I modified this workflow to 'customise' it.

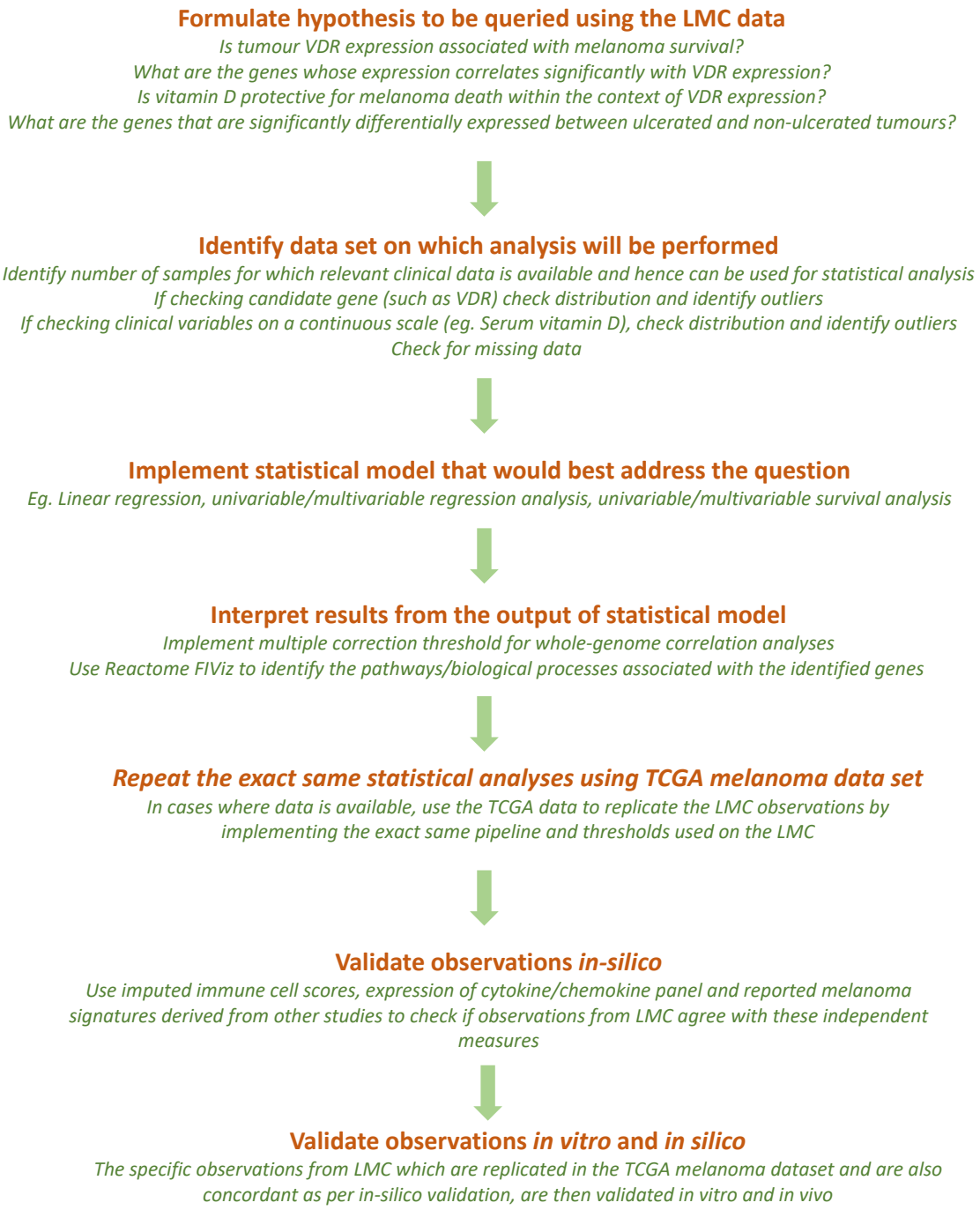


Figure 7.2: Overview of analysis pipeline which I have used in my PhD projects

7.1 Bioinformatics-based approaches used in this thesis

The primary approach I have taken with regards to biological interpretation of a given set of genes was to perform pathway enrichment analysis using Reactome FIViz. This step was of importance because the pathways identified by enrichment analysis informed the subsequent steps of my analysis. For instance, the enrichment analyses performed on genes correlated with tumour *VDR* expression, the identified pathways instructed the hypothesis to be tested *in vitro*and *in vivo*. Thus, it was necessary to choose an enrichment tool/database which enabled the implementation of stringent thresholds, thus minimising the change of false discovery. Upon comparing multiple pathway/ gene set enrichment tools such as STRING, Metacore and KEGG, I elected to use Reactome FIViz owing to the fact that it was based on the Functional Interaction network. The FI network is composed of pairwise gene-gene/protein-protein interaction annotations from both high-coverage, low-reliability and low-coverage, highly-curated databases. Performing enrichment analysis on the 'background' of this FI network enabled me to reliably identify the pathways enriched for in my genes of interest. However, it is to be noted that I have used lists of genes to perform pathway enrichment. This does not account for the variation in expression of these genes. For instance, if a set *n* genes are identified by Reactome FIViz to be enriched for a *Pathway x*, this does not account for the variable contribution of each of the *n* genes. One of the solutions for this issue is the use of another feature of Reactome FIViz: the 'Gene Expression analysis'. This feature allows the user to submit a particular set of genes and their corresponding gene expression, which are used to identify sub-networks ('Modules') of genes that are known not only to be functionally related (as per current literature) but also are highly correlated with each other in the input data set. This is done using Markov Clustering. Furthermore, every sample in the data set can be assigned a Module Score (mean gene expression value for all genes in a given module) enabling estimation of a module-specific impact on melanoma survival i.e. if a Module is protective for melanoma death. Though I have explored this approach in detail, I have not used it for the analyses described in this thesis.

I have also used reported melanoma signatures to classify the LMC primary melanoma transcriptomes. These signatures are composed of biologically-relevant sets of genes and were derived from datasets independent from the LMC. This approach enabled the 'replication' of the described molecular phenotypes. While this approach provides valuable insight with regards to the broadly classifying the LMC transcriptomes, it did not offer much insight in terms of the composite pathways. For instance, the overlap between 'pigmentation' and 'proliferation' subgroup derived from the Lund melanoma subtypes was not evident. It was owing to this reason that I used

these signature-derived molecular subtypes to validate my findings, rather than as my primary mode of hypothesis generation.

I have used imputed immune cell scores in my thesis, with the view of identifying the specific immune cell components associated with a particular feature. For instance, if immune-related pathways were identified agnostically to be correlated with a particular feature (say *VDR* expression, ulceration status etc), then I used the imputed immune cells scores to identify which particular immune cells could be associated. This has been a useful approach, but comes with a caveat, which is the process of generating the immune cells scores themselves. While the approach described in Pozniak et al is largely robust, issues such as lack of inverse correlation between cell types with known inhibitory function (such as T-regs and MDSCs) and T-cells was present. Nevertheless, the immune cell scores were a valuable measure for hypothesis validation.

In Chapter 5, which focuses on understanding the transcriptomic basis of ulceration, I have presented the findings produced by adopting the 'workflow' described in Figure 7-B. However, in addition to this approach, I also elected to use a machine learning-based approach, which I have not described in this thesis, owing to paucity of time in performing and interpreting the results. However, the rationale for choosing to use the machine learning-based approach are:

1. The necessity to use a statistically robust approach which takes into account the non-linearity of biological data (in this case: the tumour transcriptomic data) by implementing a non-reductionist approach.
2. In terms of validation, the largest primary melanoma data set besides the LMC, is the TCGA data. However, ulceration status is not well documented for these tumours. So, validating findings from the LMC in an independent cohort can prove to be difficult. Hence, an approach which would identify the genes/CNAs which best predict ulceration status could circumvent this issue. Combined with validation of the identified nodal genes using IHC, this approach could identify important genomic features associated with melanoma ulceration.
3. The use of the machine learning-approach enables the assessment of the ulceration phenomenon on the level of both gene expression and CNAs i.e, the combined data of gene expression and CNAs were used to interrogate and assess the most significant features contributing to ulceration.

7.2 Vitamin D signalling in the LMC

The association of serum vitamin D levels with stage at diagnosis of melanoma and survival has been reported in epidemiological studies where higher serum vitamin D was associated with thinner tumours and better survival (169, 288, 292). *In-vitro* studies also indicated an anti-proliferative response in melanoma cell lines to vitamin D treatment. More recently, a study by Weinstein et al assessed the effect of serum 25(OH)D years prior to cancer diagnosis on survival after cancer diagnosis of multiple cancer types. The findings from this study were concordant with previous observations, in that: melanoma patients with higher 25(OH)D prior to diagnosis experienced significantly longer survival after cancer diagnosis (431). Though there are multiple studies that have looked at the association of serum vitamin D with cancer incidence and/or mortality, there are relatively fewer studies which have assessed the effect of vitamin D supplementation on cancer risk. To this effect, a recent study reported findings from a randomised clinical trial of vitamin D supplementation in adults in New Zealand. Participants received 100,000 IU of vitamin D_3 monthly for a median of 3.3 years. The aim of this study was to assess if vitamin D supplementation contributes to cancer prevention. This study reported that the frequency of cancer diagnoses was not statistically significant between participants who received the vitamin D bolus (163 cancer diagnoses out of 2550 participants: 6.4%) and placebo (165 cancer diagnoses out of 2558 participants: 6.5%). The authors conclude that based on their findings, high-dose vitamin D supplementation does not contribute to cancer prevention. However, they acknowledge that different dosages and duration of intake could have an effect i.e. the administered vitamin D bolus and 3 years of follow may not be suitable to influence physiologically relevant levels. The study also does not allude to number of participants who were deficient for serum vitamin D at the start of the study.

Nevertheless, the necessity to understand the role of vitamin D signalling in melanomas is owing to the following reasons:

- Sun avoidance in melanoma patients is a serious issue and clarifying the role of vitamin D can help drive lifestyle changes that could benefit the patient.
- The concerns around the precise definitions of physiologically relevant levels of ‘low’ and ‘high’ serum vitamin D are not fully understood and need to be clarified.
- If proven to be efficient, vitamin D could satisfy the necessity for low-toxicity adjuvants in melanoma therapy

In the case of my PhD project, the LMC dataset was an ideal data set to explore the complex relationship between vitamin D and melanoma owing to the following reasons:

i) The availability of serum vitamin D and tumour-derived *VDR* expression at diagnosis enabled me to study the effects of both the ligand and the receptor, being able to explore both the individual and synergistic effects.

ii) The serum vitamin D levels, though available only at a single time point: at diagnosis, was adjusted for seasonal effects which reduced the effect of seasonal variations on comparisons between individuals and because of the hypothesis that the trough level might be more important.

iii) Well curated clinical data pertaining the tumours such as Breslow thickness, mitotic rate and TILs enabled me to estimate of any of the observed effects on the transcriptomic level could be confounded by any of these variables.

iv) Data pertaining to patient survival, in particular the Melanoma Specific Survival information, enabled me to assess the prognostic significance of variables such as *VDR* expression and serum vitamin D levels on death from the cancer rather than overall death rates. The clinical data mentioned above, also enabled me to estimate if the effects on survival were independent of these factors (using multivariate analyses). Low vitamin D levels are reported to be associated with a number of different health problems so that distinguishing melanoma deaths from any cause of death was very important.

v) Finally, the sample size of the data set is the largest melanoma transcriptomic data set to date. This enabled performance of stratified analyses (for example, the vitamin D-VDR subgroup analysis) without a significant adverse effect on sample size.

To conclude my findings on this project: this project integrates clinical, transcriptomic, histopathological and functional validation data to provide a novel insight into vitamin D-*VDR* signalling in melanoma. I report evidence that, in addition to vitamin D-mediated protective effect which is apparent in a subset of participants, ligand-independent *VDR*-expression bestows a prognostic benefit for melanoma patients with evidence to support a causal relationship involving inhibition of Wnt/β-catenin signalling and increasing immune cell infiltration.

The primary reasons for studying ulceration in the LMC primary melanomas, as part of my thesis was to gain a deeper understanding of the biological basis of ulceration.

This approach has been relatively (relative to the vitamin D-VDR project) exploratory and is work in progress.

The collaborative initiatives which I have undertaken during my PhD have been, in a manner of speaking, the opposite approach to the ‘pipeline’ I’ve used in my projects. In that, I have sought evidence for hypotheses generated from *in-vitro* and *in-vivo* experiments, in the LMC primary transcriptomes.

Chapter 8

Appendix A

A.1 Tables of output from enrichment analyses performed in Chapter 3

In the case of tables containing output from enrichment analyses, pathways are presented based on variable FDR thresholds. The reason for this being, repetitive pathways appear at lower thresholds and hence the necessity to drop the display of these pathways.

Table T3-2: Differential expression of 154 cytokine and chemokine genes compared between the high and low vitamin D participants in the intermediate-VDR group.

P-value derived from T-test and FDR (False Discovery Rate) derived from Benjamini-Hochberg method for multiple correction.

Cytokine/ chemokine genes	P-value	Mean expression in high vitamin D group	Mean expression in low vitamin D group	FDR
CXCL2	0.0002	7.801	8.363	0.032
CCL3L3	0.0035	7.763	8.377	0.231
IL24	0.0078	7.796	8.431	0.346
IL2RA	0.0197	7.697	8.137	0.656
IL16	0.0324	7.872	7.669	0.862
CCL5	0.0357	7.914	8.282	0.792
IL10RA	0.0382	7.824	8.164	0.727
CCL8	0.0467	7.804	8.222	0.776
IL1F10	0.0758	8.421	8.998	1.000
CXCL1	0.0774	7.888	8.147	1.000
CCL7	0.0849	7.980	7.778	1.000
CXCR4	0.0868	7.812	8.080	0.962
IL12RB2	0.0871	7.806	7.457	0.891
CXCR6	0.0888	7.668	8.119	0.843
IL6	0.0934	7.641	8.052	0.828
IL21R	0.1035	7.999	8.343	0.860
IL2RG	0.1093	7.921	8.237	0.855
IL7R	0.1165	7.888	8.258	0.861
IL32	0.1217	7.799	8.049	0.852
IL22	0.1234	7.883	7.756	0.820

Cytokine/ chemokine genes	P-value	Mean expression in high vitamin D group	Mean expression in low vitamin D group	FDR
CCL3L1	0.1271	7.984	8.174	0.805
CCR3	0.1274	7.856	7.749	0.770
XCR1	0.1305	7.883	8.061	0.754
IL18RAP	0.1313	8.034	8.355	0.728
CCR7	0.1465	7.992	8.390	0.780
CCL4L1	0.1477	7.815	8.127	0.756
CCL2	0.1518	7.819	7.972	0.748
IL13RA2	0.1612	7.820	7.976	0.766
IL21	0.1631	7.956	7.823	0.748
IL23A	0.1656	7.864	7.686	0.734
IL1R1	0.1670	7.842	7.995	0.716
IL31RA	0.1699	7.884	7.778	0.706
CCR4	0.1726	7.964	8.262	0.695
IL1RAPL2	0.1819	7.955	7.796	0.712
IL17A	0.1851	7.814	7.905	0.703
IL1RAP	0.1963	7.768	7.594	0.725
IL1A	0.2010	7.933	8.108	0.723
IL17RD	0.2014	7.804	7.875	0.705
CXCL9	0.2041	8.208	8.594	0.696
IL12RB1	0.2066	7.823	8.127	0.687
IL15RA	0.2068	7.968	8.148	0.671
CXCR3	0.2089	8.014	8.242	0.661
CX3CR1	0.2182	7.904	8.102	0.675
CXCL5	0.2488	7.880	7.986	0.752
CCR1	0.2548	7.841	7.967	0.753
IL15	0.2675	7.857	8.032	0.773
IL34	0.2703	7.856	8.052	0.765
IL22RA1	0.2796	7.905	7.823	0.775
IL19	0.2838	7.858	7.754	0.770
CCL3	0.2938	7.830	8.050	0.782
IL2	0.2941	7.891	7.806	0.767
IL3	0.2978	7.809	7.892	0.762
CXCL3	0.3075	7.776	7.889	0.772
IL13RA1	0.3148	7.759	7.846	0.775
CCL1	0.3189	7.834	7.908	0.771
IL27	0.3209	7.811	7.919	0.762
CCR6	0.3246	7.832	7.839	0.757
IL5RA	0.3248	7.878	7.802	0.745
IL2RB	0.3462	7.924	8.104	0.780

Cytokine/ chemokine genes	P-value	Mean expression in high vitamin D group	Mean expression in low vitamin D group	FDR
IL9	0.3503	7.841	7.771	0.777
IL17RE	0.3586	7.706	7.540	0.782
CXCL16	0.3663	7.889	7.956	0.786
CCL21	0.3704	7.633	7.428	0.782
IL11	0.3762	7.980	7.875	0.782
IL17B	0.3777	7.959	7.794	0.773
IL31	0.4015	7.771	7.838	0.809
CCL23	0.4087	8.039	8.215	0.811
IL22RA2	0.4295	7.876	7.818	0.840
IL1RN	0.4297	7.948	8.141	0.828
CXCL10	0.4380	7.803	8.032	0.832
CXCL11	0.4445	8.385	8.583	0.833
IL33	0.5049	7.783	7.903	0.933
IL17RA	0.5178	7.741	7.810	0.943
XCL1	0.5224	7.797	7.751	0.939
CCR2	0.5280	7.931	7.843	0.936
CCL14	0.5399	7.959	8.095	0.945
CCRL2	0.5525	7.842	7.887	0.954
IL1R2	0.5819	7.915	7.868	0.992
CCL18	0.5943	7.883	7.994	1.000
CCL28	0.5973	7.827	7.867	0.993
PITPNM3	0.6013	7.900	7.975	0.987
IL5	0.6091	7.810	7.849	0.988
CCR8	0.6202	7.960	8.016	0.994
IL25	0.6294	7.832	7.868	0.997
CCL26	0.6305	7.950	8.001	0.987
IL4	0.6348	7.782	7.749	0.982
IL17D	0.6354	7.590	7.692	0.971
CCL16	0.6383	7.965	7.912	0.965
CCL25	0.6447	7.898	7.963	0.963
CCR5	0.6496	7.838	7.796	0.960
IL12B	0.6547	7.959	8.001	0.957
CCR10	0.6664	7.797	7.862	0.963
IL20RA	0.6687	8.069	8.010	0.956
CCL22	0.6731	8.051	7.957	0.952
IL27RA	0.6751	7.907	7.839	0.945
CCL24	0.6791	7.912	7.986	0.941
IL18R1	0.6793	7.915	7.840	0.931
IL6R	0.6907	7.797	7.837	0.937

Cytokine/ chemokine genes	P-value	Mean expression in high vitamin D group	Mean expression in low vitamin D group	FDR
IL10	0.7099	7.832	7.810	0.954
CCL13	0.7101	7.879	7.961	0.944
IL20RB	0.7218	8.104	8.029	0.950
CCL19	0.7349	7.825	7.920	0.958
IL3RA	0.7390	7.858	7.826	0.954
IL12A	0.7419	7.826	7.856	0.949
CCL17	0.7424	7.965	7.916	0.940
IL18	0.7493	7.918	7.891	0.940
IL7	0.7581	7.850	7.872	0.942
IL17RC	0.7602	7.786	7.757	0.936
IL13	0.7602	7.804	7.774	0.928
IL11RA	0.7821	7.724	7.764	0.946
CX3CL1	0.7830	7.789	7.755	0.938
IL1B	0.7935	7.726	7.778	0.942
CXCL14	0.7941	7.748	7.694	0.935
CXCL17	0.8037	8.132	8.092	0.938
IL6ST	0.8080	7.772	7.796	0.934
CCL20	0.8113	8.030	7.993	0.930
IL17F	0.8160	7.900	7.927	0.928
CXCL12	0.8241	7.843	7.873	0.929
IL1RL1	0.8377	7.846	7.802	0.936
IL17C	0.8429	7.965	7.944	0.934
IL26	0.8694	7.814	7.827	0.956
CCL11	0.8755	7.968	7.989	0.954
CXCL13	0.8986	8.132	8.161	0.972
IL1RAPL1	0.9049	8.031	8.059	0.971
IL10RB	0.9173	7.739	7.728	0.976
CXCR5	0.9190	8.328	8.350	0.970
CXCR1	0.9415	7.843	7.850	0.986
CCR9	0.9469	7.860	7.855	0.984
IL1RL2	0.9570	7.954	7.946	0.987
IL20	0.9805	7.885	7.887	1.000
IL17RB	0.9812	7.839	7.836	0.996
CCL27	0.9844	8.504	8.509	0.992
CXCL6	0.9989	7.922	7.922	0.999

Table T3-3: Full list of pathways enriched for 1383 genes negatively correlated (at FDR<0.00001) with tumour *VDR* expression. Output from Reactome FIViz

The source database for each of the enriched pathways are indicated by a letter in parentheses after each pathway gene set name. The source database annotations are: C - CellMap, R – Reactome, K – KEGG, N – NCI PID, P - Panther, and B – BioCarta.

The description of column headers are: *Ratio of protein in gene set*: ratios of numbers of genes contained in pathways to total genes in the Reactome FI network ; Number of protein in gene set: numbers of genes in pathways; *Protein from network*: numbers of hit genes from the query gene list; *Nodes:* nodal genes from my input query gene list, which 'matches' with the genes in a particular pathway; P-value was estimated by Reactome FIViz using hypergeometric test and the corresponding FDR was estimated by Benjamini Hochberg multiple correction.

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
Cadherin signalling pathway(P)	0.0098	100	16	2.81E-07	1.82E-04	PCDHA3, PCDHA9, PCDHA6, PCDHGA11, PCDHB2, PCDHB5, PCDH7, CTNNA2, PCDHB12, PCDHB11, PCDHB10, PCDHB16, PCDHGB7, PCDHGB3, PCDHGA8, PCDHGA3
Wnt signalling pathway(P)	2.62E-02	2.68E+02	25	3.37E-06	1.09E-03	PRKCA, CTBP2, PCDHA3, PCDHA9, PCDHA6, PCDHGA11, PCDHB2, PCDHB5, EN2, TLE1, BCL9, SMARCE1, PCDH7, APC, NLK, CTNNA2, PCDHB12, PCDHB11, PCDHB10, PCDHB16, PCDHGB7, PCDHGB3, PCDHGA8, PCDHGA3, TGFBR1
Meiotic recombination(R)	3.10E-03	3.20E+01	8	1.27E-05	2.75E-03	NBN, MSH4, BLM, CDK4, BRCA1, RAD51C, RPA1, MLH3
Fanconi anemia pathway(N)	4.40E-03	4.50E+01	9	2.12E-05	3.43E-03	RMI1, USP1, NBN, BLM, FANCC, CHEK1, BRCA1, RPA1, RAD1
Mitochondrial translation(R)	8.70E-03	8.90E+01	12	4.53E-05	5.62E-03	MRPL10, MRPL11, MRPL47, MRPL50, MRPL2, MRPL1, GADD45GIP1, MRPS18B, MRPS11, MRPS23, MRPS35, MRPS31
SRP-dependent cotranslational protein targeting to membrane(R)	1.03E-02	1.05E+02	13	5.21E-05	5.62E-03	RPL34, RPL29, RPL28, RPS4Y1, SSR1, SRP9, RPL13, RPL7A, RPL36A, SEC11A, SEC11C, RPL35A, SRPRB
Assembly of the primary cilium(R)	1.68E-02	1.72E+02	17	6.35E-05	5.81E-03	CEP72, PLK1, WDR60, CEP57, TCTN3, TCTN1, NINL, EXOC8, IFT140, CC2D2A, KIF24, SSNA1, RPGRIP1L, KIF3A, AHI1, TMEM216, ASAP1
Cell cycle(K)	1.21E-02	1.24E+02	14	7.17E-05	5.81E-03	PRKDC, PLK1, CDC7, CDK4, TFDP2, MCM2, TTK, E2F3, CDC23, CHEK1, CCNB3, SMAD2, BUB1, WEE1

Table T3-4: Full list of pathways enriched for 2025 genes positively correlated with tumour VDR expression (FDR<0.00001). Output from Reactome FIViz

The source database for each of the enriched pathways are indicated by a letter in parentheses after each pathway gene set name. The source database annotations are: C - CellMap, R – Reactome, K – KEGG, N – NCI PID, P - Panther, and B – BioCarta.

The description of column headers are: *Ratio of protein in gene set*: ratios of numbers of genes contained in pathways to total genes in the Reactome FI network; Number of protein in gene set: numbers of genes in pathways; *Protein from network*: numbers of hit genes from the query gene list; *Nodes:* nodal genes from my input query gene list, which 'matches' with the genes in a particular pathway; P-value was estimated by Reactome FIViz using hypergeometric test and the corresponding FDR was estimated by Benjamini Hochberg multiple correction

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
Extracellular matrix organization(R)	0.0243	248	67	1.11E-16	2.28E-14	HSPG2, COL13A1, ADAMTS14, FN1, COL1A2, TPSAB1, MMP10, MMP11, MMP13, MMP19, COL3A1, EFEMP1, COL17A1, ITGB4, ITGB2, ITGAL, ITGAX, ITGB8, ITGB6, ITGA1, ITGA8, ITGA5, COL15A1, FURIN, ICAM2, CTSS, CTSG, LOXL1, PCOLCE, COL5A1, COL5A3, FBN1, LAMC2, MMP7, MMP1, MMP2, MMP3, MMP8, MMP9, LAMA2, LAMA3, LAMB3, COL7A1, SERPINE1, COL12A1, ELANE, ELN, FMOD, TNC, COL27A1, ADAM8, ADAMTS2, PECAM1, MFAP2, KLK7, CEACAM6, COMP, COL4A2, COL4A4, FBLN1, FBLN5, TGFB1, DCN, COL6A2, COL6A1, COL6A3, LUM
Staphylococcus aureus infection(K)	5.40E-03	5.50E+01	32	1.11E-16	2.28E-14	FPR3, C2, C3, HLA-DRB5, HLA-DRB4, HLA-DRB3, HLA-DRB1, HLA-DPA1, PTAFR, HLA-DMA, HLA-DMB, HLA-DPB1, HLA-DOA, HLA-DOB, HLA-DRA, ITGB2, ITGAL, FCGR3A, DSG1, C1S, C1R, FCGR2B, SELPLG, MASP1, CFB, CFD, CFH, HLA-DQA2, HLA-DQA1, C1QB, C1QA, SELP
Cytokine-cytokine receptor interaction(K)	2.59E-02	2.65E+02	74	1.11E-16	2.28E-14	IL20RB, NGFR, EGFR, IL12RB1, CXCL13, CXCL14, TNFRSF4, TNFRSF11B, TNFRSF9, TNFRSF10D, CXCL10, CXCL11, CXCL12, FLT3, FLT4, FASLG, TNFSF13B, TNFRSF1B, IL2RG, CCL5, CCL3, CCL2, IL2RA, IL2RB, TNFRSF6B, TNFSF10, TNFSF14, TNFSF13, CD27, CD40, TNFRSF13B, IL24, CCR7, CCR4, CCR2, TNFRSF12A, IL18, OSMR, IL1A, IL1B, IL18R1, PDGFRA, INHBA, IL7R, CCL14, CCL13, CCL19, CCL18, CCL24, CCL22, CCL21, IL10RA, CCL27, CSF3R, CSF1, TNF, CSF1R, CX3CR1, IL21R, IL15RA, IL1R1, FLT3LG, LIF, CXCR6, CXCR3, IL18RAP, CSF2RB, TGFB1, BMPR1B, CXCL9, CXCL2, IFNE, LTA, LTB
Hematopoietic cell lineage(K)	8.50E-03	8.70E+01	38	1.11E-16	2.28E-14	HLA-DRB5, HLA-DRB4, HLA-DRB3, HLA-DRB1, HLA-DRA, FLT3, ITGA1, ITGA5, IL2RA, ANPEP, CD1C, CD1B, CD1A, CD19, CD24, CD3G, CD3E, CD3D, CD38, CD37, CD36, CD33, IL1A, IL1B, CD8B, CD8A, IL7R, MS4A1, CD2, CD5, CD7, CD9, CSF3R, CSF1, TNF, CSF1R, IL1R1, FLT3LG

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
Cell adhesion molecules (CAMs)(K)	1.39E-02	1.42E+02	46	1.67E-15	2.73E-13	CLDN1, CLDN5, HLA-DRB5, HLA-DRB4, HLA-DRB3, HLA-DRB1, CTLA4, HLA-DPA1, HLA-DMA, HLA-DMB, HLA-DPB1, SIGLEC1, HLA-DOA, HLA-DOB, CD274, PTPRC, HLA-DRA, ITGB2, ITGAL, SPN, ITGB8, ITGA8, PVRL2, PVRL1, ICAM2, CD226, CD40, SELPLG, CD80, CD8B, CD8A, CD2, CD6, ICOS, HLA-DQA2, HLA-DQA1, CDH5, PECAM1, TIGIT, ICOSLG, HLA-B, HLA-F, PDCD1LG2, SELE, SELP, SELL
Osteoclast differentiation(K)	1.28E-02	1.31E+02	42	4.00E-14	5.44E-12	SPI1, STAT1, NCF2, NCF4, TNFRSF11B, SOCS3, SOCS1, NFKB2, PIK3CG, FCGR3A, LCP2, FCGR2B, BTK, PIK3R5, PPP3CC, IL1A, IL1B, RELB, CYBA, OSCAR, TYROBP, MAPK13, MAPK11, JUNB, PLCG2, NFATC2, FOSL2, FOSL1, FOSB, CSF1, TNF, CSF1R, MAP3K14, LCK, IL1R1, LILRA1, LILRB2, LILRB4, CAMK4, SIRPG, TGFB1, BLNK
Graft-versus-host disease(K)	4.00E-03	4.10E+01	24	7.07E-14	7.35E-12	HLA-DRB5, HLA-DRB4, HLA-DRB3, HLA-DRB1, HLA-DPA1, HLA-DMA, HLA-DMB, KIR3DL2, HLA-DPB1, HLA-DOA, HLA-DOB, HLA-DRA, FASLG, GZMB, PRF1, IL1A, IL1B, CD80, HLA-DQA2, HLA-DQA1, TNF, HLA-B, HLA-F, KLRD1
Costimulation by the CD28 family(R)	6.20E-03	6.30E+01	29	7.21E-14	7.35E-12	LYN, HLA-DRB5, HLA-DRB4, HLA-DRB3, HLA-DRB1, CTLA4, HLA-DPA1, HLA-DPB1, CD274, HLA-DRA, CDC42, PTPN6, CD247, CD3G, CD3E, CD3D, PAK1, CD80, ICOS, HLA-DQA2, HLA-DQA1, HLA-DQB2, GRAP2, BTLA, MAP3K14, LCK, ICOSLG, PDCD1LG2, VAV1
Interferon gamma signalling(R)	7.20E-03	7.40E+01	31	1.14E-13	1.04E-11	HLA-DRB5, HLA-DRB4, HLA-DRB3, HLA-DRB1, HLA-DPA1, PTAFR, STAT1, HLA-DPB1, SOCS3, SOCS1, HLA-DRA, CIITA, PTPN6, IFI30, TRIM22, IRF8, IRF5, IRF6, HLA-DQA2, HLA-DQA1, HLA-DQB2, PML, HLA-B, HLA-F, OASL, OAS1, OAS2, OAS3, CAMK2D, GBP2, GBP1
Inflammatory bowel disease (IBD)(K)	6.40E-03	6.50E+01	29	1.54E-13	1.26E-11	GATA3, HLA-DRB5, HLA-DRB4, HLA-DRB3, HLA-DRB1, HLA-DPA1, HLA-DMA, HLA-DMB, STAT4, IL12RB1, STAT1, HLA-DPB1, RORC, RORA, HLA-DOA, HLA-DOB, HLA-DRA, IL2RG, IL18, IL1A, IL1B, IL18R1, HLA-DQA2, HLA-DQA1, TNF, NOD2, IL21R, IL18RAP, TGFB1

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
IL12-mediated signalling events(N)	6.00E-03	6.10E+01	28	2.07E-13	1.53E-11	HLA-DRB1, STAT4, IL12RB1, STAT1, SOCS1, NFKB2, HLA-DRA, FASLG, IL2RG, CCL3, GZMA, GZMB, IL2RA, IL2RB, CD247, CD3G, CD3E, CD3D, IL18, IL1B, RELB, IL18R1, CD8B, CD8A, LCK, IL1R1, GADD45G, IL18RAP
Tuberculosis(K)	1.73E-02	1.77E+02	48	2.52E-13	1.71E-11	C3, HLA-DRB5, HLA-DRB4, HLA-DRB3, HLA-DRB1, CASP10, HLA-DPA1, HLA-DMA, HLA-DMB, STAT1, HLA-DPB1, HLA-DOA, HLA-DOB, PLA2R1, CORO1A, HLA-DRA, CIITA, ITGB2, FCGR3A, ITGAX, FCGR2B, CD209, CTSS, PPP3CC, FCER1G, IL18, IL1A, IL1B, CD74, IRAK2, MAPK13, MAPK11, BCL10, HLA-DQA2, HLA-DQA1, MRC2, IL10RA, CARD9, TNF, CLEC7A, TRADD, LSP1, NOD2, CALML5, CALML3, TGFB1, CASP8, CAMK2D
Intestinal immune network for IgA production(K)	4.60E-03	4.70E+01	24	1.22E-12	7.66E-11	HLA-DRB5, HLA-DRB4, HLA-DRB3, HLA-DRB1, HLA-DPA1, HLA-DMA, HLA-DMB, HLA-DPB1, HLA-DOA, HLA-DOB, CXCL12, HLA-DRA, TNFSF13B, TNFSF13, CD40, TNFRSF13B, CD80, ICOS, HLA-DQA2, HLA-DQA1, MAP3K14, ICOSLG, IL15RA, TGFB1
Type I diabetes mellitus(K)	4.20E-03	4.30E+01	23	1.42E-12	8.25E-11	HLA-DRB5, HLA-DRB4, HLA-DRB3, HLA-DRB1, HLA-DPA1, HLA-DMA, HLA-DMB, HLA-DPB1, HLA-DOA, HLA-DOB, HLA-DRA, FASLG, GZMB, PRF1, IL1A, IL1B, CD80, HLA-DQA2, HLA-DQA1, TNF, HLA-B, HLA-F, LTA
Rheumatoid arthritis(K)	8.70E-03	8.90E+01	32	2.44E-12	1.32E-10	HLA-DRB5, HLA-DRB4, HLA-DRB3, HLA-DRB1, CTLA4, HLA-DPA1, HLA-DMA, HLA-DMB, HLA-DPB1, HLA-DOA, HLA-DOB, CXCL12, HLA-DRA, ITGB2, ITGAL, TNFSF13B, CCL5, CCL3, CCL2, TNFSF13, IL18, IL1A, IL1B, CD80, MMP1, MMP3, HLA-DQA2, HLA-DQA1, CSF1, TNF, TGFB1, LTB
Allograft rejection(K)	3.60E-03	3.70E+01	21	4.49E-12	2.29E-10	HLA-DRB5, HLA-DRB4, HLA-DRB3, HLA-DRB1, HLA-DPA1, HLA-DMA, HLA-DMB, HLA-DPB1, HLA-DOA, HLA-DOB, HLA-DRA, FASLG, GZMB, PRF1, CD40, CD80, HLA-DQA2, HLA-DQA1, TNF, HLA-B, HLA-F

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
TCR signalling in naïve CD4+ T cells(N)	6.60E-03	6.70E+01	27	1.03E-11	4.96E-10	HLA-DRB1, SHC1, PTPRC, HLA-DRA, LCP2, FYB, CDC42, PTPN6, CD247, RASSF5, CD3G, CD3E, CD3D, CD80, CARD11, RASGRP1, PRKCQ, BCL10, ITK, GRAP2, MAP3K14, LAT, LCK, INPP5D, VAV1, WAS, TRPV6
TNF signalling pathway(K)	1.08E-02	1.10E+02	34	2.82E-11	1.27E-09	JAG1, CASP10, RIPK3, SOCS3, CXCL10, PIK3CG, TNFRSF1B, BIRC3, CCL5, CCL2, PIK3R5, IL1B, IL18R1, MMP3, MMP9, MAPK13, MAPK11, CREB3L1, JUNB, CSF1, TNF, BCL3, MAP3K14, TRADD, TRAF1, NOD2, LIF, EDN1, MLKL, SELE, CXCL2, CASP7, CASP8, LTA
Antigen processing and presentation(K)	7.50E-03	7.70E+01	28	4.42E-11	1.90E-09	HLA-DRB5, HLA-DRB4, HLA-DRB3, HLA-DRB1, HLA-DPA1, HLA-DMA, HLA-DMB, KIR3DL2, HLA-DPB1, HLA-DOA, HLA-DOB, HLA-DRA, CIITA, IFI30, CTSS, CD74, CD8B, CD8A, TAPBP, HLA-DQA2, HLA-DQA1, TNF, TAP2, TAP1, HLA-B, HLA-F, KLRC3, KLRD1
Leishmaniasis(K)	7.00E-03	7.20E+01	27	5.03E-11	2.06E-09	C3, HLA-DRB5, HLA-DRB4, HLA-DRB3, HLA-DRB1, HLA-DPA1, HLA-DMA, HLA-DMB, STAT1, HLA-DPB1, NCF2, NCF4, HLA-DOA, HLA-DOB, HLA-DRA, ITGB2, FCGR3A, PTPN6, IL1A, IL1B, CYBA, MAPK13, MAPK11, HLA-DQA2, HLA-DQA1, TNF, TGFB1
Pathways in cancer(K)	3.89E-02	3.97E+02	72	6.30E-11	2.46E-09	SPI1, JUP, FN1, PPARD, EGFR, STAT1, WNT7B, WNT7A, ZBTB16, NFKB2, CXCL12, FLT3, FASLG, PIK3CG, BIRC3, CDC42, RASSF5, LPAR1, TGFA, LPAR4, PIK3R5, IGF1, RAC2, PDGFRA, TCF7, RASGRP1, RASGRP4, RASGRP3, WNT3, FGF22, WNT4, FGF11, LAMC2, MMP1, MMP2, MMP9, LAMA2, LAMA3, LAMB3, LAMB4, DAPK2, CTBP1, PLCG2, WNT10A, GLI2, PLCB3, ADCY4, RUNX1T1, CDKN2B, CSF3R, FGF1, PGF, CSF1R, PLEKHG5, TRAF1, CEBPA, PML, CBLC, EDNRA, FLT3LG, PTGER4, PTGER2, PTGER3, WNT5A, COL4A2, COL4A4, TGFB1, WNT3A, FGFR3, FGFR2, FGFR1, CASP8
Phagosome(K)	1.50E-02	1.53E+02	40	7.36E-11	2.72E-09	C3, HLA-DRB5, HLA-DRB4, HLA-DRB3, HLA-DRB1, HLA-DPA1, HLA-DMA, HLA-DMB, HLA-DPB1, NCF2, NCF4, HLA-DOA, HLA-DOB, TUBA4A, PLA2R1, CORO1A, HLA-DRA, ITGB2, FCGR3A, ITGA5, C1R, FCGR2B, CD209, CTSS, CD36, SFTPD, CYBA,

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
						HLA-DQA2, HLA-DQA1, MRC2, TAP2, TAP1, CLEC7A, HLA-B, HLA-F, MARCO, RAB7B, COMP, NOS1, THBS2
Viral myocarditis(K)	5.70E-03	5.80E+01	24	8.55E-11	2.99E-09	CXADR, HLA-DRB5, HLA-DRB4, HLA-DRB3, HLA-DRB1, HLA-DPA1, HLA-DMA, HLA-DMB, HLA-DPB1, HLA-DOA, HLA-DOB, HLA-DRA, ITGB2, ITGAL, PRF1, CD40, RAC2, CD80, LAMA2, HLA-DQA2, HLA-DQA1, HLA-B, HLA-F, CASP8
NF-kappa B signalling pathway(K)	8.90E-03	9.10E+01	30	9.22E-11	3.13E-09	LYN, TICAM1, NFKB2, CXCL12, TNFSF13B, BIRC3, BTK, TNFSF14, CD40, IL1B, RELB, PLAU, CARD11, PRKCQ, BCL10, PLCG2, CCL13, CCL19, CCL21, TNF, MAP3K14, LAT, TRADD, TRAF1, LCK, IL1R1, CXCL2, BLNK, LTA, LTB
T cell receptor signalling pathway(K)	1.02E-02	1.04E+02	32	1.20E-10	3.76E-09	CTLA4, PTPRC, PIK3CG, LCP2, CDC42, PTPN6, CD247, PIK3R5, CD3G, CD3E, CD3D, PPP3CC, PAK1, PAK6, CD8B, CD8A, CARD11, RASGRP1, MAPK13, MAPK11, PRKCQ, BCL10, ICOS, NFATC2, ITK, TNF, GRAP2, MAP3K14, LAT, LCK, CBLC, VAV1
TCR signalling in naïve CD8+ T cells(N)	5.30E-03	5.40E+01	23	1.21E-10	3.76E-09	SHC1, PTPRC, LCP2, PTPN6, CD247, PRF1, RASSF5, CD3G, CD3E, CD3D, CD80, CD8B, CD8A, CARD11, RASGRP1, PRKCQ, BCL10, GRAP2, MAP3K14, LAT, LCK, VAV1, TRPV6
HTLV-I infection(K)	2.53E-02	2.58E+02	54	1.32E-10	3.96E-09	SPI1, HLA-DRB5, HLA-DRB4, HLA-DRB3, HLA-DRB1, JAK3, HLA-DPA1, HLA-DMA, HLA-DMB, WNT7B, WNT7A, HLA-DPB1, HLA-DOA, HLA-DOB, NFKB2, HLA-DRA, ITGB2, ITGAL, PIK3CG, IL2RG, IL2RA, IL2RB, PIK3R5, NRP1, CD40, CD3G, CD3E, CD3D, PPP3CC, RELB, ETS2, CCND2, PDGFRA, WNT3, WNT4, HLA-DQA2, HLA-DQA1, WNT10A, NFATC2, FOSL1, MRAS, ADCY4, CDKN2B, TNF, MAP3K14, LCK, IL15RA, IL1R1, HLA-B, HLA-F, WNT5A, TGFB1, WNT3A, LTA
ECM-receptor interaction(K)	8.50E-03	8.70E+01	29	1.47E-10	4.27E-09	HSPG2, FN1, COL1A2, COL3A1, ITGB4, ITGB8, ITGB6, ITGA1, ITGA8, ITGA5, CD36, COL5A1, COL5A3, LAMC2, LAMA2, LAMA3, LAMB3, LAMB4, TNC, COL27A1, TNXB, COMP, COL4A2, COL4A4, COL6A2, COL6A1, COL6A3, COL6A6, THBS2

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
Primary immunodeficiency(K)	3.50E-03	3.60E+01	19	1.54E-10	4.32E-09	JAK3, PTPRC, CIITA, IL2RG, BTK, CD19, CD40, TNFRSF13B, CD3E, CD3D, CD8B, CD8A, IL7R, ICOS, TAP2, TAP1, LCK, CD79A, BLNK
Asthma(K)	2.90E-03	3.00E+01	17	4.92E-10	1.32E-08	HLA-DRB5, HLA-DRB4, HLA-DRB3, HLA-DRB1, HLA-DPA1, HLA-DMA, HLA-DMB, HLA-DPB1, HLA-DOA, HLA-DOB, HLA-DRA, CD40, FCER1A, FCER1G, HLA-DQA2, HLA-DQA1, TNF
Focal adhesion(K)	2.03E-02	2.07E+02	46	5.09E-10	1.32E-08	MYLK, FN1, COL1A2, EGFR, FLNB, FLNC, COL3A1, SHC1, ITGB4, FLT4, PIK3CG, ITGB8, ITGB6, ITGA1, ITGA8, ITGA5, BIRC3, CDC42, PIK3R5, IGF1, RAC2, PAK1, PAK6, CCND2, PDGFRA, COL5A1, COL5A3, LAMC2, LAMA2, LAMA3, LAMB3, LAMB4, TNC, COL27A1, PGF, PARVG, VAV1, TNXB, COMP, COL4A2, COL4A4, COL6A2, COL6A1, COL6A3, COL6A6, THBS2
T cell activation(P)	7.90E-03	8.10E+01	27	6.31E-10	1.58E-08	HLA-DPA1, HLA-DMA, HLA-DMB, HLA-DOA, NFKB2, PTPRC, HLA-DRA, PIK3CG, LCP2, CDC42, CD247, CD3G, CD3E, CD3D, PPP3CC, RAC2, PAK1, CD74, PRKCQ, HLA-DQA2, HLA-DQA1, NFATC2, GRAP2, LAT, LCK, VAV1, WAS
Toxoplasmosis(K)	1.16E-02	1.18E+02	33	6.64E-10	1.59E-08	HLA-DRB5, HLA-DRB4, HLA-DRB3, HLA-DRB1, HLA-DPA1, HLA-DMA, HLA-DMB, STAT1, HLA-DPB1, HLA-DOA, HLA-DOB, SOCS1, HLA-DRA, CIITA, PIK3CG, BIRC3, PIK3R5, CD40, LAMC2, LAMA2, LAMA3, LAMB3, LAMB4, MAPK13, MAPK11, HLA-DQA2, HLA-DQA1, ALOX5, LDLR, IL10RA, TNF, TGFB1, CASP8
Cell surface interactions at the vascular wall(R)	9.80E-03	1.00E+02	30	8.14E-10	1.95E-08	CXADR, LYN, FN1, SHC1, ITGB2, ITGAL, SPN, ITGAX, ITGA5, PTPN6, CD244, GRB7, FCER1G, CD48, SELPLG, CD74, PROC, MMP1, CD2, GAS6, THBD, LCK, PECAM1, INPP5D, CEACAM6, SIRPG, SELE, SELP, SELL, AMICA1
Beta1 integrin cell surface interactions(N)	6.50E-03	6.60E+01	24	1.07E-09	2.45E-08	FN1, COL1A2, COL3A1, F13A1, ITGA1, ITGA8, ITGA5, PLAU, TGM2, COL5A1, FBN1, LAMC2, LAMA2, LAMA3, LAMB3, COL7A1, TNC, COL4A4, TGFBI, PLAUR, COL6A2, COL6A1, COL6A3, THBS2

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
Natural killer cell mediated cytotoxicity(K)	1.31E-02	1.34E+02	35	1.16E-09	2.56E-08	HCST, KIR3DL2, SHC1, TNFRSF10D, ITGB2, FASLG, ITGAL, PIK3CG, FCGR3A, LCP2, GZMB, PTPN6, CD247, CD244, PRF1, ICAM2, TNFSF10, PIK3R5, PPP3CC, RAC2, FCER1G, CD48, PAK1, MICA, TYROBP, PLCG2, NFATC2, TNF, LAT, LCK, HLA-B, SH2D1A, VAV1, KLRC3, KLRD1
Validated transcriptional targets of AP1 family members Fra1 and Fra2(N)	3.60E-03	3.70E+01	18	1.64E-09	3.61E-08	COL1A2, ITGB4, CCL2, PLAU, MMP1, MMP2, MMP9, LAMA3, GJA1, JUNB, NFATC2, FOSL2, FOSL1, IVL, THBD, LIF, PLAUR, DCN
Autoimmune thyroid disease(K)	5.10E-03	5.20E+01	21	1.94E-09	4.07E-08	HLA-DRB5, HLA-DRB4, HLA-DRB3, HLA-DRB1, CTLA4, HLA-DPA1, HLA-DMA, HLA-DMB, HLA-DPB1, HLA-DOA, HLA-DOB, HLA-DRA, FASLG, GZMB, PRF1, CD40, CD80, HLA-DQA2, HLA-DQA1, HLA-B, HLA-F
Chemokine signalling pathway(K)	1.83E-02	1.87E+02	42	2.15E-09	4.51E-08	LYN, JAK3, STAT1, CXCL13, CXCL14, SHC1, CXCL10, CXCL11, CXCL12, ARRB2, PIK3CG, CDC42, CCL5, CCL3, CCL2, PIK3R5, RAC2, CCR7, CCR4, CCR2, PAK1, ITK, PLCB3, CCL14, CCL13, ADCY4, CCL19, CCL18, CCL24, CCL22, CCL21, CCL27, CX3CR1, TIAM1, VAV1, HCK, WAS, CXCR6, CXCR3, CXCL9, CXCL2, FGR
Rap1 signalling pathway(K)	2.07E-02	2.11E+02	45	2.71E-09	5.41E-08	SKAP1, NGFR, EGFR, FLT4, ITGB2, ITGAL, PIK3CG, LCP2, PARD6G, FYB, CDC42, RASSF5, EPHA2, LPAR1, LPAR4, PIK3R5, IGF1, ID1, RAC2, GRIN1, PDGFRA, RASGRP3, FGF22, FGF11, MAPK13, MAPK11, PLCB3, MRAS, ADCY4, CSF1, FGF1, MAGI1, PGF, CSF1R, LAT, P2RY1, CALML5, CALML3, TIAM1, EFNA5, EFNA3, FGFR3, FGFR2, FGFR1, RGS14
Influenza A(K)	1.71E-02	1.75E+02	40	3.23E-09	6.45E-08	HLA-DRB5, HLA-DRB4, HLA-DRB3, HLA-DRB1, HLA-DPA1, HLA-DMA, HLA-DMB, STAT1, HLA-DPB1, HLA-DOA, HLA-DOB, TICAM1, SOCS3, TNFRSF10D, CXCL10, HLA-DRA, CIITA, FASLG, PIK3CG, FURIN, CCL5, CCL2, TNFSF10, RSAD2, PIK3R5, IL18, IL1A, IL1B, MAPK13, MAPK11, PRSS3, HLA-DQA2, HLA-DQA1, TNF, PYCARD, PML, TLR3, OAS1, OAS2, OAS3

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
PI3K-Akt signalling pathway(K)	3.38E-02	3.45E+02	60	1.19E-08	2.26E-07	NGFR, FN1, COL1A2, JAK3, EGFR, COL3A1, ITGB4, FLT4, FASLG, PIK3CG, ITGB8, ITGB6, ITGA1, ITGA8, ITGA5, IL2RG, IL2RA, IL2RB, EPHA2, LPAR1, LPAR4, PIK3R5, CD19, IGF1, OSMR, CCND2, PDGFRA, FGF22, COL5A1, COL5A3, IL7R, FGF11, LAMC2, LAMA2, LAMA3, LAMB3, LAMB4, CREB3L1, CSF3R, CSF1, TNC, FGF1, COL27A1, PGF, CSF1R, TNXB, COMP, COL4A2, COL4A4, EFNA5, EFNA3, FGFR3, FGFR2, FGFR1, COL6A2, COL6A1, COL6A3, COL6A6, THBS2, PPP2R2C
Herpes simplex infection(K)	1.80E-02	1.84E+02	40	1.25E-08	2.37E-07	C3, HLA-DRB5, HLA-DRB4, HLA-DRB3, HLA-DRB1, HLA-DPA1, HLA-DMA, HLA-DMB, STAT1, HLA-DPB1, HLA-DOA, HLA-DOB, TICAM1, SOCS3, HLA-DRA, FASLG, CCL5, CCL2, PVRL2, PVRL1, TNFSF14, IL1B, CD74, CFP, HLA-DQA2, HLA-DQA1, TNF, TAP2, TAP1, TRAF1, PML, HLA-B, HLA-F, TLR3, ARNTL, CASP8, OAS1, OAS2, OAS3, LTA
Axon guidance(K)	1.24E-02	1.27E+02	31	4.59E-08	8.27E-07	UNC5B, NGEF, CXCL12, NTN1, CDC42, EPHB6, EPHB2, EPHA4, EPHA1, EPHA2, NRP1, PPP3CC, RAC2, PAK1, PAK6, SLIT3, NFATC2, SEMA6B, SEMA3A, SEMA3F, ABLIM1, ABLIM2, SEMA4A, SEMA4D, RHOD, EFNA5, EFNB2, EFNB1, EFNA3, LIMK2, LIMK1
Proteoglycans in cancer(K)	1.99E-02	2.03E+02	41	5.91E-08	1.04E-06	HSPG2, FN1, EGFR, FLNB, FLNC, WNT7B, WNT7A, FASLG, PIK3CG, ITGA5, CDC42, PTPN6, ANK3, ANK1, PIK3R5, IGF1, PAK1, GPC1, PLAU, WNT3, WNT4, MMP2, MMP9, MAPK13, MAPK11, PLCG2, WNT10A, MRAS, HCLS1, TNF, ITPR2, CBLC, TIAM1, WNT5A, TGFB1, WNT3A, FGFR1, PLAUR, DCN, CAMK2D, LUM
Ras signalling pathway(K)	2.22E-02	2.27E+02	44	6.09E-08	1.04E-06	NGFR, EGFR, SHC1, FLT4, FASLG, PIK3CG, CDC42, RASSF5, EPHA2, PIK3R5, IGF1, RAC2, PAK1, PAK6, GRIN1, ETS2, PDGFRA, RASGRP1, RASGRP4, RASGRP3, FGF22, FGF11, PLA2G4F, PLA2G4D, PLA2G4E, PLA2G2A, PLCG2, MRAS, CSF1, FGF1, PGF, CSF1R, RASAL1, RASAL3, LAT, CALML5, CALML3, TIAM1, EFNA5, EFNA3, FGFR3, FGFR2, FGFR1, RASA2

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
Amoebiasis(K)	1.06E-02	1.08E+02	28	6.12E-08	1.04E-06	FN1, COL1A2, SERPINB13, COL3A1, ITGB2, PIK3CG, PIK3R5, CTSG, IL1B, COL5A1, COL5A3, LAMC2, LAMA2, LAMA3, LAMB3, LAMB4, SERPINB3, SERPINB4, SERPINB2, PLCB3, TNF, COL27A1, IL1R1, RAB7B, COL4A2, COL4A4, TGFB1, GNA15
Downstream signalling in naïve CD8+ T cells(N)	6.30E-03	6.40E+01	21	6.41E-08	1.09E-06	STAT4, TNFRSF4, TNFRSF9, FASLG, IL2RG, GZMB, IL2RA, IL2RB, PTPN7, CD247, PRF1, CD3G, CD3E, CD3D, CD8B, CD8A, JUNB, PRKCQ, NFATC2, FOSL1, TNF
MAPK signalling pathway(K)	2.50E-02	2.55E+02	47	9.27E-08	1.48E-06	CACNA2D1, CACNA2D3, DUSP1, EGFR, FLNB, FLNC, MAP3K6, MAPKAPK2, NFKB2, FASLG, ARRB2, CDC42, NTRK2, PTPN7, PPP3CC, RAC2, IL1A, IL1B, RELB, PAK1, PDGFRA, CACNA1C, RASGRP1, RASGRP4, RASGRP3, CACNA1I, FGF22, FGF11, PLA2G4F, PLA2G4D, PLA2G4E, MKNK1, MAPK13, MAPK11, MRAS, FGF1, TNF, CACNB1, MAP3K14, DUSP10, IL1R1, GADD45G, TGFB1, FGFR3, FGFR2, FGFR1, RASA2
GPVI-mediated activation cascade(R)	4.80E-03	4.90E+01	18	1.09E-07	1.75E-06	LYN, JAK3, SHC1, PIK3CG, LCP2, IL2RG, IL2RA, IL2RB, PTPN6, PIK3R5, RAC2, FCER1G, PLCG2, LAT, LCK, VAV1, PDPN, CSF2RB
CXCR4-mediated signalling events(N)	7.80E-03	8.00E+01	23	1.57E-07	2.51E-06	LYN, HLA-DRB1, STAT1, PTPRC, CXCL12, HLA-DRA, ARRB2, CDC42, PTPN6, CD247, CD3G, CD3E, CD3D, PAK1, RGS1, MMP9, PLCB3, LCK, INPP5D, VAV1, HCK, FGR, LIMK1
NOD-like receptor signalling pathway(K)	5.60E-03	5.70E+01	19	2.12E-07	3.18E-06	PYDC1, BIRC3, CCL5, CCL2, NAIP, IL18, IL1B, CARD18, MAPK13, MAPK11, MEFV, CARD9, TNF, PYCARD, NOD1, NOD2, NLRP1, CXCL2, CASP8
IL12 signalling mediated by STAT4(N)	3.00E-03	3.10E+01	14	2.52E-07	3.79E-06	HLA-DRB1, STAT4, HLA-DRA, IL2RA, CD247, PRF1, CD3G, CD3E, CD3D, IL18, IL18R1, CD80, IL18RAP, TGFB1
Interferon alpha/beta signalling(R)	6.50E-03	6.60E+01	20	4.48E-07	6.72E-06	STAT1, SOCS3, SOCS1, PTPN6, RSAD2, IFI27, ISG15, ISG20, IRF8, IRF5, IRF6, IFITM1, IFIT3, HLA-B, HLA-F, OASL, OAS1, OAS2, OAS3, GBP2
Integrin signalling pathway(P)	1.55E-02	1.58E+02	33	5.60E-07	7.84E-06	COL13A1, FN1, COL1A2, MICALL1, FLNB, COL3A1, COL17A1, SHC1, ITGB4, ITGB2, ITGAL, PIK3CG, ITGAX, ITGB8, ITGB6,

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
						ITGA1, ITGA8, ITGA5, COL15A1, CDC42, GRAP, RAC2, COL5A1, COL5A3, MAPK13, COL7A1, COL12A1, PIK3C2B, COL4A2, COL4A4, COL6A2, COL6A1, COL6A3
GPCR ligand binding(R)	3.81E-02	3.89E+02	60	6.61E-07	9.25E-06	FPR3, C3, PTH1R, PTAFR, P2RY13, P2RY10, GPR68, GPR65, CXCL13, CXCL10, CXCL11, CXCL12, CCL5, CCL3, CCL2, LPAR1, LPAR4, CCR7, CCR4, CCR2, GCGR, WNT3, LTB4R, PTHLH, GPR132, GRP, CCL19, CCL22, CCL21, GPR4, CCL27, CYSLTR1, LTB4R2, ADRA2A, CX3CR1, P2RY2, P2RY1, EDNRA, NTSR1, NMU, FFAR2, EDN1, PTGER4, PTGER2, PTGER3, CXCR6, CXCR3, F2RL1, F2RL2, WNT5A, ADRB2, HRH2, WNT3A, CXCL9, CXCL2, PTGIR, S1PR1, S1PR3, S1PR2, S1PR4

Table T3-5: Full list of pathways enriched for genes negatively correlated (at FDR<0.01) with tumour *VDR* expression, after adjusting for *FLG2*. Output from Reactome FIViz

The source database for each of the enriched pathways are indicated by a letter in parentheses after each pathway gene set name. The source database annotations are: C - CellMap, R – Reactome, K – KEGG, N – NCI PID, P - Panther, and B – BioCarta.

The description of column headers are: *Ratio of protein in gene set*: ratios of numbers of genes contained in pathways to total genes in the Reactome FI network; Number of protein in gene set: numbers of genes in pathways; *Protein from network*: numbers of hit genes from the query gene list; *Nodes:* nodal genes from my input query gene list, which 'matches' with the genes in a particular pathway; P-value was estimated by Reactome FIViz using hypergeometric test and the corresponding FDR was estimated by Benjamini Hochberg multiple correction

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
Mitochondrial translation(R)	0.0087	89	33	6.13E-08	3.46E-05	GADD45GIP1, MRPS17, MRPS14, MRPS11, MRPS28, MRPS26, MRPS27, MRPS24, MRPS23, MRPS35, MRPS33, MRPS31, GFM1, DAP3, MRPL18, MRPL16, MRPL14, MRPL15, MRPL13, MRPL10, MRPL11, MRPL24, MRPL22, MRPL42, MRPL47, MRPL50, MRPL3, MRPL2, MRPL1, MRPS7, MRPS18B, MRRF, PTCD3
Mitotic Prometaphase(R)	0.0097	99	35	7.63E-08	3.46E-05	NUP107, CDCA5, CENPF, CENPH, CENPI, CENPM, CENPN, CENPQ, NDC80, SMC2, PLK1, PPP1CC, NSL1, XPO1, ZWILCH, BUB1, KIF2A, KIF2C, MIS12, SPC25, ERCC6L, PDS5B, CCNB2, NUP43, NUP37, SKA1, RAD21, NDE1, CLIP1, RANBP2, KNTC1, KIF18A, ITGB3BP, CLASP1, CLASP2
Fanconi anemia pathway(K)	0.0052	53	23	4.44E-07	1.34E-04	STRA13, ATRIP, RMI1, BLM, REV1, WDR48, USP1, FANCL, FANCC, FANCE, FANCG, FANCF, MLH1, EME1, MUS81, PMS2, BRCA1, RPA1, RPA3, RAD51C, POLK, POLH, POLN
Mitotic Metaphase and Anaphase(R)	0.016	163	45	1.15E-06	2.59E-04	NUP107, CDCA5, CENPF, CENPH, CENPI, CENPM, CENPN, CENPQ, NDC80, CDC23, CDC16, PLK1, PPP1CC, NSL1, XPO1, ZWILCH, BUB1, PSMD9, PSMD5, PSME3, PSMF1, PSMA6, RPS27A, PSMC5, ANAPC5, KIF2A, KIF2C, MIS12, SPC25, ERCC6L, PDS5B, NUP43, NUP37, SKA1, RAD21, EMD, NDE1, CLIP1, RANBP2, PTTG1, KNTC1, KIF18A, ITGB3BP, CLASP1, CLASP2
RNA Polymerase I, RNA Polymerase III, and Mitochondrial Transcription(R)	0.0087	89	30	1.63E-06	2.95E-04	POLRMT, RRN3, MTA1, MTA3, ERCC2, KAT2B, NFIA, GTF3A, POLR3GL, POLR1B, POLR1C, POLR1D, POLR1E, POLR2F, POLR3C, POLR3E, POLR3F, POLR3G, POU2F1, CBX3, SNAPC3, GTF3C6, CHD3, TAF1D, CD3EAP, BRF2, GTF2H4, RBBP4, RBBP7, TFAM
The citric acid (TCA) cycle and respiratory electron transport(R)	0.0156	159	42	7.36E-06	1.11E-03	SCO1, ATP5A1, UQCRB, UQCRQ, ETFA, NDUFAF2, NDUFAF1, COX7B, COX5A, NDUFAB1, TACO1, SDHC, SDHD, NDUFB10, COX11, PDHA1, LDHC, IDH3B, PDK2, PDHX, SUCLG2, PDPR, NDUFA13, NDUFA11, ATP5G3, ATP5G1, UQCRC2, DLD, NDUFC2, NDUFB8, NDUFB4, NDUFB2, NDUFA8, NDUFA7,

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
						NDUFA4, ATP5E, NDUFV2, CYCS, UQCRFS1, NDUFS4, NDUFS2, LRPPRC
Cell cycle(K)	0.0121	124	35	1.09E-05	1.40E-03	E2F3, CDC23, CDC16, PCNA, ANAPC13, PRKDC, PLK1, CDKN1C, CDKN1B, CDKN2C, CDKN2A, MCM7, SMC1B, TFDP2, MCM3, MCM2, TTK, BUB1, WEE1, WEE2, ANAPC5, CCNB3, CCNB2, RAD21, CDC7, CUL1, ABL1, CDK6, CDK4, CDK2, CHEK1, SKP2, SKP1, PTTG1, SMAD2
Wnt signalling pathway(P)	0.0262	268	60	1.49E-05	1.58E-03	PCDH7, SMARCAL1, CER1, SMARCA1, SMARCA4, PCDHB14, PCDHB13, PCDHB12, PCDHB11, PCDHB10, PCDHB16, GNG3, GNG2, GNG4, GNG7, PRKCI, PRKCE, PRKCA, CTBP2, PCDHA5, PCDHA4, PCDHA3, PCDHA2, PCDHA9, PCDHA8, PCDHA6, PCDHB2, PCDHB6, PCDHB5, PCDHB3, PCDHB7, TLE1, PYGO1, BCL9, NKD1, PCDHA12, PCDHA11, LRP5L, PPP2R5A, MYH3, CDH10, PCDHGA11, DCHS1, LRP6, FZD9, EN2, CDH4, SMARCD1, SMARCE1, APC, NLK, SIAH1, TBL1Y, CTNNA2, TBL1XR1, PCDHGB7, PCDHGB3, PCDHGA8, PCDHGA3, TGFBR1
Ribosome(K)	0.0134	137	37	1.58E-05	1.58E-03	MRPS17, MRPS14, MRPS11, RPS4Y1, RPS15A, RSL24D1, RPS13, RPS23, RPS3A, MRPL18, MRPL16, MRPL14, MRPL15, MRPL13, MRPL10, MRPL11, MRPL24, MRPL22, RPS27A, MRPL3, MRPL2, MRPL1, MRPS7, RPL34, RPL7, RPL22, RPL26, RPL29, RPL28, RPL10, RPL13, RPL15, RPS7, RPS5, RPL7A, RPL36A, RPL35A
Oxidative phosphorylation(K)	0.013	133	36	1.96E-05	1.65E-03	ATP5A1, UQCRB, UQCRQ, ATP6V1E1, ATP6V1E2, ATP6V1G1, ATP6V1C1, COX7B, COX5A, NDUFAB1, SDHC, SDHD, NDUFB10, COX15, COX11, ATP6V0D2, ATP6V0A1, NDUFA13, NDUFA11, ATP5G3, ATP5G1, UQCRC2, NDUFC2, NDUFB8, NDUFB4, NDUFB2, NDUFA8, NDUFA7, NDUFA4, ATP5E, NDUFV2, UQCRFS1, NDUFS4, NDUFS2, ATP6V1A, ATP6V1F
Assembly of the primary cilium(R)	0.0168	172	43	2.09E-05	1.65E-03	NINL, IFT140, NEK2, BBS1, PCM1, SSNA1, AHI1, TMEM216, ASAP1, CDK5RAP2, CEP70, CEP72, PLK1, WDR60, CEP57, TCTN3, TCTN1, EXOC8, EXOC2, WDR19, CEP135, CEP164,

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
						CEP152, DYNC2H1, TUBB, RAB11A, LZTFL1, ARL6, FGFR1OP, CCT3, KIF24, CCT5, RPGRIP1L, KIF3A, NPHP1, DCTN2, NDE1, INPP5E, CC2D2A, BBS12, TRIP11, PDE6D, CLASP1
Fanconi anemia pathway(N)	0.0044	45	18	2.21E-05	1.65E-03	ATRIP, RMI1, BLM, RAD1, WDR48, USP1, FANCL, FANCC, FANCE, FANCG, FANCF, RFC3, TOPBP1, XRCC3, BRCA1, RPA1, NBN, CHEK1
BARD1 signalling events(N)	0.0028	29	14	2.44E-05	1.69E-03	PCNA, PRKDC, BARD1, FANCL, FANCC, FANCE, FANCG, FANCF, CSTF1, TOPBP1, XRCC5, BRCA1, NBN, CDK2
S Phase(R)	0.0117	120	33	3.15E-05	2.02E-03	CDCA5, APEX1, POLA1, PCNA, POLE2, CDKN1B, MCM7, RFC3, RFC1, MCM3, MCM2, ESCO1, ESCO2, DNA2, GINS2, WEE1, PSMD9, PSMD5, PSME3, PSMF1, PSMA6, RPS27A, PSMC5, RPA1, RPA3, PDS5B, RAD21, CUL1, CDK4, CDK2, SKP2, SKP1, CKS1B
Cadherin signalling pathway(P)	0.0098	100	29	3.73E-05	2.09E-03	PCDH7, PCDHB14, PCDHB13, PCDHB12, PCDHB11, PCDHB10, PCDHB16, PCDHA5, PCDHA4, PCDHA3, PCDHA2, PCDHA9, PCDHA8, PCDHA6, PCDHB2, PCDHB6, PCDHB5, PCDHB3, PCDHB7, PCDHA12, PCDHA11, PCDHGA11, DCHS1, FZD9, CTNNA2, PCDHGB7, PCDHGB3, PCDHGA8, PCDHGA3
SUMOylation(R)	0.0098	100	29	3.73E-05	2.09E-03	NUP107, RING1, BLM, NOP58, PCNA, RAE1, UBE2I, NUP155, UBA2, MTA1, SMC6, POM121C, CDKN2A, PCGF2, PARP1, CBX4, SAE1, SCMH1, NSMCE2, BRCA1, RPA1, AURKA, NUP43, NUP35, NUP37, RAD21, MITF, RANBP2, SENP2
Mitotic G2-G2/M phases(R)	0.0109	111	31	4.09E-05	2.17E-03	LIN54, NINL, CENPF, NEK2, E2F3, PCM1, SSNA1, CDK5RAP2, CEP70, CEP72, PLK1, CEP57, CEP135, CEP164, CEP152, XPO1, TUBB, WEE1, RPS27A, TUBGCP5, FGFR1OP, PPP1R12A, CCNB2, DCTN2, AURKA, NDE1, CUL1, CDK2, RBBP4, SKP1, CLASP1
RNA transport(K)	0.0168	172	42	4.36E-05	2.18E-03	NUP107, POP1, POP4, EIF1AY, EIF5B, EIF1B, PABPC5, RAE1, RPP30, RPP21, UBE2I, NUP155, TGS1, POM121C, PAIP1, EIF2B5, EIF2B4, MAGOH, SRRM1, XPO1, RGPD5, EIF2S2, EIF2S1, EIF2S3, NXF3, CLNS1A, NXT1, SAP18, EIF4A2, EIF4A3,

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
						GEMIN4, GEMIN6, NCBP2, NUP43, NUP35, NUP37, UPF3B, FMR1, NMD3, PRMT5, RANBP2, SENP2
Resolution of Abasic Sites (AP sites)(R)	0.003	31	14	4.96E-05	2.33E-03	APEX1, PCNA, SMUG1, NTHL1, RFC3, RFC1, PARP1, LIG3, UNG, XRCC1, RPA1, RPA3, POLB, MUTYH
Processing of Capped Intron-Containing Pre-mRNA(R)	0.0165	169	41	6.14E-05	2.77E-03	NUP107, NUDT21, ZCRB1, FUS, BCAS2, SNRPD1, RAE1, HNRNPA2B1, NUP155, PABPN1, HNRNPU, RNPC3, POM121C, HNRNPF, PAPOLA, SNRPG, SNRPE, SNRNP25, POLR2C, POLR2F, MAGOH, SRRM1, CSTF1, PRPF3, CPSF2, CWC15, CD2BP2, PCBP1, NCBP2, NUP43, NUP35, NUP37, UPF3B, GTF2F2, LSM5, LSM3, SNRPA1, SNRPB2, PRPF19, PCF11, RANBP2
Huntington's disease(K)	0.0189	193	45	6.83E-05	2.94E-03	BBC3, ATP5A1, UQCRB, NRF1, UQCRQ, SP1, COX7B, COX5A, NDUFAB1, SDHC, SDHD, NDUFB10, DNAL4, GRM5, POLR2C, POLR2F, NDUFA13, NDUFA11, ATP5G3, ATP5G1, TBPL1, UQCRC2, PPARGC1A, BDNF, NDUFC2, SLC25A5, SLC25A4, NDUFB8, NDUFB4, NDUFB2, NDUFA8, NDUFA7, NDUFA4, ATP5E, NDUFV2, CYCS, UQCRFS1, CREB3, NDUFS4, NDUFS2, DCTN2, DCTN4, TAF4, TFAM, VDAC2
Mitotic G1-G1/S phases(R)	0.0123	126	33	7.85E-05	3.22E-03	LIN54, E2F3, POLA1, PCNA, POLE2, CDKN1B, CDKN2C, CDKN2A, MCM7, TFDP2, MCM3, MCM2, TYMS, WEE1, PSMD9, PSMD5, PSME3, PSMF1, PSMA6, RPS27A, PSMC5, MCM10, RPA1, RPA3, CDC7, CUL1, CDK6, CDK4, CDK2, RBBP4, SKP2, SKP1, CKS1B
HDR through Homologous Recombination (HR) or Single Strand Annealing (SSA)(R)	0.0083	85	25	1.00E-04	3.90E-03	ATRIP, RMI1, BLM, UIMC1, RAD1, UBE2I, UBE2N, BARD1, RFC3, DNA2, EME1, MUS81, TOPBP1, SIRT6, RPS27A, XRCC2, XRCC3, BRCA1, RPA1, RPA3, NBN, ABL1, CDK2, CHEK1, RAD51C

Table T3-6: Full list of pathways enriched for genes positively correlated (at FDR<0.00001) with tumour *VDR* expression, after adjusting for *FLG2*. Output from Reactome FIViz

The source database for each of the enriched pathways are indicated by a letter in parentheses after each pathway gene set name. The source database annotations are: C - CellMap, R – Reactome, K – KEGG, N – NCI PID, P - Panther, and B – BioCarta.

The description of column headers are: *Ratio of protein in gene set*: ratios of numbers of genes contained in pathways to total genes in the Reactome FI network; Number of protein in gene set: numbers of genes in pathways; *Protein from network*: numbers of hit genes from the query gene list; *Nodes:* nodal genes from my input query gene list, which 'matches' with the genes in a particular pathway; P-value was estimated by Reactome FIViz using hypergeometric test and the corresponding FDR was estimated by Benjamini Hochberg multiple correction

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
Extracellular matrix organization(R)	0.0243	248	103	1.11E-16	5.51E-14	COL13A1, PDGFB, ADAMTS14, FN1, TPSAB1, BGN, MMP10, MMP11, MMP13, MMP16, MMP15, MMP19, COL3A1, EFEMP1, COL17A1, ITGB4, ITGB2, ITGAL, ITGAX, ITGB8, ITGB6, ITGA2, ITGA1, ITGA8, ITGA5, ICAM2, ICAM3, ICAM4, KDR, ACTN1, SDC3, LOXL1, VCAM1, FBN1, COL7A1, COL10A1, TNC, COL27A1, ADAMTS5, ADAMTS2, PECAM1, MFAP4, MFAP2, COL18A1, COMP, LOX, FBLN1, FBLN2, FBLN5, DCN, COL6A2, COL6A1, COL6A3, THBS1, LUM, HSPG2, COL1A1, COL1A2, COL11A1, COL15A1, FURIN, COL9A3, CTSS, CTSG, CTSD, CTSB, PCOLCE, PSEN1, BMP2, BMP1, COL5A1, COL5A3, COL5A2, LAMC2, MMP7, MMP1, MMP2, MMP3, MMP8, MMP9, VCAN, LAMA2, LAMA3, LAMB3, SERPINE1, COL12A1, SERPINH1, ELANE, ELN, FMOD, ADAM9, ADAM8, SPARC, COL8A2, COL8A1, CEACAM6, COL4A2, COL4A1, COL4A4, COL4A3, LTBP4, NID2, TGFB1
Cytokine-cytokine receptor interaction(K)	0.0259	265	109	1.11E-16	5.51E-14	PDGFB, PDGFC, EGFR, IL12RB1, TNFRSF4, TNFRSF17, TNFRSF9, TNFSF13B, TNFRSF1B, IL2RG, IL2RA, IL2RB, TNFRSF6B, TNFSF10, KDR, TNFSF14, TNFSF13, IL24, IL11, IL15, IL18, IL1A, IL1B, VEGFC, VEGFA, KITLG, IL18R1, IL7R, CNTF, IL10RB, IL10RA, CLCF1, CSF2, CSF1, TNF, IFNAR2, BMPR2, IL21R, IL15RA, IL1R1, FLT3LG, LIF, CCL4L1, BMPR1B, LTA, LTB, IL20RB, NGFR, CXCL13, CXCL14, CXCL16, TNFRSF11B, TNFRSF10B, TNFRSF10D, CXCL10, CXCL11, CXCL12, FLT3, FLT4, FASLG, CCL8, CCL7, CCL5, CCL3, CCL2, ACVR1, CD27, CD40, TNFRSF13B, CCR7, CCR4, CCR2, TNFRSF12A, OSMR, CD70, PDGFRB, PDGFRA, CCL3L1, INHBB, INHBA, IL6, BMP2, TNFSF4, CCL14, CCL13, CCL19, CCL18, CCL17, CCL24, CCL22, CCL21, CSF3R, CSF1R, IL3RA, CX3CR1, CXCR5, CXCR4, CXCR6, CXCR3, FAS, IL18RAP, CSF2RB, CSF2RA, TGFB1, CXCL9, CXCL1, CXCL2, TGFBR2, IFNE

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
Hematopoietic cell lineage(K)	0.0085	87	50	2.22E-16	7.33E-14	HLA-DRB4, HLA-DRB3, HLA-DRB1, ITGA2, ITGA1, ITGA5, IL2RA, ANPEP, IL11, IL1A, IL1B, KITLG, IL7R, MS4A1, CD2, CD4, CD5, CD7, CD9, CSF2, CSF1, TNF, CR1, IL1R1, FLT3LG, HLA-DRA, FLT3, FCGR1A, CD1D, CD1C, CD1B, CD1A, CD19, CD22, CD3G, CD3E, CD3D, CD38, CD37, CD36, CD33, MME, CD8B, CD8A, IL6, FCER2, CSF3R, CSF1R, IL3RA, CSF2RA
Staphylococcus aureus infection(K)	0.0054	55	38	3.55E-15	8.81E-13	HLA-DRB4, HLA-DRB3, HLA-DRB1, PTAFR, ITGB2, ITGAL, C1S, C1R, SELPLG, CFB, CFD, CFH, CFI, HLA-DQA2, HLA-DQA1, HLA-DQB1, C1QB, C1QA, C1QC, SELP, FPR1, FPR3, C2, C3, HLA-DPA1, HLA-DMA, HLA-DMB, HLA-DPB1, HLA-DOA, HLA-DOB, HLA-DRA, FCGR3A, DSG1, FCGR1A, FCGR2B, FCGR2C, MASP1, C5AR1
Pathways in cancer(K)	0.0389	397	120	9.33E-15	1.85E-12	SPI1, BCR, PDGFB, FN1, JAK1, PPARD, EGFR, EGLN1, EGLN3, WNT7B, WNT7A, ITGA2, ROCK1, CDC42, E2F2, EPAS1, LPAR1, TGFA, LPAR4, STAT5B, IGF1, RAC2, ARHGEF1, RELA, GNAI2, VEGFC, VEGFA, KITLG, TCF7, FGF11, CTBP1, ADCY4, ADCY7, RUNX1T1, CDKN2B, FGF1, PGF, PLEKHG5, CEBPA, HDAC2, PML, CBLC, FLT3LG, PTGER4, PTGER2, PTGER3, F2RL3, WNT5B, WNT5A, WNT3A, FGFR3, FGFR2, FGFR1, WNT2B, CASP8, MAPK8, GNA12, AR, STAT1, MAX, NFKB2, CXCL12, FLT3, PIK3CD, FASLG, PIK3CG, BIRC3, RASSF5, PIK3R5, BRCA2, PDGFRB, PDGFRA, RASGRP1, RASGRP4, RASGRP3, WNT4, IL6, BMP2, LAMC2, MMP1, MMP2, MMP9, LAMA2, LAMA3, LAMB3, DAPK2, HIF1A, PLCG2, FZD1, WNT10B, WNT10A, FZD7, RET, GLI2, PLCB3, PAX8, CSF3R, CSF1R, TRAF2, TRAF1, TRAF3, SOS1, SOS2, EDNRA, RUNX1, GNG10, GNG11, CXCR4, FAS, COL4A2, COL4A1, COL4A4, COL4A3, CSF2RA, TGFB1, PTCH2, NFKBIA, BDKRB2, TGFBR2, BAD
Osteoclast differentiation(K)	0.0128	131	59	2.00E-14	3.01E-12	SPI1, JAK1, LCP2, BTK, PPP3CC, SYK, IL1A, IL1B, RELA, RELB, TYROBP, NFATC2, FOSL2, FOSL1, FOSB, CSF1, TNF, IFNAR2, MAP3K14, LCK, IL1R1, BLNK, MAPK8, STAT1, NCF1, NCF2, NCF4, TNFRSF11B, SOCS3, SOCS1, NFKB2, FHL2, PIK3CD,

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
						PIK3CG, FCGR3A, FCGR1A, FCGR2B, FCGR2C, PIK3R5, CYBA, OSCAR, SIRPB1, MAPK13, MAPK11, JUNB, PLCG2, CSF1R, TRAF2, LILRA1, LILRA5, LILRB1, LILRB2, LILRB4, LILRB5, CAMK4, SIRPG, TGFB1, NFKBIA, TGFBR2
Graft-versus-host disease(K)	0.004	41	32	2.13E-14	3.01E-12	HLA-DRB4, HLA-DRB3, HLA-DRB1, KIR3DL2, KIR2DL5A, GZMB, IL1A, IL1B, HLA-DQA2, HLA-DQA1, HLA-DQB1, TNF, HLA-B, HLA-F, HLA-E, HLA-DPA1, HLA-DMA, HLA-DMB, HLA-DPB1, HLA-DOA, HLA-DOB, HLA-DRA, FASLG, PRF1, CD28, CD86, CD80, KIR2DL1, IL6, FAS, KLRC1, KLRD1
TNF signalling pathway(K)	0.0108	110	53	3.28E-14	4.06E-12	JAG1, CASP10, MAP3K8, MAP3K5, TNFRSF1B, IL15, IL1B, RELA, VEGFC, IL18R1, VCAM1, TNFAIP3, CSF2, CSF1, TNF, MAP2K3, BCL3, MAP3K14, NOD2, LIF, EDN1, MLKL, SELE, CASP7, CASP8, LTA, MAPK8, RIPK3, PGAM5, SOCS3, CXCL10, PIK3CD, PIK3CG, BIRC3, CCL5, CCL2, PIK3R5, IL6, MMP3, MMP9, MAPK13, MAPK11, CREB3L1, JUNB, TRADD, TRAF2, TRAF1, TRAF3, CFLAR, FAS, NFKBIA, CXCL1, CXCL2
Tuberculosis(K)	0.0173	177	68	3.21E-13	3.53E-11	HLA-DRB4, HLA-DRB3, HLA-DRB1, CASP10, JAK2, JAK1, PLA2R1, CORO1A, CIITA, ITGB2, ITGAX, CD209, PPP3CC, CAMP, FCER1G, SYK, IL18, IL1A, IL1B, RELA, IRAK1, IRAK2, BCL10, HLA-DQA2, HLA-DQA1, HLA-DQB1, IL10RB, IL10RA, TNF, TCIRG1, CR1, NOD2, CALML5, CALML3, TLR9, TLR4, TLR2, CASP8, MAPK8, C3, HLA-DPA1, HLA-DMA, HLA-DMB, STAT1, HLA-DPB1, HLA-DOA, HLA-DOB, HLA-DRA, MYD88, FCGR3A, KSR1, FCGR1A, FCGR2B, FCGR2C, CTSS, CTSD, CD74, IL6, MAPK13, MAPK11, MRC2, CARD9, CLEC7A, TRADD, LSP1, TGFB1, CAMK2D, BAD
Rheumatoid arthritis(K)	0.0087	89	45	5.70E-13	5.65E-11	HLA-DRB4, HLA-DRB3, HLA-DRB1, ITGB2, ITGAL, TNFSF13B, ATP6V1C2, TNFSF13, IL11, IL15, IL18, IL1A, IL1B, VEGFA, HLA-DQA2, HLA-DQA1, HLA-DQB1, CSF2, CSF1, TNF, TCIRG1, TLR4, TLR2, LTB, CTLA4, HLA-DPA1, HLA-DMA, HLA-DMB, HLA-DPB1, HLA-DOA, HLA-DOB, CXCL12, HLA-DRA, CCL5,

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
						CCL3, CCL2, CD28, CD86, CD80, CCL3L1, IL6, MMP1, MMP3, TGFB1, CXCL1
Cell adhesion molecules (CAMs)(K)	0.0139	142	58	1.67E-12	1.50E-10	HLA-DRB4, HLA-DRB3, HLA-DRB1, SIGLEC1, CD274, ITGB2, ITGAL, SPN, ITGB8, ITGA8, ICAM2, ICAM3, CD226, SDC3, SELPLG, VCAM1, CD2, CD4, CD6, ALCAM, ICOS, HLA-DQA2, HLA-DQA1, HLA-DQB1, PECAM1, ICOSLG, HLA-B, HLA-F, HLA-E, PDCD1LG2, SELE, SELP, SELL, CLDN5, CTLA4, HLA-DPA1, HLA-DMA, HLA-DMB, HLA-DPB1, HLA-DOA, HLA-DOB, PTPRC, HLA-DRA, PVRL2, PVRL1, CD28, CD22, CD40, CD58, CD86, CD80, CD8B, CD8A, VCAN, CDH5, ESAM, TIGIT, NFASC
Leishmaniasis(K)	0.007	72	39	2.64E-12	2.16E-10	HLA-DRB4, HLA-DRB3, HLA-DRB1, JAK2, JAK1, ITGB2, IL1A, IL1B, RELA, IRAK1, HLA-DQA2, HLA-DQA1, HLA-DQB1, TNF, CR1, TLR4, TLR2, C3, HLA-DPA1, HLA-DMA, HLA-DMB, STAT1, HLA-DPB1, NCF1, NCF2, NCF4, HLA-DOA, HLA-DOB, HLA-DRA, MYD88, FCGR3A, FCGR1A, PTPN6, FCGR2C, CYBA, MAPK13, MAPK11, TGFB1, NFKBIA
Chemokine signalling pathway(K)	0.0183	187	68	3.59E-12	2.73E-10	JAK2, JAK3, SHC1, ROCK1, CDC42, STAT5B, RAC2, RELA, GNAI2, GRK5, GRK6, PTK2B, ADCY4, ADCY7, TIAM1, VAV1, HCK, CCL4L1, LYN, STAT1, NCF1, CXCL13, CXCL14, CXCL16, CXCL10, CXCL11, CXCL12, PIK3CD, ARRB2, PIK3CG, CCL8, CCL7, CCL5, CCL3, CCL2, PIK3R5, CCR7, CCR4, CCR2, PAK1, CCL3L1, ITK, PLCB3, CCL14, CCL13, CCL19, CCL18, CCL17, CCL24, CCL22, CCL21, SOS1, SOS2, CX3CR1, GNG10, GNG11, WAS, CXCR5, CXCR4, CXCR6, CXCR3, ELMO1, NFKBIA, CXCL9, CXCL1, CXCL2, FGR, DOCK2
NF-kappa B signalling pathway(K)	0.0089	91	44	4.34E-12	3.04E-10	TNFSF13B, BTK, TNFSF14, SYK, IL1B, RELA, RELB, PLAU, CARD11, IRAK1, VCAM1, TNFAIP3, PRKCQ, BCL10, TNF, MAP3K14, LAT, LCK, IL1R1, TLR4, CCL4L1, BLNK, ZAP70, LTA, LTB, LYN, TICAM1, NFKB2, CXCL12, MYD88, BIRC3, CD40, PLCG2, CCL13, CCL19, CCL21, TRADD, TRAF2, TRAF1, TRAF3, CFLAR, GADD45B, NFKBIA, CXCL2

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
TCR signalling in naïve CD4+ T cells(N)	0.0066	67	37	5.52E-12	3.57E-10	HLA-DRB1, MAP3K8, SHC1, LCP2, FYB, CDC42, CD247, CARD11, PDPK1, CD4, PRKCQ, BCL10, MAP3K14, LAT, LCK, VAV1, TRPV6, ZAP70, PTPRC, HLA-DRA, SH3BP2, PTPN6, RASSF5, CD28, CD3G, CD3E, CD3D, CD86, CD80, RASGRP1, ITK, GRAP2, STIM1, SOS1, INPP5D, MAP4K1, WAS
Interferon gamma signalling(R)	0.0072	74	39	5.89E-12	3.57E-10	HLA-DRB4, HLA-DRB3, HLA-DRB1, JAK2, JAK1, PTAFR, CIITA, MT2A, TRIM22, VCAM1, HLA-DQA2, HLA-DQA1, HLA-DQB2, HLA-DQB1, PML, HLA-B, HLA-F, HLA-E, OASL, OAS1, OAS2, OAS3, GBP2, GBP1, HLA-DPA1, STAT1, HLA-DPB1, SOCS3, SOCS1, HLA-DRA, FCGR1A, PTPN6, IFI30, IRF2, IRF7, IRF8, IRF5, IRF6, CAMK2D
Focal adhesion(K)	0.0203	207	72	6.16E-12	3.57E-10	PDGFB, PDGFC, FN1, EGFR, COL3A1, SHC1, ITGB4, ITGB8, ITGB6, ITGA2, ITGA1, ITGA8, ITGA5, ROCK1, CDC42, KDR, ACTN1, IGF1, RAC2, VEGFC, VEGFA, PDPK1, TNC, COL27A1, PGF, PARVG, VAV1, COMP, DIAPH1, COL6A2, COL6A1, COL6A3, COL6A6, THBS2, THBS1, MAPK8, MYLK, MYL9, COL1A1, COL1A2, COL11A1, FLNB, FLNC, FLT4, PIK3CD, PIK3CG, PPP1R12C, BIRC3, PIK3R5, PAK1, PAK6, PPP1CA, CCND2, PDGFRB, PDGFRA, COL5A1, COL5A3, COL5A2, LAMC2, LAMA2, LAMA3, LAMB3, ZYX, VWF, SOS1, SOS2, TNXB, COL4A2, COL4A1, COL4A4, COL4A3, BAD
HTLV-I infection(K)	0.0253	258	83	7.33E-12	4.03E-10	SPI1, PDGFB, HLA-DRB4, HLA-DRB3, HLA-DRB1, JAK3, JAK1, WNT7B, WNT7A, ITGB2, ITGAL, IL2RG, E2F2, IL2RA, IL2RB, CDC20, STAT5B, XBP1, NRP1, PPP3CC, IL15, RELA, RELB, ETS2, VCAM1, POLD3, POLD4, HLA-DQA2, HLA-DQA1, NFATC2, NFATC4, FOSL1, HLA-DQB1, ADCY4, ADCY7, CDKN2B, CSF2, TNF, MAP3K14, LCK, IL15RA, IL1R1, HLA-B, HLA-F, HLA-E, WNT5B, WNT5A, WNT3A, WNT2B, LTA, VAC14, MAPK8, HLA-DPA1, HLA-DMA, HLA-DMB, HLA-DPB1, HLA-DOA, HLA-DOB, ANAPC2, NFKB2, HLA-DRA, PIK3CD, PIK3CG, PIK3R5, CD40, CD3G, CD3E, CD3D, CCND2, PDGFRB,

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
						PDGFRA, WNT4, IL6, FZD1, WNT10B, WNT10A, FZD7, MRAS, ZFP36, PTTG2, TGFB1, NFKBIA, TGFBR2
Intestinal immune network for IgA production(K)	0.0046	47	30	1.85E-11	9.60E-10	HLA-DRB4, HLA-DRB3, HLA-DRB1, TNFRSF17, TNFSF13B, TNFSF13, IL15, ICOS, HLA-DQA2, HLA-DQA1, HLA-DQB1, MAP3K14, ICOSLG, IL15RA, HLA-DPA1, HLA-DMA, HLA-DMB, HLA-DPB1, HLA-DOA, HLA-DOB, CXCL12, HLA-DRA, CD28, CD40, TNFRSF13B, CD86, CD80, IL6, CXCR4, TGFB1
Integrin signalling pathway(P)	0.0155	158	59	3.54E-11	1.73E-09	COL13A1, FN1, COL3A1, LIMS1, MAP3K5, COL17A1, SHC1, ARPC3, ITGB4, ITGB2, ITGAL, ITGAX, ITGB8, ITGB6, ITGA2, ITGA1, ARFGAP1, ITGA8, ITGA5, CDC42, ACTN1, GRAP, RAC2, MAP3K2, PTK2B, COL7A1, ARPC1B, COL10A1, MAP2K3, PIK3C2B, COL6A2, COL6A1, COL6A3, MAPK8, COL1A1, COL1A2, COL11A1, MICALL1, FLNB, PIK3CD, PIK3CG, COL15A1, COL9A3, COL5A1, COL5A3, COL5A2, MAPK13, COL12A1, SOS1, SOS2, COL8A2, COL8A1, COL4A2, COL4A1, COL4A4, COL4A3, ELMO1, ELMO2, ARHGAP26
Natural killer cell mediated cytotoxicity(K)	0.0131	134	53	4.75E-11	2.23E-09	HCST, KIR3DL2, SHC1, KIR2DL5A, ITGB2, ITGAL, LCP2, GZMB, CD247, CD244, ICAM2, TNFSF10, PPP3CC, RAC2, FCER1G, SYK, TYROBP, PTK2B, NFATC2, CSF2, TNF, IFNAR2, LAT, LCK, HLA-B, HLA-E, SH2D1A, VAV1, ZAP70, TNFRSF10B, TNFRSF10D, PIK3CD, FASLG, PIK3CG, FCGR3A, SH3BP2, PTPN6, PRF1, PIK3R5, NCR1, CD48, PAK1, MICA, MICB, KIR2DL1, KIR2DL4, PLCG2, SOS1, SOS2, FAS, KLRC1, KLRC3, KLRD1
Costimulation by the CD28 family(R)	0.0062	63	34	7.13E-11	3.21E-09	HLA-DRB4, HLA-DRB3, HLA-DRB1, MAP3K8, CD274, CDC42, CD247, PDPK1, CD4, ICOS, HLA-DQA2, HLA-DQA1, HLA-DQB2, HLA-DQB1, BTLA, MAP3K14, LCK, ICOSLG, PDCD1LG2, VAV1, LYN, CTLA4, HLA-DPA1, HLA-DPB1, HLA-DRA, PTPN6, CD28, CD3G, CD3E, CD3D, PAK1, CD86, CD80, GRAP2
IL12-mediated signalling events(N)	0.006	61	33	1.27E-10	5.46E-09	HLA-DRB1, JAK2, IL12RB1, IL2RG, GZMA, GZMB, IL2RA, IL2RB, CD247, IL18, IL1B, RELA, RELB, IL18R1, CD4, MAP2K3, LCK,

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
						IL1R1, HLX, STAT4, STAT1, SOCS1, NFKB2, HLA-DRA, FASLG, CCL3, CD3G, CD3E, CD3D, CD8B, CD8A, GADD45B, IL18RAP
Inflammatory bowel disease (IBD)(K)	0.0064	65	34	1.58E-10	6.50E-09	GATA3, HLA-DRB4, HLA-DRB3, HLA-DRB1, IL12RB1, IL2RG, IL18, IL1A, IL1B, RELA, IL18R1, HLA-DQA2, HLA-DQA1, HLA-DQB1, TNF, NOD2, IL21R, TLR4, TLR2, HLA-DPA1, HLA-DMA, HLA-DMB, STAT4, STAT1, HLA-DPB1, RORC, RORA, HLA-DOA, HLA-DOB, MAF, HLA-DRA, IL6, IL18RAP, TGFB1
Beta1 integrin cell surface interactions(N)	0.0065	66	34	2.33E-10	9.10E-09	FN1, COL3A1, ITGA2, ITGA1, ITGA8, ITGA5, VEGFA, PLAU, TGM2, VCAM1, FBN1, COL7A1, TNC, COL18A1, PLAUR, COL6A2, COL6A1, COL6A3, THBS2, THBS1, COL1A1, COL1A2, COL11A1, F13A1, COL5A1, COL5A2, LAMC2, LAMA2, LAMA3, LAMB3, COL4A1, COL4A4, COL4A3, TGFBI
Antigen processing and presentation(K)	0.0075	77	37	2.53E-10	9.43E-09	HLA-DRB4, HLA-DRB3, HLA-DRB1, LGMN, KIR3DL2, KIR2DL5A, CIITA, TAPBP, CD4, HLA-DQA2, HLA-DQA1, HLA-DQB1, TNF, HLA-B, HLA-F, HLA-E, PSME2, HLA-DPA1, HLA-DMA, HLA-DMB, HLA-DPB1, HLA-DOA, HLA-DOB, HLA-DRA, IFI30, CTSS, CTSB, CD74, CD8B, CD8A, KIR2DL1, KIR2DL4, TAP2, TAP1, KLRC1, KLRC3, KLRD1
Type I diabetes mellitus(K)	0.0042	43	27	2.62E-10	9.43E-09	HLA-DRB4, HLA-DRB3, HLA-DRB1, GZMB, IL1A, IL1B, HLA-DQA2, HLA-DQA1, HLA-DQB1, TNF, HLA-B, HLA-F, HLA-E, LTA, HLA-DPA1, HLA-DMA, HLA-DMB, HLA-DPB1, HLA-DOA, HLA-DOB, HLA-DRA, FASLG, PRF1, CD28, CD86, CD80, FAS
Allograft rejection(K)	0.0036	37	25	2.74E-10	9.60E-09	HLA-DRB4, HLA-DRB3, HLA-DRB1, GZMB, HLA-DQA2, HLA-DQA1, HLA-DQB1, TNF, HLA-B, HLA-F, HLA-E, HLA-DPA1, HLA-DMA, HLA-DMB, HLA-DPB1, HLA-DOA, HLA-DOB, HLA-DRA, FASLG, PRF1, CD28, CD40, CD86, CD80, FAS
PI3K-Akt signalling pathway(K)	0.0338	345	96	3.77E-10	1.28E-08	PDGFB, PDGFC, FN1, JAK2, JAK3, JAK1, EGFR, COL3A1, ITGB4, ITGB8, ITGB6, ITGA2, ITGA1, ITGA8, ITGA5, IL2RG, IL2RA, IL2RB, KDR, LPAR1, LPAR4, PCK2, IGF1, SYK, RELA, VEGFC, VEGFA, KITLG, IL7R, FGF11, PDPK1, CSF1, TNC, FGF1, IFNAR2, COL27A1, PGF, MCL1, TLR4, TLR2, COMP,

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
						EFNA5, EFNA3, FGFR3, FGFR2, FGFR1, COL6A2, COL6A1, COL6A3, COL6A6, THBS2, THBS1, NOS3, PPP2R3B, PPP2R2C, NGFR, COL1A1, COL1A2, COL11A1, BCL2L11, FLT4, PIK3CD, FASLG, PIK3CG, EPHA2, PIK3R5, CD19, OSMR, CCND2, PDGFRB, PDGFRA, IL6, COL5A1, COL5A3, COL5A2, LAMC2, LAMA2, LAMA3, LAMB3, CREB3L1, PHLPP2, CSF3R, CSF1R, VWF, IL3RA, SOS1, SOS2, GNG10, GNG11, RPTOR, TNXB, COL4A2, COL4A1, COL4A4, COL4A3, BAD
Toxoplasmosis(K)	0.0116	118	47	4.70E-10	1.51E-08	HLA-DRB4, HLA-DRB3, HLA-DRB1, JAK2, JAK1, CIITA, RELA, GNAI2, IRAK1, PDPK1, HLA-DQA2, HLA-DQA1, HLA-DQB1, ALOX5, LDLR, IL10RB, IL10RA, TNF, MAP2K3, TLR4, TLR2, CASP8, MAPK8, HLA-DPA1, HLA-DMA, HLA-DMB, STAT1, HLA-DPB1, HLA-DOA, HLA-DOB, SOCS1, HLA-DRA, MYD88, PIK3CD, PIK3CG, BIRC3, PIK3R5, CD40, LAMC2, LAMA2, LAMA3, LAMB3, MAPK13, MAPK11, TGFB1, NFKBIA, BAD
TCR signalling in naïve CD8+ T cells(N)	0.0053	54	30	4.72E-10	1.51E-08	MAP3K8, SHC1, LCP2, CD247, CARD11, PDPK1, PRKCQ, BCL10, MAP3K14, LAT, LCK, VAV1, TRPV6, ZAP70, PTPRC, PTPN6, PRF1, RASSF5, CD28, CD3G, CD3E, CD3D, CD86, CD80, CD8B, CD8A, RASGRP1, GRAP2, STIM1, SOS1
T cell receptor signalling pathway(K)	0.0102	104	43	8.63E-10	2.67E-08	MAP3K8, LCP2, CDC42, CD247, PPP3CC, RELA, CARD11, PDPK1, CD4, PRKCQ, BCL10, ICOS, NFATC2, CSF2, TNF, MAP3K14, LAT, LCK, CBLC, VAV1, ZAP70, CTLA4, PTPRC, PIK3CD, PIK3CG, PTPN6, PIK3R5, CD28, CD3G, CD3E, CD3D, PAK1, PAK6, CD8B, CD8A, RASGRP1, MAPK13, MAPK11, ITK, GRAP2, SOS1, SOS2, NFKBIA
T cell activation(P)	0.0079	81	37	9.71E-10	2.91E-08	LCP2, CDC42, CD247, PPP3CC, RAC2, PRKCQ, HLA-DQA2, HLA-DQA1, NFATC2, NFATC4, LAT, LCK, VAV1, ZAP70, MAPK8, HLA-DPA1, HLA-DMA, HLA-DMB, HLA-DOA, NFKB2, PTPRC, HLA-DRA, PIK3CD, PIK3CG, CD28, CD3G, CD3E, CD3D, PAK1, CD74, CD86, GRAP2, ITPR1, SOS1, SOS2, WAS, NFKBIA
Proteoglycans in cancer(K)	0.0199	203	65	1.69E-09	4.89E-08	FRS2, FN1, EGFR, WNT7B, WNT7A, ITGA2, ITGA5, ROCK1, CDC42, KDR, IGF1, ARHGEF1, VEGFA, PLAU, PDPK1, TNF,

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
						CBLC, TIAM1, TLR4, TLR2, WNT5B, WNT5A, WNT3A, FGFR1, WNT2B, PLAUR, DCN, THBS1, LUM, HSPG2, FLNB, FLNC, PIK3CD, FASLG, PIK3CG, PPP1R12C, PTPN6, ANK3, ANK1, PIK3R5, PAK1, GPC1, PPP1CA, WNT4, MMP2, MMP9, MAPK13, MAPK11, HIF1A, PLCG2, FZD1, WNT10B, WNT10A, FZD7, MRAS, HCLS1, HBEGF, ITPR1, ITPR2, SOS1, SOS2, EZR, FAS, TGFB1, CAMK2D
Viral myocarditis(K)	0.0057	58	30	2.38E-09	6.66E-08	CXADR, HLA-DRB4, HLA-DRB3, HLA-DRB1, ITGB2, ITGAL, RAC2, HLA-DQA2, HLA-DQA1, HLA-DQB1, HLA-B, HLA-F, HLA-E, CASP8, MYH6, HLA-DPA1, HLA-DMA, HLA-DMB, HLA-DPB1, HLA-DOA, HLA-DOB, HLA-DRA, PRF1, CD28, CD40, CD86, CD80, LAMA2, ABL2, EIF4G1
Chagas disease (American trypanosomiasis)(K)	0.0102	104	42	2.63E-09	7.11E-08	CD247, IL1B, RELA, GNAI2, IRAK1, TNF, TLR9, TLR4, TLR2, C1QB, C1QA, C1QC, CASP8, GNA15, MAPK8, PPP2R2C, C3, TICAM1, MYD88, PIK3CD, FASLG, PIK3CG, CCL5, CCL3, CCL2, PIK3R5, CD3G, CD3E, CD3D, CCL3L1, IL6, MAPK13, MAPK11, SERPINE1, ACE, PLCB3, CFLAR, FAS, TGFB1, NFKBIA, BDKRB2, TGFBR2
Herpes simplex infection(K)	0.018	184	60	3.73E-09	9.70E-08	HLA-DRB4, HLA-DRB3, HLA-DRB1, JAK2, JAK1, TNFSF14, IL15, IL1B, RELA, CFP, HLA-DQA2, HLA-DQA1, HLA-DQB1, TNF, IFNAR2, PML, IFIT1, HLA-B, HLA-F, HLA-E, TLR9, TLR3, TLR2, CASP8, OAS1, OAS2, OAS3, LTA, MAPK8, C3, HLA-DPA1, HLA-DMA, HLA-DMB, STAT1, HLA-DPB1, HLA-DOA, HLA-DOB, TICAM1, SOCS3, HLA-DRA, MYD88, FASLG, CCL5, CCL2, PVRL2, PVRL1, CD74, PPP1CA, IL6, IRF7, TAF13, MED8, TAP2, TAP1, TRAF2, TRAF1, TRAF3, ARNTL, FAS, NFKBIA
GPVI-mediated activation cascade(R)	0.0048	49	27	4.05E-09	1.05E-07	JAK2, JAK3, JAK1, SHC1, LCP2, IL2RG, IL2RA, IL2RB, RAC2, FCER1G, SYK, PDPK1, CSF2, LAT, LCK, VAV1, PDPN, LYN, PIK3CD, PIK3CG, PTPN6, PIK3R6, PIK3R5, PLCG2, IL3RA, CSF2RB, CSF2RA
Influenza A(K)	0.0171	175	57	9.49E-09	2.37E-07	HLA-DRB4, HLA-DRB3, HLA-DRB1, JAK2, JAK1, CIITA, TNFSF10, IL18, IL1A, IL1B, RELA, IL33, PRSS3, NUP98, HLA-

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
						DQA2, HLA-DQA1, HLA-DQB1, TNF, IFNAR2, MAP2K3, PML, TLR4, TLR3, CASP1, OAS1, OAS2, OAS3, MAPK8, HLA-DPA1, HLA-DMA, HLA-DMB, STAT1, HLA-DPB1, HLA-DOA, HLA-DOB, TICAM1, SOCS3, TNFRSF10B, TNFRSF10D, CXCL10, HLA-DRA, MYD88, PIK3CD, FASLG, PIK3CG, FURIN, CCL5, CCL2, RSAD2, PIK3R5, IL6, IRF7, MAPK13, MAPK11, PYCARD, FAS, NFKBIA
Phagosome(K)	0.015	153	52	1.11E-08	2.66E-07	HLA-DRB4, HLA-DRB3, HLA-DRB1, PLA2R1, CORO1A, ITGB2, ITGA2, ITGA5, C1R, ATP6V1C2, MSR1, CD209, HLA-DQA2, HLA-DQA1, HLA-DQB1, TCIRG1, HLA-B, HLA-F, HLA-E, RAB7B, TLR4, TLR2, COMP, NOS1, THBS2, THBS1, C3, TUBA1A, HLA-DPA1, HLA-DMA, HLA-DMB, HLA-DPB1, NCF1, NCF2, NCF4, HLA-DOA, HLA-DOB, TUBA4A, HLA-DRA, FCGR3A, FCGR1A, FCGR2B, FCGR2C, CTSS, CD36, STX18, CYBA, MRC2, TAP2, TAP1, CLEC7A, MARCO
Signalling by Interleukins(R)	0.0274	280	78	1.65E-08	3.81E-07	FRS2, PDGFB, NEFL, JAK2, JAK3, JAK1, DUSP5, DUSP7, EGFR, MAP3K8, SHC1, NRG1, IL2RG, IL2RA, IL2RB, IL27RA, LGALS9, STAT5B, IL11, SYK, IL18, IL1A, IL1B, RELA, GRIN1, KITLG, FBXW11, IRAK1, IRAK2, IL7R, PTK2B, CNTF, TOLLIP, CLCF1, CSF2, FGF1, LAT, NOD1, NOD2, IL1RN, IL1R1, LIF, VAV1, FGFR3, FGFR2, FGFR1, BLNK, CASP1, RASA2, PSME2, PELI3, STAT1, SOCS3, NFKB2, MYD88, RASGRF2, PIK3CD, KSR1, PTPN6, OSMR, PDGFRB, PDGFRA, RASGRP1, RASGRP4, RASGRP3, IL6, HAVCR2, HBEGF, RASAL1, RASAL3, IL3RA, SOS1, EBI3, INPP5D, CSF2RB, CSF2RA, MARK3, CAMK2D
MAPK signalling pathway(K)	0.025	255	73	1.66E-08	3.81E-07	CACNA2D1, CACNA2D3, CACNA2D4, DDIT3, PDGFB, DUSP5, DUSP1, DUSP7, EGFR, TAOK2, MAP3K8, MAP3K6, MAP3K5, CDC42, NTRK2, PPP3CC, RAC2, IL1A, IL1B, RELA, RELB, MAP3K2, CACNA1C, CACNA1I, FGF11, MKNK1, FGF1, TNF, MAP2K3, MAP3K14, RPS6KA2, RPS6KA1, IL1R1, FGFR3, FGFR2, FGFR1, RASA2, MAPK8, GNA12, ECSIT, FLNB, FLNC,

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
						MAPKAPK2, MAX, NFKB2, RASGRF2, FASLG, ARRB2, PPM1A, PTPN7, PTPRR, PAK1, PDGFRB, PDGFRA, RASGRP1, RASGRP4, RASGRP3, PLA2G4F, PLA2G4E, MAPK13, MAPK11, MRAS, CACNB1, DUSP10, TRAF2, SOS1, SOS2, GADD45B, GADD45A, MAP4K1, FAS, TGFB1, TGFBR2
Primary immunodeficiency(K)	0.0035	36	22	1.83E-08	4.21E-07	JAK3, CIITA, IL2RG, BTK, IL7R, CD4, ICOS, LCK, CD79A, BLNK, ZAP70, AIRE, PTPRC, CD19, CD40, TNFRSF13B, CD3E, CD3D, CD8B, CD8A, TAP2, TAP1
ECM-receptor interaction(K)	0.0085	87	36	1.94E-08	4.26E-07	FN1, COL3A1, ITGB4, ITGB8, ITGB6, ITGA2, ITGA1, ITGA8, ITGA5, TNC, COL27A1, COMP, COL6A2, COL6A1, COL6A3, COL6A6, THBS2, THBS1, HSPG2, COL1A1, COL1A2, COL11A1, CD36, COL5A1, COL5A3, COL5A2, LAMC2, LAMA2, LAMA3, LAMB3, VWF, TNXB, COL4A2, COL4A1, COL4A4, COL4A3
Amoebiasis(K)	0.0106	108	41	2.17E-08	4.78E-07	FN1, COL3A1, ITGB2, ACTN1, IL1B, RELA, CSF2, TNF, COL27A1, IL1R1, RAB7B, TLR4, TLR2, GNA15, COL1A1, COL1A2, SERPINB13, COL11A1, PIK3CD, PIK3CG, CD1D, PIK3R5, CTSG, IL6, COL5A1, COL5A3, COL5A2, LAMC2, LAMA2, LAMA3, LAMB3, SERPINB3, SERPINB4, SERPINB2, PLCB3, COL4A2, COL4A1, COL4A4, COL4A3, TGFB1, CXCL1
CXCR4-mediated signalling events(N)	0.0078	80	34	2.55E-08	5.35E-07	HLA-DRB1, JAK2, BLK, CDC42, CD247, STAT5B, GNAI2, GRK6, PDPK1, PTK2B, CD4, LCK, VAV1, HCK, LIMK1, LYN, STAT1, PTPRC, CXCL12, HLA-DRA, ARRB2, PTPN6, CD3G, CD3E, CD3D, PAK1, RGS1, SSH1, MMP9, PLCB3, INPP5D, CXCR4, FGR, BAD
NOD-like receptor signalling pathway(K)	0.0056	57	27	8.40E-08	1.76E-06	PYDC1, NAIP, IL18, IL1B, RELA, CARD18, TNFAIP3, TNF, NOD1, NLRC4, NOD2, NLRP1, CASP8, CASP1, MAPK8, BIRC3, CCL5, CCL2, IL6, MAPK13, MAPK11, MEFV, CARD9, PYCARD, NFKBIA, CXCL1, CXCL2
Platelet activation(K)	0.0127	130	44	1.62E-07	3.25E-06	COL3A1, ITGA2, LCP2, ROCK1, BTK, FCER1G, SYK, ARHGEF1, GNAI2, ADCY4, ADCY7, COL27A1, P2RY1, F2RL3, GUCY1B3, GUCY1A3, GUCY1A2, NOS3, MYLK, LYN, COL1A1, COL1A2,

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
						FERMT3, COL11A1, PTGS1, PIK3CD, PIK3CG, PIK3R5, PPP1CA, RASGRP1, COL5A1, COL5A3, COL5A2, PLA2G4F, PLA2G4E, MAPK13, MAPK11, PLCG2, PLCB3, STIM1, ITPR1, ITPR2, VWF, PTGIR
B cell activation(P)	0.0058	59	27	1.64E-07	3.28E-06	BTK, GRAP, RAC2, SYK, MAP3K2, NFATC2, NFATC4, CD79B, CD79A, VAV1, BLNK, MAPK8, LYN, NFKB2, PTPRC, PIK3CD, PIK3CG, CD19, CD22, MAPK13, MAPK11, PLCG2, ITPR1, ITPR2, SOS1, SOS2, NFKBIA
Autoimmune thyroid disease(K)	0.0051	52	25	1.90E-07	3.61E-06	HLA-DRB4, HLA-DRB3, HLA-DRB1, GZMB, HLA-DQA2, HLA-DQA1, HLA-DQB1, HLA-B, HLA-F, HLA-E, CTLA4, HLA-DPA1, HLA-DMA, HLA-DMB, HLA-DPB1, HLA-DOA, HLA-DOB, HLA-DRA, FASLG, PRF1, CD28, CD40, CD86, CD80, FAS
Signalling by NGF(R)	0.0382	390	95	2.32E-07	4.41E-06	FRS2, PDGFB, MOV10, NEFL, JAK2, JAK3, JAK1, DUSP5, DUSP7, EGFR, SHC1, NRG1, ADORA2A, IL2RG, NTRK2, IL2RA, IL2RB, NET1, ARHGEF3, ARHGEF4, ARHGEF1, ARHGEF7, RELA, GRIN1, KITLG, IRAK1, PDPK1, ADCY4, ADCY7, FGD3, CSF2, FGF1, LAT, PLEKHG5, LCK, HDAC2, TNRC6A, RPS6KA2, RPS6KA1, TIAM1, MAGED1, PSENEN, VAV1, FGFR3, FGFR2, FGFR1, RASA2, MAPK8, MEF2A, PRKAR1B, PSME2, NGEF, PRKAR2B, NGFR, MAPKAPK2, BCL2L11, MYD88, RASGRF2, PIK3CD, KSR1, FURIN, CD19, CD28, AP2A1, PCSK6, AATF, CD86, CD80, PDGFRB, PDGFRA, RASGRP1, RASGRP4, RASGRP3, ARHGEF17, MAPK13, MAPK11, HBEGF, PHLPP2, ITPR1, ITPR2, RASAL1, RASAL3, IL3RA, SOS1, SOS2, PDE1B, PDE1A, CAMK4, CSF2RB, CSF2RA, MARK3, NFKBIA, CAMK2D, BAD, TRAT1
Malaria(K)	0.0048	49	24	2.39E-07	4.54E-06	ITGB2, ITGAL, IL18, IL1B, VCAM1, TNF, CR1, PECAM1, TLR9, TLR4, TLR2, COMP, SELE, SELP, THBS2, THBS1, MYD88, LRP1, CCL2, CD40, CD36, IL6, KLRB1, TGFB1
Toll-like receptor signalling pathway(K)	0.0104	106	38	2.84E-07	5.11E-06	MAP3K8, IL1B, RELA, IRAK1, TOLLIP, TNF, IFNAR2, MAP2K3, TLR9, TLR8, TLR4, TLR3, TLR2, CCL4L1, CASP8, MAPK8, STAT1, TICAM1, CXCL10, CXCL11, MYD88, PIK3CD, PIK3CG,

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
						CCL5, CCL3, PIK3R5, CD40, CD86, CD80, CCL3L1, IL6, IRF7, IRF5, MAPK13, MAPK11, TRAF3, NFKBIA, CXCL9
Inflammation mediated by chemokine and cytokine signalling pathway(P)	0.0071	73	30	3.28E-07	5.91E-06	JAK2, CISH, ITGAL, ROCK1, GRAP, RAC2, MAP3K2, GRK6, PDPK1, PTK2B, VAV1, ALOX12, COL6A1, GNA15, MYLK, SOCS3, NFKB2, PIK3CD, PIK3CG, CCR4, PLCG2, PLCB3, PLCD3, PLCD1, VWF, SOS1, PLCL2, NFKBIA, RGS13, ALOX5AP
Ras signalling pathway(K)	0.0222	227	63	4.45E-07	7.69E-06	PDGFB, PDGFC, EGFR, SHC1, CDC42, KDR, IGF1, RAC2, RELA, VEGFC, GRIN1, VEGFA, KITLG, ETS2, PLD2, FGF11, CSF1, FGF1, PGF, LAT, CALML5, CALML3, TIAM1, EFNA5, EFNA3, FGFR3, FGFR2, FGFR1, ZAP70, RASA2, MAPK8, NGFR, FLT4, RASGRF2, PIK3CD, FASLG, PIK3CG, KSR1, RASSF5, EPHA2, PIK3R5, PAK1, PAK6, PDGFRB, PDGFRA, RASGRP1, RASGRP4, RASGRP3, PLA2G4F, PLA2G4E, PLA2G2A, PLCG2, REL, MRAS, CSF1R, RASAL1, RASAL3, ABL2, SOS1, SOS2, GNG10, GNG11, BAD
Asthma(K)	0.0029	30	18	4.52E-07	7.69E-06	HLA-DRB4, HLA-DRB3, HLA-DRB1, FCER1A, FCER1G, MS4A2, HLA-DQA2, HLA-DQA1, HLA-DQB1, TNF, HLA-DPA1, HLA-DMA, HLA-DMB, HLA-DPB1, HLA-DOA, HLA-DOB, HLA-DRA, CD40
Signalling by VEGF(R)	0.0268	274	72	5.02E-07	8.54E-06	FRS2, PDGFB, SHB, NEFL, JAK2, JAK3, JAK1, DUSP5, DUSP7, EGFR, SHC1, NRG1, ROCK1, IL2RG, CDC42, IL2RA, IL2RB, KDR, NRP1, NRP2, VEGFC, GRIN1, VEGFA, KITLG, PDPK1, PTK2B, NCKAP1L, CSF2, FGF1, PGF, LAT, SH2D2A, VAV1, FGFR3, FGFR2, FGFR1, RASA2, NOS3, PSME2, NCF1, NCF2, NCF4, MAPKAPK2, FLT4, RASGRF2, KSR1, BAIAP2, PAK1, PDGFRB, PDGFRA, CYBA, RASGRP1, RASGRP4, RASGRP3, MAPK13, MAPK11, CYFIP2, HBEGF, CDH5, ITPR1, ITPR2, RASAL1, RASAL3, IL3RA, SOS1, WASF2, ELMO1, ELMO2, CSF2RB, CSF2RA, MARK3, CAMK2D
Validated transcriptional targets of AP1 family members Fra1 and Fra2(N)	0.0036	37	20	5.33E-07	9.06E-06	ITGB4, PLAU, GJA1, HMOX1, NFATC2, FOSL2, FOSL1, THBD, LIF, PLAUR, DCN, NOS3, COL1A2, CCL2, IL6, MMP1, MMP2, MMP9, LAMA3, JUNB

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
Pertussis(K)	0.0073	75	30	5.68E-07	9.08E-06	ITGB2, ITGA5, C1S, C1R, IL1A, IL1B, RELA, GNAI2, IRAK1, TNF, NOD1, CALML5, CALML3, TLR4, C1QB, C1QA, C1QC, CASP7, CASP1, MAPK8, C2, C3, TICAM1, MYD88, IL6, IRF8, MAPK13, MAPK11, SERPING1, PYCARD
Fc gamma R-mediated phagocytosis(K)	0.009	92	34	6.06E-07	9.70E-06	ARPC3, CDC42, RAC2, SYK, PLD2, ASAP3, ARPC1B, GSN, LAT, PPAP2B, VAV1, HCK, AMPH, DNM2, LIMK2, LIMK1, LYN, NCF1, PTPRC, PIK3CD, PIK3CG, FCGR3A, FCGR1A, FCGR2B, PIK3R5, PAK1, PLA2G4F, PLA2G4E, BIN1, PLCG2, WASF2, INPP5D, WAS, DOCK2

Table T3-7: Full list of pathways enriched for genes positively correlated (at FDR<0.00001) with tumour *VDR* expression in the TCGA metastatic melanomas

The source database for each of the enriched pathways are indicated by a letter in parentheses after each pathway gene set name. The source database annotations are: C - CellMap, R – Reactome, K – KEGG, N – NCI PID, P - Panther, and B – BioCarta.

The description of column headers are: *Ratio of protein in gene set*: ratios of numbers of genes contained in pathways to total genes in the Reactome FI network; Number of protein in gene set: numbers of genes in pathways; *Protein from network*: numbers of hit genes from the query gene list; *Nodes:* nodal genes from my input query gene list, which 'matches' with the genes in a particular pathway; P-value was estimated by Reactome FIViz using hypergeometric test and the corresponding FDR was estimated by Benjamini Hochberg multiple correction

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
Chemokine signalling pathway(K)	0.0183	187	74	1.11E-16	1.57E-14	LYN, JAK3, STAT1, NCF1, CXCL13, CXCL14, CXCL16, CXCL10, CXCL11, CXCL12, PIK3CD, ARRB2, PIK3CG, CCL8, CCL7, CCL5, CCL4, CCL3, CCL2, PIK3R5, CCR1, GNGT2, RAC2, CCR8, CCR7, CCR6, CCR5, CCR4, CCR3, CCR2, CCL3L1, RASGRP2, PTK2B, PRKCB, ITK, CCL14, CCL13, CCL11, ADCY4, ADCY2, ADCY1, ADCY7, CCL19, CCL18, CCL17, CCL23, CCL22, CCL21, CCL20, CCL26, XCL2, XCL1, CX3CR1, TIAM1, XCR1, VAV1, HCK, CXCR5, WAS, CXCR4, CXCR6, CXCR1, CXCR3, CXCR2, ELMO1, NFKBIA, CXCL6, CXCL9, CXCL1, CXCL3, CXCL2, CXCL5, FGR, DOCK2
Extracellular matrix organization(R)	0.0243	248	96	1.11E-16	1.57E-14	F11R, COL13A1, PDGFB, ADAMTS14, FN1, COL1A1, COL1A2, TPSAB1, BGN, MMP10, MMP12, MMP11, MMP13, COL11A1, JAM2, COL3A1, EFEMP1, IBSP, COL17A1, ITGAM, ITGB4, ITGB3, ITGB2, ITGAL, ITGAX, ITGB8, ITGB7, ITGA4, ITGA3, ITGAD, ITGA7, ITGA5, COL15A1, ICAM2, ICAM3, ICAM4, ICAM1, ACTN1, SDC1, CTSS, CTSK, CTSG, CTSD, CTSB, LOXL1, PCOLCE, VCAM1, COL5A1, COL5A3, COL5A2, LAMC2, MMP7, MMP1, MMP2, MMP3, MMP8, MMP9, VCAN, PLEC, LAMA2, LAMA3, LAMB3, COL7A1, SERPINE1, COL12A1, ELN, COL10A1, TIMP1, FMOD, TNC, COL27A1, ADAM8, VTN, ITGA11, COL14A1, ADAMTS2, PECAM1, MFAP5, MFAP4, TLL1, COL8A2, COL8A1, KLK7, COMP, COL4A4, COL4A3, LTBP1, FBLN2, FBLN5, TGFB1, DCN, COL6A2, COL6A1, COL6A3, THBS1, LUM
GPCR ligand binding(R)	0.0381	389	107	1.11E-16	1.57E-14	HTR7, C3AR1, FPR1, FPR3, FPR2, C3, GPR18, PTAFR, P2RY13, P2RY10, GPR68, GPR65, CXCL13, CXCL16, ADORA1, UTS2, HTR1F, CXCL10, CXCL11, CXCL12, ADORA2A, CCL7, CCL5, CCL4, CCL3, CCL2, LPAR1, CCR1, MC1R, LPAR5, CCR8, CCR7, CCR6, CCR5, CCR4, CCR3, CCR2, RAMP3, RAMP1, CCRL2, CCL3L1, WNT1, WNT2, GRM2, SSTR2, SSTR3, GRPR, ADM, GPR132, ANXA1, TACR1, CCL11, CCL19, CCL17, CCL23, CCL22, CCL21, CCL20, AGT, ADM2, CYSLTR1, CYSLTR2, ADRA2A, XCL2, XCL1, CX3CR1, P2RY6, P2RY2, PTGFR,

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
						NTSR1, XCR1, FFAR2, EDN1, OPRD1, PTGER2, PTGER3, CXCR5, CXCR4, CXCR6, CXCR1, CXCR3, CXCR2, SAA1, F2RL1, GHRL, F2RL3, PMCH, WNT5A, APLNR, SUCNR1, C5AR1, ADRB2, HRH1, HRH2, NMUR1, CXCL6, CXCL9, CXCL1, CXCL3, CXCL2, CXCL5, BDKRB2, BDKRB1, PTGIR, S1PR1, S1PR4, POMC
Cytokine-cytokine receptor interaction(K)	0.0259	265	133	1.11E-16	1.57E-14	PDGFB, IL20RA, IL20RB, IL22RA2, NGFR, IL22RA1, CTF1, EGFR, IL12RB1, IL12RB2, CXCL13, CXCL14, CXCL16, TNFRSF8, TNFRSF4, TNFRSF11B, TNFRSF11A, TNFRSF17, TNFRSF9, TNFRSF10C, TNFRSF10A, TNFRSF10D, CXCL10, CXCL11, CXCL12, FLT3, FASLG, TNFSF13B, TNFRSF1B, IL2RG, CCL8, CCL7, CCL5, CCL4, CCL3, CCL2, IL2RA, IL2RB, TNFRSF6B, TNFSF10, TNFSF11, IL12B, IL12A, CCR1, TNFSF14, TNFSF13, CD27, CD40, TNFRSF13B, IL24, CCR8, CCR7, CCR6, CCR5, CCR4, CCR3, CCR2, IL10, IL11, TNFRSF12A, IL15, IL18, IL1A, IL1B, CD70, OSM, VEGFC, KITLG, IL18R1, PDGFRA, IL4R, CCL3L1, INHBA, IL6, IL7, IL7R, IL9R, TNFSF4, TNFSF8, CCL14, CCL13, CCL11, CCL19, CCL18, CCL17, CCL23, CCL22, CCL21, CCL20, IL10RA, CLCF1, CCL26, CSF3, CSF3R, CSF2, CSF1, TNF, EDAR, CSF1R, IL3RA, XCL2, XCL1, CX3CR1, IL21R, IL15RA, IL1R1, IL1R2, FLT3LG, IL23A, XCR1, LIF, CXCR5, CXCR4, CXCR6, CXCR1, CXCR3, CXCR2, CD40LG, IL18RAP, HGF, CSF2RB, CSF2RA, TGFB1, BMPR1B, CXCL6, CXCL9, CXCL1, CXCL3, CXCL2, CXCL5, IFNG, LTA, LTB
Osteoclast differentiation(K)	0.0128	131	61	1.11E-16	1.57E-14	SPI1, STAT1, NCF1, NCF2, NCF4, TNFRSF11B, TNFRSF11A, SOCS3, SOCS1, NFKB2, ITGB3, FHL2, PIK3CD, PIK3CG, FCGR3A, FCGR3B, LCP2, FCGR1A, FCGR2A, FCGR2B, TNFSF11, BTK, PIK3R5, CTSK, SYK, IL1A, IL1B, RELB, CYBB, CYBA, OSCAR, TYROBP, IRF9, MAPK13, TEC, JUNB, PLCG2, FOSL2, FOSL1, CYLD, CSF1, TNF, CSF1R, TREM2, LCK, IL1R1, LILRA6, LILRA1, LILRA3, LILRA4, LILRA5, LILRB1, LILRB2, LILRB4, LILRB5, CAMK4, SIRPG, TGFB1, NFKBIA, BLNK, IFNG

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
Hematopoietic cell lineage(K)	0.0085	87	57	1.11E-16	1.57E-14	ITGAM, FLT3, ITGB3, ITGA4, ITGA3, ITGA5, FCGR1A, IL2RA, ANPEP, CD1E, CD1C, CD1B, CD1A, CD19, CD14, CD24, CD22, CD3G, CD3E, CD3D, CD38, CD37, CD33, IL11, MME, IL1A, IL1B, KITLG, IL4R, CD8B, CD8A, IL6, IL7, IL7R, MS4A1, IL9R, CD2, CD4, CD5, CD7, CD9, GP5, FCER2, CSF3, CSF3R, CSF2, CSF1, TNF, CSF1R, CR2, CR1, IL3RA, IL1R1, IL1R2, FLT3LG, GP1BA, CSF2RA
NF-kappa B signalling pathway(K)	0.0089	91	45	3.33E-16	3.53E-14	LYN, PTGS2, TICAM2, LY96, TNFRSF11A, NFKB2, CXCL12, TNFSF13B, BIRC3, CCL4, ICAM1, TNFSF11, BTK, CD14, TNFSF14, CD40, SYK, IL1B, RELB, PLAU, CARD11, VCAM1, TNFAIP3, BCL2A1, PRKCB, PRKCQ, PLCG2, CCL13, CCL19, CCL21, TNF, LAT, LBP, TRAF1, LCK, IL1R1, GADD45B, TLR4, CD40LG, NFKBIA, CXCL2, BLNK, ZAP70, LTA, LTB
IL12-mediated signalling events(N)	0.006	61	37	3.33E-16	3.53E-14	STAT4, IL12RB1, IL12RB2, STAT1, SOCS1, NFKB2, FASLG, IL2RG, CCL4, CCL3, GZMA, GZMB, IL2RA, IL2RB, CD247, IL12B, IL12A, CD3G, CD3E, CD3D, CCR5, IL18, IL1B, RELB, IL18R1, EOMES, CD8B, CD8A, TBX21, CD4, LCK, B2M, IL1R1, GADD45B, GADD45G, IL18RAP, IFNG
Rheumatoid arthritis(K)	0.0087	89	42	1.49E-14	1.40E-12	CTLA4, TNFRSF11A, CXCL12, ITGB2, ITGAL, TNFSF13B, CCL5, CCL3, CCL2, ICAM1, TNFSF11, TNFSF13, CD28, CTSK, IL11, IL15, IL18, IL1A, IL1B, CD86, CD80, CCL3L1, IL6, MMP1, MMP3, TEK, CCL20, CSF2, CSF1, TNF, TCIRG1, IL23A, ATP6V0D2, ATP6V0A4, TLR4, TLR2, TGFB1, CXCL6, CXCL1, CXCL5, IFNG, LTB
Natural killer cell mediated cytotoxicity(K)	0.0131	134	52	2.39E-14	2.01E-12	KLRK1, HCST, KIR3DL1, KIR3DL2, TNFRSF10C, TNFRSF10A, TNFRSF10D, ITGB2, PIK3CD, FASLG, ITGAL, PIK3CG, FCGR3A, FCGR3B, LCP2, GZMB, PTPN6, CD247, CD244, PRF1, ICAM2, ICAM1, TNFSF10, PIK3R5, RAC2, FCER1G, SYK, NCR3, CD48, MICB, TYROBP, KIR2DL3, KIR2DL4, PTK2B, PRKCB, PLCG2, KIR2DS4, CSF2, TNF, LAT, LCK, SH2D1A, SH2D1B, RAET1E, RAET1G, VAV1, KLRC1, KLRC2, KLRD1, ULBP2, ZAP70, IFNG

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
Amoebiasis(K)	0.0106	108	46	3.10E-14	2.39E-12	FN1, COL1A1, COL1A2, SERPINB13, COL11A1, COL3A1, ITGAM, ITGB2, PIK3CD, PIK3CG, IL12B, IL12A, ACTN1, PIK3R5, CD14, C8G, CTSG, IL10, IL1B, IL6, COL5A1, COL5A3, COL5A2, LAMC2, LAMA2, LAMA3, LAMB3, LAMB4, PRKCB, SERPINB3, SERPINB4, SERPINB2, ADCY1, CSF2, TNF, COL27A1, IL1R1, IL1R2, TLR4, TLR2, COL4A4, COL4A3, TGFB1, CXCL1, IFNG, GNA15
Malaria(K)	0.0048	49	30	1.53E-13	1.07E-11	KLRK1, ITGB2, ITGAL, CCL2, ICAM1, IL12A, SDC1, CD40, IL10, IL18, IL1B, VCAM1, IL6, CSF3, TNF, CR1, PECAM1, TLR9, TLR4, TLR2, CD40LG, COMP, HGF, KLRB1, TGFB1, SELE, SELP, IFNG, THBS2, THBS1
TNF signalling pathway(K)	0.0108	110	45	2.30E-13	1.49E-11	CASP10, RIPK3, PTGS2, MAP3K8, MAP3K5, SOCS3, CXCL10, PIK3CD, PIK3CG, TNFRSF1B, BIRC3, CCL5, CCL2, ICAM1, PIK3R5, IL15, IL1B, VEGFC, IL18R1, VCAM1, IL6, TNFAIP3, MMP3, MMP9, MAPK13, CREB3L1, JUNB, CCL20, CSF2, CSF1, TNF, BCL3, TRAF1, CEBPB, NOD2, LIF, EDN1, MLKL, SELE, NFKBIA, CXCL1, CXCL3, CXCL2, CXCL5, LTA
Beta1 integrin cell surface interactions(N)	0.0065	66	34	4.47E-13	2.68E-11	FN1, COL1A1, COL1A2, COL11A1, JAM2, COL3A1, ITGA4, ITGA3, ITGA7, ITGA5, CD14, PLAU, TGM2, NPNT, VCAM1, COL5A1, COL5A2, LAMC2, LAMA2, LAMA3, LAMB3, COL7A1, TNC, VTN, ITGA11, COL4A4, COL4A3, TGFBI, PLAUR, COL6A2, COL6A1, COL6A3, THBS2, THBS1
Staphylococcus aureus infection(K)	0.0054	55	31	5.00E-13	2.80E-11	C3AR1, FPR1, FPR3, FPR2, C2, C3, PTAFR, ITGAM, ITGB2, ITGAL, FCGR3A, FCGR3B, MBL2, DSG1, C1S, C1R, FCGR1A, FCGR2A, FCGR2B, ICAM1, IL10, SELPLG, MASP1, CFB, CFD, CFH, C1QB, C1QA, C1QC, C5AR1, SELP
Tuberculosis(K)	0.0173	177	58	7.81E-13	4.14E-11	C3, CASP10, SPHK1, STAT1, PLA2R1, CORO1A, CIITA, ITGAM, ITGB2, FCGR3A, FCGR3B, ITGAX, FCGR1A, FCGR2A, FCGR2B, IL12B, IL12A, CD14, CD209, CTSS, CTSD, IL10, FCER1G, SYK, IL18, IL1A, IL1B, CD74, IRAK2, IL6, MAPK13, PLK3, MRC2, MRC1, IL10RA, CARD9, TNF, CLEC4E, CLEC7A, LSP1, TCIRG1, LBP, CR1, CEBPB, NOD2, CALML5, CALML3,

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
						IL23A, ATP6V0D2, ATP6V0A4, TLR1, TLR9, TLR6, TLR4, TLR2, TGFB1, IFNG, CAMK2A
ECM-receptor interaction(K)	0.0085	87	38	2.55E-12	1.25E-10	FN1, COL1A1, COL1A2, COL11A1, COL3A1, IBSP, ITGB4, ITGB3, ITGB8, ITGB7, ITGA4, ITGA3, ITGA7, ITGA5, SDC1, COL5A1, COL5A3, COL5A2, LAMC2, LAMA2, LAMA3, LAMB3, LAMB4, GP5, TNC, COL27A1, VTN, ITGA11, GP1BA, TNXB, COMP, COL4A4, COL4A3, COL6A2, COL6A1, COL6A3, THBS2, THBS1
Beta2 integrin cell surface interactions(N)	0.0028	29	22	4.70E-12	2.21E-10	F11R, THY1, C3, ITGAM, ITGB2, ITGAL, ITGAX, ITGAD, FCGR2A, SPON2, ICAM2, ICAM3, ICAM4, ICAM1, PLAU, VCAM1, PROC, CYR61, GP1BA, CD40LG, TGFBI, PLAUR
Signalling by Interleukins(R)	0.0274	280	75	5.86E-12	2.58E-10	PSME2, PDGFB, GRIN2A, GRIN2D, JAK3, DUSP5, CTF1, EGFR, IL12RB2, STAT1, PSMB8, PSMB9, MAP3K8, SOCS3, NFKB2, RASGRF2, RASGRF1, PIK3CD, NRG1, IL2RG, IL2RA, IL2RB, PTPN6, IL27RA, IL12A, LGALS9, EREG, IL27, IL11, SYK, IL18, IL1A, IL1B, OSM, GRIN1, KITLG, PDGFRA, RASGRP1, RASGRP4, RASGRP3, IRAK2, IL6, IL7, IL7R, RASGEF1A, IRAK3, PSMB10, PTK2B, TEC, TEK, HAVCR2, CLCF1, CSF2, FGF1, FGF7, RASAL1, RASAL3, LAT, IL3RA, NOD2, IL1RN, IL1R1, IL1R2, EBI3, INPP5D, LIF, VAV1, HGF, CSF2RB, CSF2RA, FGFR2, BLNK, CASP1, RASA4, CAMK2A
Cell adhesion molecules (CAMs)(K)	0.0139	142	49	8.22E-12	3.45E-10	F11R, CLDN5, CLDN4, CLDN3, L1CAM, CTLA4, JAM2, SIGLEC1, CD274, PTPRC, CNTNAP2, ITGAM, ITGB2, ITGAL, SPN, ITGB8, ITGB7, ITGA4, PDCD1, ICAM2, ICAM3, ICAM1, CD226, CD28, SDC1, CD22, CD40, SELPLG, CD86, CD80, CD8B, CD8A, VCAM1, VCAN, CD2, CD4, CD6, ICOS, PECAM1, NFASC, ICOSLG, CLDN23, PDCD1LG2, CLDN14, CD40LG, SELE, SELP, SELL, CDH15
Cell surface interactions at the vascular wall(R)	0.0098	100	40	9.56E-12	3.83E-10	F11R, CXADR, LYN, FN1, L1CAM, JAM2, ITGAM, ITGB3, ITGB2, ITGAL, SPN, ITGAX, ITGA4, ITGA3, ITGA5, PTPN6, CD244, GRB7, FCER1G, CD48, SELPLG, CD74, PROCR, DOK2, PROC,

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
						MMP1, CD2, TEK, GAS6, CD177, TREM1, THBD, LCK, PECAM1, INPP5D, SIRPG, SELE, SELP, SELL, AMICA1
Measles(K)	0.0131	134	46	4.22E-11	1.60E-09	JAK3, IKBKE, STAT1, TNFRSF10C, TNFRSF10A, TNFRSF10D, PIK3CD, FASLG, PIK3CG, IL2RG, IL2RA, IL2RB, FCGR2B, TNFSF10, IL12B, IL12A, PIK3R5, SLAMF1, CD209, CD28, TP73, CD3G, CD3E, CD3D, IL1A, IL1B, CCND2, IL6, TNFAIP3, IRF7, IRF9, MX1, PRKCQ, IFIH1, TACR1, SH2D1A, TLR9, TLR7, TLR4, TLR2, HSPA6, NFKBIA, OAS1, IFNG, OAS2, OAS3
Primary immunodeficiency(K)	0.0035	36	23	4.48E-11	1.61E-09	JAK3, AICDA, PTPRC, CIITA, IL2RG, BTK, CD19, CD40, TNFRSF13B, CD3E, CD3D, CD8B, CD8A, IL7R, CD4, ICOS, TAP2, TAP1, LCK, CD79A, CD40LG, BLNK, ZAP70
Pertussis(K)	0.0073	75	33	5.74E-11	2.01E-09	C2, C3, TICAM2, LY96, ITGAM, ITGB2, ITGA5, C1S, C1R, IL12B, IL12A, CD14, IL10, IL1A, IL1B, IL6, IRF1, IRF8, MAPK13, SERPING1, TNF, PYCARD, CALML5, CALML3, IL23A, TLR4, C1QB, C1QA, C1QC, NLRP3, CXCL6, CXCL5, CASP1
TCR signalling in naïve CD8+ T cells(N)	0.0053	54	27	2.11E-10	6.96E-09	MAP3K8, PTPRC, LCP2, PTPN6, CD247, PRF1, RASSF5, CD28, CD3G, CD3E, CD3D, CD86, CD80, CD8B, CD8A, CARD11, RASGRP2, RASGRP1, PRKCB, PRKCQ, GRAP2, LAT, LCK, B2M, VAV1, TRPV6, ZAP70
IL23-mediated signalling events(N)	0.0035	36	22	2.66E-10	8.44E-09	STAT4, IL12RB1, STAT1, SOCS3, ITGA3, CCL2, IL12B, IL24, CD3E, IL18, IL1B, IL18R1, IL6, CD4, TNF, IL23A, ALOX12B, IL18RAP, NFKBIA, CXCL9, CXCL1, IFNG
TCR signalling in naïve CD4+ T cells(N)	0.0066	67	30	2.83E-10	8.44E-09	SLA2, MAP3K8, PTPRC, LCP2, FYB, PTPN6, CD247, RASSF5, CD28, CD3G, CD3E, CD3D, CD86, CD80, CARD11, RASGRP2, RASGRP1, CD4, PRKCB, PRKCQ, ITK, GRAP2, LAT, LCK, INPP5D, MAP4K1, VAV1, WAS, TRPV6, ZAP70
Integrin signalling pathway(P)	0.0155	158	49	2.91E-10	8.44E-09	COL13A1, FN1, COL1A1, COL1A2, COL11A1, COL3A1, MAP3K5, COL17A1, ITGAM, ITGB4, ITGB3, ITGB2, PIK3CD, ITGAL, PIK3CG, ITGAX, ITGB8, ITGB7, ITGA4, ITGA3, ITGAD, ITGA7, ITGA5, COL15A1, ACTN1, RRAS, ITGBL1, GRAP, RAC2, COL5A1, COL5A3, COL5A2, RND1, PTK2B, MAPK13, COL7A1,

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
						COL12A1, COL10A1, ITGA11, COL14A1, PIK3C2B, COL8A2, COL8A1, COL4A4, COL4A3, ELMO1, COL6A2, COL6A1, COL6A3
Jak-STAT signalling pathway(K)	0.0155	158	49	2.91E-10	8.44E-09	IL20RA, IL20RB, IL22RA2, IL22RA1, JAK3, CTF1, STAT4, IL12RB1, IL12RB2, STAT1, PIM1, CISH, SOCS3, SOCS1, PIK3CD, PIK3CG, IL2RG, IL2RA, IL2RB, PTPN6, IL27RA, IL12B, IL12A, PIK3R5, IL24, IL10, IL11, IL15, OSM, CCND2, IL4R, IL6, IL7, IL7R, IL9R, IRF9, IL10RA, CSF3, CSF3R, CSF2, IL3RA, IL21R, IL15RA, IL23A, LIF, CSF2RB, CSF2RA, AOX1, IFNG
Toll-like receptor signalling pathway(K)	0.0104	106	38	6.44E-10	1.80E-08	IKBKE, STAT1, MAP3K8, TICAM2, LY96, CXCL10, CXCL11, PIK3CD, PIK3CG, CCL5, CCL4, CCL3, IL12B, IL12A, PIK3R5, CD14, CD40, CTSK, IL1B, CD86, CD80, CCL3L1, IL6, IRF7, IRF5, MAPK13, TNF, LBP, TLR1, TLR9, TLR8, TLR7, TLR6, TLR5, TLR4, TLR2, NFKBIA, CXCL9
Gastrin-CREB signalling pathway via PKC and MAPK(R)	0.0354	362	82	1.66E-09	4.48E-08	PSME2, PDGFB, FPR2, GRIN2A, GRIN2D, JAK3, DUSP5, PTAFR, EGFR, P2RY10, PSMB8, PSMB9, GPR68, GPR65, UTS2, RASGRF2, RASGRF1, NRG1, IL2RG, IL2RA, IL2RB, EREG, LPAR1, LPAR5, GRIN1, KITLG, RGS2, PDGFRA, RASGRP2, RASGRP1, RASGRP4, RASGRP3, RASGEF1A, MMP3, PSMB10, TEK, GRPR, PRKCH, PRKCQ, GPR132, ANXA1, TACR1, CCL23, AGT, DGKG, CSF2, DGKA, FGF1, CYSLTR1, FGF7, CYSLTR2, RASAL1, RASAL3, LAT, IL3RA, XCL2, XCL1, P2RY6, P2RY2, RPS6KA2, RPS6KA1, PTGFR, NTSR1, XCR1, FFAR2, EDN1, SAA1, F2RL1, GHRL, F2RL3, PMCH, CSF2RB, CSF2RA, HRH1, NMUR1, FGFR2, BDKRB2, BDKRB1, RASA4, RGS18, CAMK2A, GNA15
PI3K-Akt signalling pathway(K)	0.0338	345	79	2.09E-09	5.44E-08	PDGFB, NGFR, FN1, COL1A1, COL1A2, JAK3, COL11A1, EGFR, COL3A1, IBSP, ITGB4, ITGB3, PIK3CD, FASLG, PIK3CG, ITGB8, ITGB7, ITGA4, ITGA3, ITGA7, ITGA5, IL2RG, IL2RA, IL2RB, PIK3AP1, LPAR1, PIK3R5, CD19, IGF1, LPAR5, GNGT2, SYK, OSM, VEGFC, KITLG, CCND2, PDGFRA, IL4R, IL6, FGF14, COL5A1, IL7, COL5A3, COL5A2, IL7R, FGF11, LAMC2, LAMA2, LAMA3, LAMB3, LAMB4, TEK, CREB3L1, MYB, CSF3, CSF3R,

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
						CSF1, TNC, FGF1, FGF7, COL27A1, CSF1R, VTN, ITGA11, IL3RA, TLR4, TLR2, TNXB, COMP, HGF, COL4A4, COL4A3, FGFR2, COL6A2, COL6A1, COL6A3, THBS2, THBS1, PPP2R2C
Pathways in cancer(K)	0.0389	397	87	2.41E-09	6.03E-08	SPI1, FZD10, PDGFB, FN1, WNT9A, PTGS2, EGFR, EGLN3, STAT1, WNT7B, WNT7A, NFKB2, CXCL12, FLT3, PIK3CD, FASLG, PIK3CG, ITGA3, BIRC3, NTRK1, RASSF5, LPAR1, TGFA, PIK3R5, IGF1, LPAR5, GNGT2, RAC2, VEGFC, KITLG, PDGFRA, TCF7, PLD1, RASGRP2, RASGRP1, RASGRP4, RASGRP3, WNT1, WNT2, IL6, FGF14, FGF11, LAMC2, MMP1, MMP2, MMP9, LAMA2, LAMA3, LAMB3, LAMB4, PRKCB, DAPK2, PLCG2, WNT10B, WNT10A, RET, PAX8, ADCY4, ADCY2, ADCY1, ADCY7, CSF3R, FGF1, FGF7, CSF1R, PLEKHG5, TRAF1, CEBPA, CBLC, FLT3LG, RUNX1, PTGER2, PTGER3, CXCR4, F2RL3, WNT5B, HGF, WNT5A, COL4A4, COL4A3, CSF2RA, TGFB1, NFKBIA, FGFR2, BDKRB2, BDKRB1, WNT16
Inflammatory bowel disease (IBD)(K)	0.0064	65	28	2.51E-09	6.03E-08	GATA3, STAT4, IL12RB1, IL12RB2, STAT1, RORC, FOXP3, IL2RG, IL12B, IL12A, IL10, IL18, IL1A, IL1B, IL18R1, IL4R, IL6, TBX21, TNF, NOD2, IL21R, IL23A, TLR5, TLR4, TLR2, IL18RAP, TGFB1, IFNG
GPVI-mediated activation cascade(R)	0.0048	49	24	3.09E-09	7.42E-08	LYN, JAK3, PIK3CD, PIK3CG, LCP2, IL2RG, IL2RA, IL2RB, PTPN6, PIK3R6, PIK3R5, RAC2, FCER1G, SYK, PLCG2, CSF2, LAT, LCK, IL3RA, RHOG, VAV1, PDPN, CSF2RB, CSF2RA
Leukocyte transendothelial migration(K)	0.0116	118	39	3.51E-09	8.08E-08	F11R, CLDN5, CLDN4, CLDN3, MYL9, THY1, JAM2, NCF1, NCF2, NCF4, CXCL12, ITGAM, ITGB2, PIK3CD, ITGAL, PIK3CG, ITGA4, ICAM1, RASSF5, ACTN1, PIK3R5, RAC2, CYBB, CYBA, VCAM1, MMP2, MMP9, PTK2B, MAPK13, PRKCB, PLCG2, ITK, PECAM1, EZR, CLDN23, RHOH, VAV1, CLDN14, CXCR4
Urokinase-type plasminogen activator (uPA) and uPAR-mediated signalling(N)	0.0041	42	22	4.26E-09	9.21E-08	FPR1, FPR3, FPR2, FN1, MMP12, MMP13, EGFR, ITGAM, ITGB3, ITGB2, ITGA3, ITGA5, CTSG, PLAU, MMP3, MMP9, SERPINE1, VTN, KLK4, HGF, TGFB1, PLAUR

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
Chagas disease (American trypanosomiasis)(K)	0.0102	104	36	4.39E-09	9.21E-08	C3, PIK3CD, FASLG, PIK3CG, CCL5, CCL3, CCL2, CD247, IL12B, IL12A, PIK3R5, CD3G, CD3E, CD3D, IL10, IL1B, CCL3L1, IL6, MAPK13, SERPINE1, ACE, ADCY1, TNF, TLR9, TLR6, TLR4, TLR2, C1QB, C1QA, C1QC, TGFB1, NFKBIA, BDKRB2, IFNG, GNA15, PPP2R2C
T cell receptor signalling pathway(K)	0.0102	104	36	4.39E-09	9.21E-08	CTLA4, MAP3K8, PTPRC, PIK3CD, PIK3CG, LCP2, PTPN6, PDCD1, CD247, PIK3R5, CD28, CD3G, CD3E, CD3D, IL10, CD8B, CD8A, CARD11, RASGRP1, MAPK13, CD4, TEC, PRKCQ, ICOS, ITK, CSF2, TNF, GRAP2, LAT, LCK, CBLC, VAV1, CD40LG, NFKBIA, ZAP70, IFNG
Leishmaniasis(K)	0.007	72	29	5.68E-09	1.19E-07	C3, PTGS2, STAT1, NCF1, NCF2, NCF4, ITGAM, ITGB2, FCGR3A, FCGR3B, ITGA4, FCGR1A, FCGR2A, PTPN6, IL12B, IL12A, IL10, IL1A, IL1B, CYBA, MAPK13, PRKCB, TNF, CR1, TLR4, TLR2, TGFB1, NFKBIA, IFNG
IL4-mediated signalling events(N)	0.0063	64	27	7.37E-09	1.47E-07	SPI1, THY1, COL1A1, COL1A2, PIGR, JAK3, SOCS3, SOCS1, AICDA, ITGB3, IL2RG, PTPN6, EGR2, IL10, IL4R, DOK2, MYB, FCER2, CCL11, CCL17, CCL26, CEBPB, INPP5D, PARP14, CD40LG, SELP, LTA
Complement and coagulation cascades(K)	0.0068	69	28	8.84E-09	1.77E-07	C3AR1, C2, C3, MBL2, C1S, C1R, C8G, PLAU, MASP1, PROC, SERPINE1, SERPINF2, SERPING1, CFB, CFD, CFH, SERPINA1, SERPIND1, THBD, CR2, CR1, C1QB, C1QA, C1QC, C5AR1, BDKRB2, BDKRB1, PLAUR
IL27-mediated signalling events(N)	0.0025	26	17	1.05E-08	2.00E-07	GATA3, STAT4, IL12RB1, IL12RB2, STAT1, IL27RA, IL12B, IL12A, IL27, IL18, IL1B, IL6, TBX21, TNF, EBI3, TGFB1, IFNG
DAP12 interactions(R)	0.0292	298	69	1.47E-08	2.80E-07	PSME2, PDGFB, GRIN2A, KLRK1, PRKAR2B, GRIN2D, JAK3, DUSP5, EGFR, PSMB8, PSMB9, RASGRF2, RASGRF1, PIK3CD, NRG1, LCP2, IL2RG, IL2RA, IL2RB, EREG, BTK, CD19, CD28, SYK, GRIN1, KITLG, CD86, CD80, PDGFRA, TYROBP, RASGRP1, RASGRP4, RASGRP3, RASGEF1A, PSMB10, TEK, PLCG2, KIR2DS4, ADCY4, ADCY2, ADCY1, ADCY7, CSF2, FGF1, FGF7, GRAP2, CLEC5A, RASAL1, RASAL3, CD300LB,

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
						LAT, TREM2, TREM1, LCK, IL3RA, B2M, SIGLEC15, SIGLEC14, VAV1, CAMK4, CD300E, CSF2RB, CSF2RA, KLRC2, KLRD1, FGFR2, RASA4, CAMK2A, TRAT1
IL12 signalling mediated by STAT4(N)	0.003	31	18	2.34E-08	4.21E-07	STAT4, IL2RA, CD247, PRF1, CD28, CD3G, CD3E, CD3D, IL18, IL18R1, CD86, CD80, IRF1, TBX21, CD4, IL18RAP, TGFB1, IFNG
Intestinal immune network for IgA production(K)	0.0046	47	22	3.03E-08	5.45E-07	PIGR, TNFRSF17, AICDA, CXCL12, TNFSF13B, ITGB7, ITGA4, TNFSF13, CD28, CD40, TNFRSF13B, IL10, IL15, CD86, CD80, IL6, ICOS, ICOSLG, IL15RA, CXCR4, CD40LG, TGFB1
Beta3 integrin cell surface interactions(N)	0.0042	43	21	3.07E-08	5.53E-07	F11R, PDGFB, THY1, FN1, L1CAM, COL1A1, COL1A2, SPHK1, IBSP, ITGB3, SDC1, PLAU, TNC, VTN, CYR61, PECAM1, COL4A4, COL4A3, TGFBI, PLAUR, THBS1
Focal adhesion(K)	0.0203	207	53	3.30E-08	5.61E-07	MYL9, PDGFB, FN1, COL1A1, COL1A2, COL11A1, EGFR, COL3A1, IBSP, ITGB4, ITGB3, RASGRF1, PIK3CD, PIK3CG, ITGB8, ITGB7, ITGA4, ITGA3, ITGA7, ITGA5, BIRC3, ACTN1, PIK3R5, IGF1, RAC2, VEGFC, CCND2, PDGFRA, COL5A1, COL5A3, COL5A2, LAMC2, LAMA2, LAMA3, LAMB3, LAMB4, PRKCB, TNC, COL27A1, VTN, ITGA11, PARVG, VAV1, TNXB, COMP, HGF, COL4A4, COL4A3, COL6A2, COL6A1, COL6A3, THBS2, THBS1
Validated transcriptional targets of AP1 family members Fra1 and Fra2(N)	0.0036	37	19	6.41E-08	1.09E-06	COL1A2, ITGB4, CCL2, PLAU, IL6, MMP1, MMP2, MMP9, LAMA3, GJA1, JUNB, HMOX1, FOSL2, FOSL1, IVL, THBD, LIF, PLAUR, DCN
Legionellosis(K)	0.0054	55	23	1.08E-07	1.73E-06	C3, NFKB2, ITGAM, ITGB2, IL12B, IL12A, CD14, IL18, IL1B, IL6, TNF, PYCARD, CR1, NLRC4, TLR5, TLR4, TLR2, HSPA6, NFKBIA, CXCL1, CXCL3, CXCL2, CASP1
Platelet activation(K)	0.0127	130	38	1.24E-07	1.99E-06	LYN, COL1A1, COL1A2, FERMT3, COL11A1, PTGS1, COL3A1, ITGB3, PIK3CD, PIK3CG, LCP2, FCGR2A, BTK, PIK3R5, FCER1G, SYK, RASGRP2, RASGRP1, COL5A1, COL5A3,

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
						COL5A2, PLA2G4F, PLA2G4D, PLA2G4E, MAPK13, PLCG2, GP5, APBB1IP, ADCY4, ADCY2, ADCY1, ADCY7, COL27A1, GP1BA, F2RL3, GUCY1B3, GUCY1A3, PTGIR
Interferon gamma signalling(R)	0.0072	74	27	1.32E-07	2.11E-06	PTAFR, STAT1, SOCS3, SOCS1, CIITA, FCGR1A, PTPN6, IFI30, ICAM1, MT2A, TRIM21, TRIM22, VCAM1, IRF1, IRF7, IRF8, IRF5, IRF9, B2M, OASL, OAS1, IFNG, OAS2, OAS3, CAMK2A, GBP2, GBP1
Toll-Like Receptors Cascades(R)	0.0129	132	38	1.80E-07	2.74E-06	LGMN, IKBKE, RIPK3, TLR10, MAP3K8, TICAM2, LY96, SOCS1, LY86, NFKB2, ITGAM, ITGB2, BIRC3, BTK, CD14, CTSS, CTSK, CTSB, IRAK2, IRAK3, IRF7, PLCG2, S100A12, CD180, LBP, NOD2, RPS6KA2, RPS6KA1, TLR1, TLR9, TLR8, TLR7, TLR6, TLR5, TLR4, TLR2, SAA1, NFKBIA
Protein digestion and absorption(K)	0.0088	90	30	1.83E-07	2.74E-06	COL13A1, COL1A1, COL1A2, FXYD2, COL11A1, COL3A1, CPA3, COL17A1, COL15A1, KCNQ1, MME, COL5A1, COL5A3, COL5A2, COL7A1, COL12A1, PRSS3, ELN, COL10A1, SLC8A1, COL27A1, COL14A1, MEP1A, COL4A4, COL4A3, DPP4, COL6A2, COL6A1, COL6A3, ATP1A3
Interferon alpha/beta signalling(R)	0.0065	66	25	1.93E-07	2.90E-06	STAT1, PSMB8, SOCS3, SOCS1, PTPN6, RSAD2, IFI27, BST2, ISG15, ISG20, IRF1, IRF7, IRF8, IRF5, IRF9, SAMHD1, MX1, IFITM1, IFI6, OASL, OAS1, OAS2, OAS3, USP18, GBP2
NOD-like receptor signalling pathway(K)	0.0056	57	23	2.01E-07	3.01E-06	BIRC3, CCL5, CCL2, IL18, IL1B, CARD18, IL6, TNFAIP3, MAPK13, MEFV, CARD9, TNF, PYCARD, PSTPIP1, NLRC4, NOD2, NLRP3, NLRP1, NFKBIA, CXCL1, CXCL2, CASP5, CASP1
Proteoglycans in cancer(K)	0.0199	203	50	2.50E-07	3.51E-06	FZD10, FN1, WNT9A, EGFR, WNT7B, WNT7A, HPSE, ITGB3, PIK3CD, FASLG, PIK3CG, ITGA5, PTPN6, IL12B, ANK3, ANK1, PIK3R5, RRAS, IGF1, SDC1, PLAU, WNT1, WNT2, MMP2, MMP9, MAPK13, PRKCB, PLCG2, WNT10B, WNT10A, HCLS1, TNF, VTN, CBLC, TIAM1, EZR, TWIST2, TLR4, TLR2, WNT5B, HGF, WNT5A, TGFB1, ESR1, PLAUR, WNT16, DCN, CAMK2A, THBS1, LUM

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
B cell receptor signalling pathway(K)	0.007	72	26	2.72E-07	3.80E-06	LYN, PIK3CD, PIK3CG, PTPN6, FCGR2B, PIK3AP1, BTK, PIK3R5, CD19, CD22, RAC2, SYK, CD72, CARD11, DAPP1, RASGRP3, PRKCB, PLCG2, CR2, IFITM1, CD79B, CD79A, INPP5D, VAV1, NFKBIA, BLNK
Fc gamma R-mediated phagocytosis(K)	0.009	92	30	2.87E-07	4.02E-06	LYN, SPHK1, NCF1, PTPRC, PIK3CD, PIK3CG, FCGR3A, FCGR1A, FCGR2A, FCGR2B, PIK3R5, RAC2, SYK, PLD1, PLA2G4F, PLA2G4D, PLA2G4E, ASAP3, PRKCB, PLCG2, LAT, INPP5D, VAV1, HCK, WAS, AMPH, SCIN, DOCK2, LIMK2, LIMK1
Rap1 signalling pathway(K)	0.0207	211	51	3.30E-07	4.63E-06	SKAP1, FPR1, PDGFB, GRIN2A, NGFR, EGFR, ITGAM, ITGB3, ITGB2, PIK3CD, ITGAL, PIK3CG, ADORA2A, LCP2, FYB, RASSF5, LPAR1, PIK3R5, RRAS, IGF1, LPAR5, ID1, RAC2, VEGFC, GRIN1, KITLG, PDGFRA, RASGRP2, RASGRP3, FGF14, FGF11, MAPK13, TEK, PRKCB, APBB1IP, ADCY4, ADCY2, ADCY1, ADCY7, CSF1, FGF1, FGF7, CSF1R, LAT, CALML5, CALML3, TIAM1, F2RL3, HGF, FGFR2, THBS1
Downstream signalling in naïve CD8+ T cells(N)	0.0063	64	24	4.01E-07	5.21E-06	STAT4, TNFRSF4, TNFRSF9, FASLG, IL2RG, GZMB, IL2RA, IL2RB, PTPN7, CD247, PRF1, CD3G, CD3E, CD3D, EOMES, CD8B, CD8A, JUNB, PRKCB, PRKCQ, FOSL1, TNF, B2M, IFNG
African trypanosomiasis(K)	0.0033	34	17	4.46E-07	5.80E-06	IDO2, IDO1, FASLG, ICAM1, IL12B, IL12A, IL10, IL18, IL1B, VCAM1, IL6, PRKCB, TNF, TLR9, F2RL1, SELE, IFNG
Phagosome(K)	0.015	153	40	9.48E-07	1.23E-05	C3, NCF1, NCF2, NCF4, TUBA4A, PLA2R1, CORO1A, ITGAM, ITGB3, ITGB2, FCGR3A, FCGR3B, MBL2, ITGA5, C1R, FCGR1A, FCGR2A, FCGR2B, CD14, MSR1, CD209, CTSS, CYBB, CYBA, MRC2, MRC1, TUBB3, TAP2, TAP1, CLEC7A, TCIRG1, ATP6V0D2, ATP6V0A4, MARCO, TLR6, TLR4, TLR2, COMP, THBS2, THBS1
Inflammatory mediator regulation of TRP channels(K)	0.0096	98	30	1.03E-06	1.34E-05	PIK3CD, PIK3CG, NTRK1, PIK3R5, IGF1, IL1B, PLA2G4F, PLA2G4D, PLA2G4E, MAPK13, PRKCH, PRKCB, PRKCQ, PLCG2, ADCY4, ADCY2, ADCY1, ADCY7, P2RY2, CALML5, CALML3, IL1R1, PTGER2, F2RL1, HRH1, TRPV2, TRPV4, BDKRB2, BDKRB1, CAMK2A

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
Calcineurin-regulated NFAT-dependent transcription in lymphocytes(N)	0.0045	46	19	1.60E-06	2.08E-05	GATA3, CTLA4, PTGS2, FOXP3, FASLG, IL2RA, IKZF1, EGR2, EGR3, TBX21, JUNB, PRKCQ, FOSL1, BATF3, CSF2, DGKA, TNF, CD40LG, IFNG
Influenza A(K)	0.0171	175	43	1.81E-06	2.17E-05	IKBKE, STAT1, SOCS3, TNFRSF10C, TNFRSF10A, TNFRSF10D, CXCL10, CIITA, PIK3CD, FASLG, PIK3CG, CCL5, CCL2, ICAM1, TNFSF10, IL12B, IL12A, RSAD2, PIK3R5, IL18, IL1A, IL1B, IL33, IL6, IRF7, IRF9, MX1, MAPK13, PRSS3, PRKCB, IFIH1, TNF, PYCARD, TLR7, TLR4, NLRP3, HSPA6, NFKBIA, CASP1, OAS1, IFNG, OAS2, OAS3
TNFR2 non-canonical NF-kB pathway(R)	0.0094	96	29	1.99E-06	2.39E-05	PSME2, PSMB8, PSMB9, TNFRSF4, TNFRSF11B, TNFRSF11A, TNFRSF17, NFKB2, FASLG, TNFSF13B, TNFRSF1B, BIRC3, TNFRSF6B, TNFSF11, TNFSF14, TNFSF13, CD27, CD40, TNFRSF13B, TNFRSF12A, RELB, CD70, PSMB10, TNFSF4, TNF, EDAR, CD40LG, LTA, LTB
Signalling by SCF-KIT(R)	0.0259	265	57	2.51E-06	3.01E-05	PSME2, PDGFB, GRIN2A, LYN, GRIN2D, JAK3, DUSP5, EGFR, STAT1, PSMB8, PSMB9, SOCS1, RASGRF2, RASGRF1, PIK3CD, NRG1, IL2RG, SH2B3, SH2B2, IL2RA, IL2RB, PTPN6, EREG, GRB7, CD19, CD28, GRAP, PTPRU, GRIN1, KITLG, CD86, CD80, PDGFRA, RASGRP1, RASGRP4, RASGRP3, RASGEF1A, MMP9, PSMB10, TEC, TEK, CSF2, FGF1, FGF7, GRAP2, RASAL1, RASAL3, LAT, LCK, IL3RA, VAV1, CSF2RB, CSF2RA, FGFR2, RASA4, CAMK2A, TRAT1
Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell(R)	0.0242	247	54	2.91E-06	3.39E-05	CXADR, C3, KLRK1, HCST, KIR3DL1, KIR3DL2, ITGB2, ITGAL, FCGR3A, ITGB7, ITGA4, FCGR1A, CD247, FCGR2B, ICAM2, ICAM3, ICAM4, ICAM1, CD226, CD19, SLAMF6, CD40, CD3G, CD3E, CD3D, CD33, NCR3, MICB, OSCAR, TYROBP, CD8B, CD8A, KIR2DL3, KIR2DL4, VCAM1, CD200R1, CD300LB, TREM2, TREM1, IFITM1, B2M, SH2D1A, SH2D1B, RAET1E, LILRA1, LILRB1, LILRB2, CD40LG, CD300E, KLRB1, KLRC1, SELL, KLRD1, AMICA1

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
Antigen processing and presentation(K)	0.0075	77	25	2.94E-06	3.39E-05	PSME2, LGMN, KIR3DL1, KIR3DL2, CIITA, IFI30, CTSS, CTSB, CD74, CD8B, CD8A, KIR2DL3, KIR2DL4, TAPBP, CD4, KIR2DS4, TNF, TAP2, TAP1, B2M, KLRC1, KLRC2, HSPA6, KLRD1, IFNG
amb2 Integrin signalling(N)	0.003	31	15	3.08E-06	3.39E-05	THY1, JAM2, CTGF, ITGAM, ITGB2, ICAM1, MST1R, SELPLG, PLAU, IL6, MMP2, MMP9, TNF, HCK, SELP
Interleukin signalling pathway(P)	0.0054	55	20	5.62E-06	6.18E-05	SPI1, SLA2, SPIB, IL20RA, JAK3, IL12RB1, IL12RB2, IL2RA, IL2RB, IL15, IL18, IL16, IL1A, IL4R, IL7, IL10RA, IL3RA, RPS6KA1, IL23A, CXCR1
Transcriptional misregulation in cancer(K)	0.0175	179	42	7.15E-06	7.86E-05	SPI1, SPINT1, NGFR, PBX3, FLI1, GRIA3, ITGAM, FLT3, ITGB7, PTCRA, BIRC3, FCGR1A, NTRK1, GZMB, IL2RB, CD14, IGF1, CD40, CCR7, CD86, CCND2, PLAU, IL6, LYL1, MMP3, MMP9, BCL2A1, PAX5, PAX8, CSF2, CSF1R, TRAF1, CEBPA, CEBPB, CEBPE, IL1R2, RUNX2, RUNX1, LMO2, MYCN, NR4A3, WNT16

Table T3-8: Full list of pathways enriched for genes negatively correlated (at FDR<0.05) with tumour *VDR* expression in the TCGA metastatic melanomas

The source database for each of the enriched pathways are indicated by a letter in parentheses after each pathway gene set name. The source database annotations are: C - CellMap, R – Reactome, K – KEGG, N – NCI PID, P - Panther, and B – BioCarta.

The description of column headers are: *Ratio of protein in gene set*: ratios of numbers of genes contained in pathways to total genes in the Reactome FI network; Number of protein in gene set: numbers of genes in pathways; *Protein from network*: numbers of hit genes from the query gene list; *Nodes:* nodal genes from my input query gene list, which 'matches' with the genes in a particular pathway; P-value was estimated by Reactome FIViz using hypergeometric test and the corresponding FDR was estimated by Benjamini Hochberg multiple correction

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
Wnt signalling pathway(P)	0.0262	268	12	3.43E-06	1.11E-03	CDH8, TLE4, PRKCA, PCDHA11, PCDHB5, PCDHB3, CDH10, GNG7, CTNNA2, TCF7L1, PCDH9, PCDH7
Metabolism of carbohydrates(R)	0.0234	239	9	2.27E-04	0.0192	HS6ST2, BCAN, CSPG5, SLC2A4, CHST9, GPC3, GPC2, GPC4, HS3ST5
Cadherin signalling pathway(P)	0.0098	100	6	2.41E-04	0.0192	PCDHA11, PCDHB5, PCDHB3, CTNNA2, PCDH9, PCDH7
Ephrin A reverse signalling(N)	0.0003	3	2	3.25E-04	0.0192	EPHA5, EFNA5
cAMP signalling pathway(K)	0.0195	199	8	3.34E-04	0.0192	CHRM1, ATP1B2, GRIA1, GRIA2, GRIA4, FXYD1, ATP1A2, ATP2B3
Heterotrimeric G-protein signalling pathway-Gq alpha and Go alpha mediated pathway(P)	0.0106	108	6	3.61E-04	0.0192	CHRM1, PRKCA, GPSM1, GRM7, GARNL3, GNG7
Glutamatergic synapse(K)	0.0112	114	6	4.80E-04	0.0221	PRKCA, GRIA1, GRIA2, GRIA4, GRM7, GNG7
Signalling by NODAL(R)	0.0019	19	3	6.28E-04	0.0229	GDF1, LEFTY1, ACVR2B
TGF-beta signalling pathway(P)	0.0078	80	5	6.74E-04	0.0229	GDF1, MSTN, LEFTY1, GDF11, ACVR2B
PI3K-Akt signalling pathway(K)	0.0338	345	10	7.89E-04	0.0229	CHRM1, IGF1R, PRKCA, EFNA5, PDGFD, FGF17, SGK3, FIGF, COL11A2, GNG7
Ras signalling pathway(K)	0.0222	227	8	7.89E-04	0.0229	IGF1R, PRKCA, SHC2, EFNA5, PDGFD, FGF17, FIGF, GNG7
Dopaminergic synapse(K)	0.0126	129	6	9.09E-04	0.0236	SCN1A, PRKCA, GRIA1, GRIA2, GRIA4, GNG7

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
Proximal tubule bicarbonate reclamation(K)	0.0023	23	3	1.09E-03	0.0261	ATP1B2, SLC9A3, ATP1A2
Circadian entrainment(K)	0.0093	95	5	1.44E-03	0.0331	PRKCA, GRIA1, GRIA2, GRIA4, GNG7
Long-term depression(K)	0.0059	60	4	1.88E-03	0.0376	IGF1R, PRKCA, GRIA1, GRIA2
Retrograde endocannabinoid signalling(K)	0.0099	101	5	1.88E-03	0.0376	PRKCA, GRIA1, GRIA2, GRIA4, GNG7
EPHA forward signalling(N)	0.003	31	3	2.53E-03	0.0456	EPHA5, EPHA6, EFNA5
Amphetamine addiction(K)	0.0066	67	4	2.79E-03	0.0474	PRKCA, GRIA1, GRIA2, GRIA4

Table T3-9: Scoring of LMC primary melanoma sections stained with anti-VDR antibody.

Slide number	Number of cytoplasmic VDR +ve tumour cells	Number of nuclear VDR +ve tumour cells	Number of VDR +ve TILs
1	2	3	none
2	0	36	none
3	2	12	none
4	1	0	none
5	1	0	none
6	2	0	none
7	1	0	none
8	2	8	none
9	3	12	none
10	3	12	none
11	0	0	none
12	3	16	some
13	2	1	none
14	3	36	some
15	0	1	some
16	2	8	none
17	3	25	lots
18	2	9	none
19	3	20	some
20	2	0	none
21	1	1	none
22	1	0	none
23	1	169	lots
24	1	9	none
25	1	0	none
26	1	10	none
27	2	0	none
28	2	2	none
29	2	40	some
30	0	0	none

A.2 Primer sequences used for Sanger Sequencing of VDR plasmid

Table T4-1: Table of primers used to check sequence of the cloned VDR plasmid, using Sanger Sequencing.

Three rounds of sequencing checks were done to span various regions of the plasmid. Tm- the melting temperature o primer sequence, Nts- number of nucleotides

Primer#	Tm	Nts	Primer Sequence
Round 1			
Primer 1	52	29	AATTGTAAGCGTTAATATTTTGTTAAAAT
Primer 2	50	18	GACAGCAGGCTGAATAAT
Primer 3	52	22	AAAATGATGTCATGGCTTTAGA
Primer 4	61	16	GCGTTGCCTTCGCCCC
Primer 5	53	19	TCCTTTGTCCCAAATCTGG
Primer 6	52	17	TGTCTGAGGAGCAACAG
Primer 7	60	14	GCTGCCGCCACCCG
Primer 8	62	20	CCACGGGGACGTGGTTTTCC
Primer 9	60	22	AGTTGCCAGCCATCTGTTGTTT
Primer 10	51	29	AACTACCCATTTTATTATATATTAGTCAC
Primer 11	62	21	TAATGAATCGGCCAACGCGCG
Primer 12	52	20	AAAAGAGTTGGTAGCTCTTG
Primer 13	52	18	GTCAGAAGTAAGTTGGCC
Round 2			
Primer 14	51	25	TCGACATTGATTATTGACTAGTTAT
Primer 15	58	24	GAAAGTTTCCTTTTATGGCGAGGC
Primer 16	58	21	CGCAGCCATTGCCTTTTATGG
Primer 17	61	20	CAAGGCCCTGTTCACCTGCC
Primer 18	58	23	TACAGCATCCAAAAGGTCATCGG
Primer 19	58	21	CCCTTGTGCTAGAGGTGTTCG
Round 3			
Primer 20	61	20	GCGGAGCCGAAATCTGGGAG
Primer 21	60	27	CTTCTTCTTTTTCCTACAGCTCCTGGG

A.3 Tables of output from enrichment analyses performed in Chapter 5

In the case of tables containing output from enrichment analyses, pathways are presented based on variable FDR thresholds. The reason for this being, repetitive pathways appear at lower thresholds and hence the necessity to drop the display of these pathways.

Table T5-1: Full list of pathways (at FDR<0.00001) enriched for genes that are expressed significantly higher in ulcerated tumours compared to non-ulcerated tumours). Output from Reactome FIViz

The source database for each of the enriched pathways are indicated by a letter in parentheses after each pathway gene set name. The source database annotations are: C - CellMap, R – Reactome, K – KEGG, N – NCI PID, P - Panther, and B – BioCarta.

The description of column headers are: *Ratio of protein in gene set*: ratios of numbers of genes contained in pathways to total genes in the Reactome FI network; *Number of protein in gene set:* numbers of genes in pathways; *Protein from network*: numbers of hit genes from the query gene list; *Nodes:* nodal genes from my input query gene list, which 'matches' with the genes in a particular pathway; P-value was estimated by Reactome FIViz using hypergeometric test and the corresponding FDR was estimated by Benjamini Hochberg multiple correction

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
Mitotic Prometaphase(R)	0.0097	99	43	1.11E-16	4.89E-14	CDCA5, NCAPG, CDCA8, CENPA, CENPE, CENPF, APITD1, CENPI, CENPK, CENPL, CENPM, CENPO, CENPQ, CASC5, NDC80, BIRC5, CDC20, DSN1, KIF2C, SPC24, MAD2L1, ERCC6L, BUB1B, CCNB2, CCNB1, SGOL2, SGOL1, AURKB, SMC4, NUP98, SKA1, SKA2, PLK1, NUF2, MAD1L1, HDAC8, ZWINT, AHCTF1, ZWILCH, KNTC1, KIF18A, ITGB3BP, BUB1
Mitotic Metaphase and Anaphase(R)	0.016	163	56	1.11E-16	4.89E-14	PSMD4, PSMD3, CDCA5, CDCA8, PSMA5, CENPA, PSMB7, PSMB4, PSMB5, PSMB2, CENPE, CENPF, PSMC6, APITD1, CENPI, CENPK, CENPL, CENPM, CENPO, CENPQ, CASC5, NDC80, BIRC5, CDC20, FBXO5, DSN1, KIF2C, SPC24, MAD2L1, ERCC6L, BUB1B, PSMD12, PSMD11, PSMD14, PSMD13, UBE2C, SGOL2, SGOL1, AURKB, NUP98, SKA1, SKA2, PLK1, NUF2, VRK1, MAD1L1, ESPL1, HDAC8, ZWINT, AHCTF1, ZWILCH, PTTG1, KNTC1, KIF18A, ITGB3BP, BUB1
Signalling by Rho GTPases(R)	0.0316	323	81	2.22E-16	6.51E-14	CDCA8, IQGAP3, CENPA, FLNA, CENPE, CENPF, APITD1, CENPI, CENPK, CENPL, CENPM, CENPO, CENPQ, FAM13B, CASC5, CDC25C, NDC80, ARHGAP11A, ARHGAP11B, BIRC5, PIK3C3, CDC42, PRC1, CDC20, HIST2H3A, ROPN1, DSN1, RACGAP1, A2M, PTK2, GRB2, KIF14, TAX1BP3, KIF2C, ARHGEF2, ARHGEF7, PAK2, SPC24, MAD2L1, ERCC6L, BUB1B, RHPN2, ECT2, SGOL2, SGOL1, YWHAG, ARHGEF11, ARHGEF17, AURKB, TRIO, WIPF2, WIPF3, DEPDC1B, NUP98, SKA1, SKA2, STARD13, PLK1, ARPC1B, NUF2, CIT, FGD1, FGD4, MAD1L1, RHOBTB1, SRGAP2, NF2, HIST1H3A, ZWINT, HIST1H4A, AHCTF1, CTNNB1, ZWILCH, KNTC1, ARHGAP26, KIF18A, DIAPH3, ARAP3, ITGB3BP, BUB1, LIMK1
Mitotic G1-G1/S phases(R)	0.0123	126	43	1.97E-13	4.13E-11	PSMD4, PSMD3, DHFR, LIN52, PSMA5, PSMB7, PSMB4, PSMB5, PSMB2, PSMC6, CDC25A, PKMYT1, MNAT1, MCM10, DBF4, E2F1, E2F3, E2F5, FBXO5, POLA1, PCNA, RPA3, CCNE2, CCNE1, PSMD12, PSMD11, PSMD14, PSMD13, CCNB1, RRM2, POLE2, CDC7, CDC6, CDK2, TFDP2, MCM3, MCM4, SKP2, CDT1, TOP2A, PRIM1, TYMS, LIN9

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
Cell Cycle Checkpoints(R)	0.0146	149	47	2.35E-13	4.13E-11	MDC1, PSMD4, PSMD3, PSMA5, PSMB7, PSMB4, PSMB5, PSMB2, PSMC6, PIAS4, CDC25C, CDC25A, PKMYT1, MCM10, SUMO1, DBF4, CDC20, BRCA1, RPA3, RAD1, MAD2L1, BUB1B, UBE2V2, PSMD12, PSMD11, PSMD14, PSMD13, CCNB2, CCNB1, UBE2C, YWHAG, RNF8, CDC7, CDC6, NBN, MAD1L1, CDK2, RFC3, RFC4, MCM3, MCM4, CHEK2, CHEK1, HIST1H4A, ATR, CLSPN, H2AFX
Synthesis of DNA(R)	0.0094	96	34	2.51E-11	3.67E-09	PSMD4, PSMD3, PSMA5, PSMB7, PSMB4, PSMB5, PSMB2, PSMC6, POLA1, PCNA, RPA3, PSMD12, PSMD11, PSMD14, PSMD13, POLD3, POLD4, POLD1, POLE2, CDC6, CDK2, RFC3, RFC4, RFC1, MCM3, MCM4, DNA2, CDT1, FEN1, PRIM1, GINS1, GINS3, GINS4, LIG1
S Phase(R)	0.0117	120	38	4.18E-11	5.22E-09	PSMD4, PSMD3, CDCA5, PSMA5, PSMB7, PSMB4, PSMB5, PSMB2, PSMC6, CDC25A, MNAT1, POLA1, PCNA, RPA3, PSMD12, PSMD11, PSMD14, PSMD13, POLD3, POLD4, POLD1, POLE2, CDC6, CDK2, RFC3, RFC4, RFC1, MCM3, MCM4, SKP2, DNA2, CDT1, FEN1, PRIM1, GINS1, GINS3, GINS4, LIG1
Mitochondrial translation(R)	0.0087	89	31	2.80E-10	3.08E-08	MRPL19, MRPL16, MRPL12, MRPL13, MRPL27, MRPL28, MRPL24, MRPL21, MRPL39, MRPL35, MRPL32, MRPL33, MRPL42, MRPL53, MRPL50, MRPL51, MRPL3, MRPS17, MRPS28, MRPS23, MRPS7, MRPS5, MRPS33, MRPS34, MRPS30, MRPS18A, GFM2, MRRF, CHCHD1, TUFM, DAP3
Cell cycle(K)	0.0121	124	37	3.76E-10	3.52E-08	CDC25C, CDC25A, CDC25B, PKMYT1, ZBTB17, DBF4, E2F1, E2F3, E2F5, CDC20, PCNA, MAD2L2, MAD2L1, BUB1B, CCNE2, CCNE1, CCNB2, CCNB1, YWHAG, CCNA2, PLK1, CDC7, CDC6, MAD1L1, CDK2, ESPL1, TFDP2, MCM3, MCM4, TTK, CHEK2, CHEK1, SKP2, ATR, PTTG1, SMAD4, BUB1
HDR through Homologous Recombination (HR)	0.0083	85	30	4.00E-10	3.52E-08	MDC1, PPP4R2, PIAS4, BLM, SUMO1, EXO1, BRIP1, XRCC3, BRCA1, BRCA2, RPA3, RAD1, UBE2V2, RNF8, ERCC4, NBN, RAD51AP1, CDK2, RFC3, RFC4, CHEK1, DNA2, HIST1H4A, ATR, RAD51C, EME1, TIMELESS, CLSPN, TOPBP1, H2AFX

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
or Single Strand Annealing (SSA)(R)						
PLK1 signalling events(N)	0.0043	44	21	1.04E-09	8.30E-08	CENPE, CDC25C, CDC25B, ODF2, NDC80, PRC1, CDC20, FBXO5, KIF20A, SPC24, ERCC6L, BUB1B, CCNB1, ECT2, SGOL1, AURKA, PLK1, TPX2, TUBG1, CLSPN, BUB1
RNA Polymerase I, RNA Polymerase III, and Mitochondrial Transcription(R)	0.0087	89	30	1.14E-09	8.33E-08	EHMT2, POLRMT, MNAT1, ZNRD1, SSB, HIST2H3A, TFB2M, TTF1, RRN3, TAF1A, CD3EAP, GTF2H1, GTF2H3, GTF2H4, UBTF, POLR1A, POLR1C, POLR1E, RBBP7, POLR2K, POLR3A, POLR3C, POLR3D, POLR3G, POLR3K, HIST1H3A, HIST1H4A, TWISTNB, CBX3, SNAPC5
DNA replication(K)	0.0035	36	19	1.29E-09	8.66E-08	RNASEH2C, RNASEH2A, POLA1, PCNA, RPA3, POLD3, POLD4, POLD1, POLE4, POLE2, RFC3, RFC4, RFC1, MCM3, MCM4, DNA2, FEN1, PRIM1, LIG1
Nucleosome assembly(R)	0.0028	29	17	2.06E-09	1.27E-07	CENPA, APITD1, CENPI, CENPK, CENPL, CENPM, CENPO, CENPQ, CASC5, SMARCA5, NPM1, HJURP, OIP5, RBBP7, HIST1H4A, RSF1, ITGB3BP
Mitotic G2-G2/M phases(R)	0.0109	111	33	3.63E-09	2.10E-07	LIN52, CENPF, CENPJ, NEK2, CDC25C, CDC25A, CDC25B, PKMYT1, ODF2, MNAT1, FOXM1, E2F1, E2F3, CEP192, CCNB2, CCNB1, YWHAG, CCNA2, DCTN1, AURKA, CEP70, CEP76, CEP78, PLK4, PLK1, CEP63, ALMS1, CDK2, CEP152, TUBB, TUBG1, SDCCAG8, LIN9
Nucleotide Excision Repair(R)	0.01	102	31	6.54E-09	3.52E-07	ACTL6A, INO80E, MNAT1, ISY1, SUMO1, ACTR8, PCNA, RPA3, UBE2V2, POLD3, POLD4, POLD1, ERCC4, ERCC8, RNF111, GTF2H1, GTF2H3, GTF2H4, COPS5, RFC3, RFC4, RFC1, POLR2B, POLR2C, POLR2G, POLR2K, ZNF830, PRPF19, PARP1, LIG1, LIG3
Aurora B signalling(N)	0.0039	40	19	6.89E-09	3.52E-07	NCAPG, CDCA8, CENPA, NDC80, BIRC5, KIF20A, RACGAP1, KIF23, KIF2C, NPM1, KLHL9, SGOL1, AURKB, AURKA, SMC4, STMN1, NCL, PPP2R5D, BUB1

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
Validated targets of C-MYC transcriptional activation(N)	0.0073	75	26	8.28E-09	3.98E-07	ACTL6A, ODC1, RUVBL2, CDC25A, BIRC5, E2F3, EIF4E, RCC1, NPM1, SUPT7L, CCNB1, HMGA1, CAD, TAF9, TAF12, PRDX3, MINA, TK1, NBN, NCL, UBTF, POLR3D, NME2, NME1, SMAD4, HSPA4
M/G1 Transition(R)	0.008	82	27	1.23E-08	5.66E-07	PSMD4, PSMD3, PSMA5, PSMB7, PSMB4, PSMB5, PSMB2, PSMC6, GMNN, MCM10, DBF4, E2F1, E2F3, POLA1, RPA3, PSMD12, PSMD11, PSMD14, PSMD13, POLE2, CDC7, CDC6, CDK2, MCM3, MCM4, CDT1, PRIM1
Fanconi anemia pathway(N)	0.0044	45	19	4.29E-08	1.89E-06	APITD1, BLM, BRIP1, XRCC3, BRCA1, BRCA2, RAD1, UBE2T, NBN, FANCI, FANCC, FANCB, RFC3, RFC4, CHEK1, ATR, FANCD2, TOPBP1, H2AFX
ATR signalling pathway(N)	0.0036	37	17	6.74E-08	2.76E-06	CDC25C, CDC25A, BRCA2, RAD1, CCNA2, PLK1, CDC6, NBN, CDK2, RFC3, RFC4, CHEK1, ATR, FANCD2, TIMELESS, CLSPN, TOPBP1
NoRC negatively regulates rRNA expression(R)	0.0048	49	19	1.56E-07	6.24E-06	SAP30BP, MNAT1, ZNRD1, HIST2H3A, SMARCA5, TTF1, TAF1A, CD3EAP, GTF2H1, GTF2H3, GTF2H4, UBTF, POLR1A, POLR1C, POLR1E, POLR2K, HIST1H3A, HIST1H4A, TWISTNB
APC/C-mediated degradation of cell cycle proteins(R)	0.008	82	25	1.75E-07	6.66E-06	PSMD4, PSMD3, PSMA5, PSMB7, PSMB4, PSMB5, PSMB2, PSMC6, NEK2, CDC20, FBXO5, MAD2L1, BUB1B, PSMD12, PSMD11, PSMD14, PSMD13, CCNB1, UBE2C, AURKB, AURKA, PLK1, CDK2, SKP2, PTTG1

Table T5-2: Full list of pathways (at FDR<0.0001) enriched for genes that are expressed significantly lower in ulcerated tumours compared to non-ulcerated tumours). Output from Reactome FIViz

The source database for each of the enriched pathways are indicated by a letter in parentheses after each pathway gene set name. The source database annotations are: C - CellMap, R – Reactome, K – KEGG, N – NCI PID, P - Panther, and B – BioCarta.

The description of column headers are: *Ratio of protein in gene set*: ratios of numbers of genes contained in pathways to total genes in the Reactome FI network; *Number of protein in gene set:* numbers of genes in pathways; *Protein from network*: numbers of hit genes from the query gene list; *Nodes:* nodal genes from my input query gene list, which 'matches' with the genes in a particular pathway; P-value was estimated by Reactome FIViz using hypergeometric test and the corresponding FDR was estimated by Benjamini Hochberg multiple correction

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
Extracellular matrix organization(R)	0.0243	248	49	1.45E-11	1.17E-08	ELANE, ELN, MMP11, MMP19, JAM3, COL3A1, FMOD, EFEMP2, EFEMP1, CDH1, COL17A1, ITGB4, ITGB2, ITGAL, ITGB6, ITGA8, LRP4, MFAP4, SPARC, COL8A2, KLK7, SDC1, COL9A2, CTSG, LOXL1, COMP, CASK, COL4A4, COL4A6, COL4A5, LTBP4, FBLN2, FBLN5, VCAM1, TGFB3, BMP7, BMP4, BMP2, LAMC2, FGA, DCN, COL6A2, COL6A1, COL6A3, LAMA5, LAMA2, LAMA3, LAMB3, COL7A1
Pathways in cancer(K)	0.0389	397	61	9.67E-10	3.70E-07	SPI1, DAPK2, AR, WNT10B, WNT10A, JUP, CDKN1A, ADCY7, EGFR, EGLN3, CDKN2B, WNT7B, WNT7A, CDH1, ZBTB16, FLT3, FASLG, PIK3CG, PLEKHG5, BIRC3, CEBPA, CBLC, RASSF5, FLT3LG, LPAR1, TGFA, RALGDS, LPAR5, NCOA4, KIT, PTGER2, PTGER3, GNAI1, RXRA, FAS, PDGFRB, COL4A4, COL4A6, COL4A5, RASGRP1, WNT3, FGF22, WNT4, WNT3A, TGFB3, PTCH1, BMP4, NFKBIA, BMP2, FGFR3, FGF11, FGFR2, LAMC2, AXIN2, LAMA5, LAMA2, LAMA3, WNT11, ERBB2, LAMB3, LAMB4
ECM-receptor interaction(K)	0.0085	87	25	1.38E-09	3.70E-07	GP6, COL3A1, TNN, ITGB4, ITGB6, ITGA8, SDC1, CD36, TNXB, COMP, COL4A4, COL4A6, COL4A5, CHAD, LAMC2, COL6A2, COL6A1, COL6A3, COL6A6, LAMA5, LAMA2, LAMA3, THBS2, LAMB3, LAMB4
Cytokine-cytokine receptor interaction(K)	0.0259	265	45	9.87E-09	1.99E-06	IL20RA, IL20RB, IL22RA1, IFNGR1, CCL13, EGFR, CCL19, CCL21, IL10RA, CCL27, EPO, CXCL14, CXCL16, CXCL10, CXCL11, FLT3, FASLG, TNFSF13B, CX3CR1, IL2RG, CCL5, TNFSF10, IL15RA, FLT3LG, XCR1, CX3CL1, CCR7, IL15, IL18, OSMR, KIT, CXCR1, CXCR3, FAS, IL18R1, IL18RAP, PDGFRB, LIFR, TGFB3, BMP7, BMP2, CXCL9, GHR, LTB, ACVR2A
Interferon gamma signalling(R)	0.0072	74	21	3.48E-08	5.60E-06	HLA-DQA2, HLA-DQA1, JAK2, HLA-DPA1, IFNGR1, HLA-DQB2, HLA-DPB1, PIAS1, HLA-DRA, PTPN6, B2M, TRIM22, HLA-B, HLA-E, VCAM1, OAS2, OAS3, IRF8, IRF6, IRF9, CAMK2G
Beta1 integrin cell surface interactions(N)	0.0065	66	19	1.25E-07	1.67E-05	COL3A1, F13A1, ITGA8, PLAU, COL4A4, COL4A6, COL4A5, VCAM1, LAMC2, FGA, COL6A2, COL6A1, COL6A3, LAMA5, LAMA2, LAMA3, THBS2, LAMB3, COL7A1

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
HTLV-I infection(K)	0.0253	258	41	2.54E-07	2.92E-05	SPI1, MYB, HLA-DQA2, HLA-DQA1, WNT10B, WNT10A, NFATC2, CDKN1A, TSPO, HLA-DPA1, ADCY7, HLA-DMB, CDKN2B, WNT7B, WNT7A, HLA-DPB1, HLA-DOB, HLA-DRA, ITGB2, ITGAL, PIK3CG, LCK, IL2RG, IL15RA, HLA-B, HLA-E, CD3G, CD3E, CD3D, PPP3CB, IL15, ETS2, CCND2, PDGFRB, WNT3, WNT4, VCAM1, WNT3A, TGFB3, NFKBIA, WNT11
Cell adhesion molecules (CAMs)(K)	0.0139	142	28	3.73E-07	3.69E-05	CLDN1, HLA-DQA2, HLA-DQA1, HLA-DPA1, HLA-DMB, JAM3, HLA-DPB1, SIGLEC1, CDH1, HLA-DOB, PTPRC, HLA-DRA, ITGB2, ITGAL, SPN, NEO1, ITGA8, PVRL1, HLA-B, HLA-E, OCLN, SDC1, PTPRF, CLDN23, CLDN16, CD8A, VCAM1, CNTN1
PI3K-Akt signalling pathway(K)	0.0338	345	49	4.57E-07	3.69E-05	MYB, CDKN1A, JAK2, EGFR, COL3A1, EPO, TNN, RPS6, FOXO3, ITGB4, FASLG, PIK3CG, ITGB6, ITGA8, IL2RG, EPHA2, LPAR1, LPAR5, OSMR, KIT, SGK2, TNXB, RXRA, COMP, CCND2, PDGFRB, COL4A4, COL4A6, COL4A5, EFNA5, CHAD, FGF22, EFNA3, FGFR3, FGF11, FGFR2, LAMC2, GHR, COL6A2, COL6A1, COL6A3, COL6A6, LAMA5, LAMA2, LAMA3, THBS2, LAMB3, LAMB4, PPP2R2C
Inflammatory bowel disease (IBD)(K)	0.0064	65	18	4.61E-07	3.69E-05	GATA3, HLA-DQA2, HLA-DQA1, HLA-DPA1, IFNGR1, HLA-DMB, HLA-DPB1, RORC, RORA, HLA-DOB, MAF, HLA-DRA, NOD2, IL2RG, IL18, IL18R1, IL18RAP, TGFB3
Axon guidance(K)	0.0124	127	25	1.60E-06	1.17E-04	DPYSL2, UNC5B, NGEF, NFATC2, PLXNC1, PLXNA3, PLXNB1, SEMA3D, SEMA4A, SEMA4D, SEMA4B, EPHB6, EPHB1, EPHA4, EPHA1, EPHA2, ROBO2, PPP3CB, RHOD, GNAI1, PAK6, EFNA5, EFNB1, EFNA3, SLIT3
Chemical carcinogenesis(K)	0.0079	81	19	2.50E-06	1.67E-04	CYP2E1, ADH7, GSTO2, CYP2C19, CYP2C18, UGT1A1, CYP3A5, UGT1A5, UGT1A3, UGT1A9, UGT1A7, UGT1A6, MGST1, MGST2, AKR1C2, GSTA4, GSTA3, ALDH3A1, ALDH3B2

GeneSet	Ratio Of Protein In GeneSet	No. Of Protein In GeneSet	Protein From Network	P-value	FDR	Nodes
Metabolism of xenobiotics by cytochrome P450(K)	0.0072	74	18	2.79E-06	1.73E-04	CYP2E1, ADH7, GSTO2, UGT1A1, CYP3A5, UGT1A5, UGT1A3, UGT1A9, UGT1A7, UGT1A6, MGST1, MGST2, AKR1C1, AKR1C2, GSTA4, GSTA3, ALDH3A1, ALDH3B2
Drug metabolism - cytochrome P450(K)	0.0067	68	17	3.67E-06	2.06E-04	CYP2E1, ADH7, GSTO2, CYP2C19, UGT1A1, CYP3A5, UGT1A5, UGT1A3, UGT1A9, UGT1A7, UGT1A6, MGST1, MGST2, GSTA4, GSTA3, ALDH3A1, ALDH3B2
IL12-mediated signalling events(N)	0.006	61	16	3.90E-06	2.06E-04	JAK2, HLA-DRA, FASLG, LCK, IL2RG, CD247, B2M, GADD45B, GADD45G, CD3G, CD3E, CD3D, IL18, IL18R1, IL18RAP, CD8A
T cell activation(P)	0.0079	81	18	9.37E-06	4.69E-04	HLA-DQA2, HLA-DQA1, NFATC2, HLA-DPA1, HLA-DMB, PTPRC, HLA-DRA, PIK3CG, LCK, CD247, B2M, CD3G, CD3E, CD3D, PPP3CB, VAV3, CD74, NFKBIA
Amoebiasis(K)	0.0106	108	21	1.28E-05	6.04E-04	SERPINB3, SERPINB4, SERPINB2, SERPINB13, COL3A1, ITGB2, PIK3CG, CTSG, RAB7B, COL4A4, COL4A6, COL4A5, TGFB3, HSPB1, LAMC2, LAMA5, LAMA2, LAMA3, GNA15, LAMB3, LAMB4
Toxoplasmosis(K)	0.0116	118	22	1.52E-05	6.69E-04	HLA-DQA2, HLA-DQA1, JAK2, HLA-DPA1, IFNGR1, HLA-DMB, IL10RA, HLA-DPB1, HLA-DOB, HLA-DRA, PIK3CG, BIRC3, GNAI1, TGFB3, NFKBIA, LAMC2, LAMA5, LAMA2, LAMA3, LAMB3, LAMB4, MAPK13
TCR signalling in naïve CD8+ T cells(N)	0.0053	54	14	1.75E-05	7.35E-04	MAP3K8, PTPRC, LCK, MALT1, PTPN6, CD247, RASSF5, B2M, CD3G, CD3E, CD3D, CD8A, RASGRP1, TRPV6
Staphylococcus aureus infection(K)	0.0054	55	14	2.13E-05	8.54E-04	CFD, CFH, C3, HLA-DQA2, HLA-DQA1, HLA-DPA1, HLA-DMB, HLA-DPB1, HLA-DOB, HLA-DRA, ITGB2, ITGAL, DSG1, C1S
Focal adhesion(K)	0.0203	207	31	2.27E-05	8.62E-04	EGFR, FLNB, COL3A1, TNN, ITGB4, PIK3CG, ITGB6, ITGA8, BIRC3, VAV3, TNXB, PAK6, COMP, CCND2, PDGFRB, COL4A4, COL4A6, COL4A5, CHAD, LAMC2, COL6A2, COL6A1, COL6A3, COL6A6, LAMA5, LAMA2, LAMA3, THBS2, ERBB2, LAMB3, LAMB4

References

1. Breslow A. Thickness, cross-sectional areas and depth of invasion in the prognosis of cutaneous melanoma. *Ann Surg.* 1970;172(5):902-8.
2. Edge SB, and Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol.* 2010;17(6):1471-4.
3. Kashani-Sabet M, Sagebiel RW, Ferreira CM, Nosrati M, and Miller JR, 3rd. Vascular involvement in the prognosis of primary cutaneous melanoma. *Arch Dermatol.* 2001;137(9):1169-73.
4. Shain AH, and Bastian BC. From melanocytes to melanomas. *Nature reviews Cancer.* 2016;16(6):345-58.
5. Bastian BC. The molecular pathology of melanoma: an integrated taxonomy of melanocytic neoplasia. *Annual review of pathology.* 2014;9(239-71.
6. UK CR. Cancer Research UK Melanoma skin cancer incidence statistics. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/melanoma-skin-cancer/incidence#heading-Zero. Accessed January, 2019.
7. Linos E, Swetter SM, Cockburn MG, Colditz GA, and Clarke CA. Increasing burden of melanoma in the United States. *J Invest Dermatol.* 2009;129(7):1666-74.
8. 2012 G. Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 http://globocan.iarc.fr.
9. Kosary CL, Altekruse SF, Ruhl J, Lee R, and Dickie L. Clinical and prognostic factors for melanoma of the skin using SEER registries: collaborative stage data collection system, version 1 and version 2. *Cancer.* 2014;120 Suppl 23(3807-14.
10. Karimkhani C, Green AC, Nijsten T, Weinstock MA, Dellavalle RP, Naghavi M, and Fitzmaurice C. The global burden of melanoma: results from the Global Burden of Disease Study 2015. *Br J Dermatol.* 2017;177(1):134-40.
11. Crombie IK. Variation of melanoma incidence with latitude in North America and Europe. *Br J Cancer.* 1979;40(5):774-81.
12. Baade P, Meng X, Youlden D, Aitken J, and Youl P. Time trends and latitudinal differences in melanoma thickness distribution in Australia, 1990-2006. *Int J Cancer.* 2012;130(1):170-8.

13. Natalie H. Matthews W-QL, Abrar A. Qureshi, Martin A. Weinstock, and Eunyoung Cho. In: Ward WH FJ ed. *Cutaneous Melanoma: Etiology and Therapy*. Brisbane: Codon Publications; 2017.
14. WHO. 2002.
15. Eide MJ, and Weinstock MA. Association of UV index, latitude, and melanoma incidence in nonwhite populations--US Surveillance, Epidemiology, and End Results (SEER) Program, 1992 to 2001. *Arch Dermatol.* 2005;141(4):477-81.
16. Levine H, Afek A, Shamiss A, Derazne E, Tzur D, Astman N, Keinan-Boker L, Mimouni D, and Kark JD. Country of origin, age at migration and risk of cutaneous melanoma: a migrant cohort study of 1,100,000 Israeli men. *Int J Cancer.* 2013;133(2):486-94.
17. Garbe C, and Leiter U. Melanoma epidemiology and trends. *Clin Dermatol.* 2009;27(1):3-9.
18. Weir HK, Marrett LD, Cokkinides V, Barnholtz-Sloan J, Patel P, Tai E, Jemal A, Li J, Kim J, and Ekwueme DU. Melanoma in adolescents and young adults (ages 15-39 years): United States, 1999-2006. *J Am Acad Dermatol.* 2011;65(5 Suppl 1):S38-49.
19. Watson M, Geller AC, Tucker MA, Guy GP, Jr., and Weinstock MA. Melanoma burden and recent trends among non-Hispanic whites aged 15-49years, United States. *Prev Med.* 2016;91(294-8.
20. Bulliard JL, and Cox B. Cutaneous malignant melanoma in New Zealand: trends by anatomical site, 1969-1993. *Int J Epidemiol.* 2000;29(3):416-23.
21. Surveillance E, and End Results (SEER). 2015.
22. Li WQ, Qureshi AA, Ma J, Goldstein AM, Giovannucci EL, Stampfer MJ, and Han J. Personal history of prostate cancer and increased risk of incident melanoma in the United States. *J Clin Oncol.* 2013;31(35):4394-9.
23. Li WQ, Cho E, Weinstock MA, Mashfiq H, and Qureshi AA. Epidemiological Assessments of Skin Outcomes in the Nurses' Health Studies. *Am J Public Health.* 2016;106(9):1677-83.
24. Downing A, Newton-Bishop JA, and Forman D. Recent trends in cutaneous malignant melanoma in the Yorkshire region of England; incidence, mortality and survival in relation to stage of disease, 1993-2003. *Br J Cancer.* 2006;95(1):91-5.
25. Whiteman DC, Green AC, and Olsen CM. The Growing Burden of Invasive Melanoma: Projections of Incidence Rates and Numbers of New Cases in Six Susceptible Populations through 2031. *J Invest Dermatol.* 2016;136(6):1161-71.

26. Chen L, and Jin S. Trends in mortality rates of cutaneous melanoma in East Asian populations. *PeerJ.* 2016;4(e2809.
27. Jemal A, Saraiya M, Patel P, Cherala SS, Barnholtz-Sloan J, Kim J, Wiggins CL, and Wingo PA. Recent trends in cutaneous melanoma incidence and death rates in the United States, 1992-2006. *J Am Acad Dermatol.* 2011;65(5 Suppl 1):S17-25 e1-3.
28. Ward-Peterson M, Acuna JM, Alkhalifah MK, Nasiri AM, Al-Akeel ES, Alkhaldi TM, Dawari SA, and Aldaham SA. Association Between Race/Ethnicity and Survival of Melanoma Patients in the United States Over 3 Decades: A Secondary Analysis of SEER Data. *Medicine (Baltimore).* 2016;95(17):e3315.
29. Wu XC, Eide MJ, King J, Saraiya M, Huang Y, Wiggins C, Barnholtz-Sloan JS, Martin N, Cokkinides V, Miller J, et al. Racial and ethnic variations in incidence and survival of cutaneous melanoma in the United States, 1999-2006. *J Am Acad Dermatol.* 2011;65(5 Suppl 1):S26-37.
30. Bradford PT, Goldstein AM, McMaster ML, and Tucker MA. Acral lentiginous melanoma: incidence and survival patterns in the United States, 1986-2005. *Arch Dermatol.* 2009;145(4):427-34.
31. Shen W, Sakamoto N, and Yang L. Melanoma-specific mortality and competing mortality in patients with non-metastatic malignant melanoma: a population-based analysis. *BMC Cancer.* 2016;16(413.
32. Joosse A, de Vries E, Eckel R, Nijsten T, Eggermont AM, Holzel D, Coebergh JW, Engel J, and Munich Melanoma G. Gender differences in melanoma survival: female patients have a decreased risk of metastasis. *J Invest Dermatol.* 2011;131(3):719-26.
33. Van Der Esch EP, Cascinelli N, Preda F, Morabito A, and Bufalino R. Stage I melanoma of the skin: evaluation of prognosis according to histologic characteristics. *Cancer.* 1981;48(7):1668-73.
34. Thompson JF, Shaw HM, Hersey P, and Scolyer RA. The history and future of melanoma staging. *J Surg Oncol.* 2004;86(4):224-35.
35. Balch CM, Buzaid AC, Soong SJ, Atkins MB, Cascinelli N, Coit DG, Fleming ID, Gershenwald JE, Houghton A, Jr., Kirkwood JM, et al. Final version of the American Joint Committee on Cancer staging system for cutaneous melanoma. *J Clin Oncol.* 2001;19(16):3635-48.
36. Balch CM, Wilkerson JA, Murad TM, Soong SJ, Ingalls AL, and Maddox WA. The prognostic significance of ulceration of cutaneous melanoma. *Cancer.* 1980;45(12):3012-7.

37. Zendee Elaba MJM, Philip Kerr, Jane M. Grant-Kels. In: Murphy MJ ed. *Diagnostic and Prognostic Biomarkers and Therapeutic Targets in Melanoma*. 2012.
38. Debniak T. Familial malignant melanoma - overview. *Hered Cancer Clin Pract.* 2004;2(3):123-9.
39. Hussussian CJ, Struewing JP, Goldstein AM, Higgins PA, Ally DS, Sheahan MD, Clark WH, Jr., Tucker MA, and Dracopoli NC. Germline p16 mutations in familial melanoma. *Nat Genet.* 1994;8(1):15-21.
40. Kamb A, Shattuck-Eidens D, Eeles R, Liu Q, Gruis NA, Ding W, Hussey C, Tran T, Miki Y, Weaver-Feldhaus J, et al. Analysis of the p16 gene (CDKN2) as a candidate for the chromosome 9p melanoma susceptibility locus. *Nat Genet.* 1994;8(1):23-6.
41. Bruce JL, Hurford RK, Jr., Classon M, Koh J, and Dyson N. Requirements for cell cycle arrest by p16INK4a. *Mol Cell.* 2000;6(3):737-42.
42. Weber HO, Samuel T, Rauch P, and Funk JO. Human p14(ARF)-mediated cell cycle arrest strictly depends on intact p53 signalling pathways. *Oncogene.* 2002;21(20):3207-12.
43. Goldstein AM, Chan M, Harland M, Gillanders EM, Hayward NK, Avril MF, Azizi E, Bianchi-Scarra G, Bishop DT, Bressac-de Paillerets B, et al. High-risk melanoma susceptibility genes and pancreatic cancer, neural system tumours, and uveal melanoma across GenoMEL. *Cancer Res.* 2006;66(20):9818-28.
44. Harland M, Taylor CF, Chambers PA, Kukalizch K, Randerson-Moor JA, Gruis NA, de Snoo FA, ter Huurne JA, Goldstein AM, Tucker MA, et al. A mutation hotspot at the p14ARF splice site. *Oncogene.* 2005;24(28):4604-8.
45. Mistry SH, Taylor C, Randerson-Moor JA, Harland M, Turner F, Barrett JH, Whitaker L, Jenkins RB, Knowles MA, Bishop JA, et al. Prevalence of 9p21 deletions in UK melanoma families. *Genes Chromosomes Cancer.* 2005;44(3):292-300.
46. Randerson-Moor JA, Harland M, Williams S, Cuthbert-Heavens D, Sheridan E, Aveyard J, Sibley K, Whitaker L, Knowles M, Bishop JN, et al. A germline deletion of p14(ARF) but not CDKN2A in a melanoma-neural system tumour syndrome family. *Hum Mol Genet.* 2001;10(1):55-62.
47. Aoude LG, Wadt KA, Pritchard AL, and Hayward NK. Genetics of familial melanoma: 20 years after CDKN2A. *Pigment Cell Melanoma Res.* 2015;28(2):148-60.
48. Zuo L, Weger J, Yang Q, Goldstein AM, Tucker MA, Walker GJ, Hayward N, and Dracopoli NC. Germline mutations in the p16INK4a binding domain of CDK4 in familial melanoma. *Nat Genet.* 1996;12(1):97-9.

49. Horn S, Figl A, Rachakonda PS, Fischer C, Sucker A, Gast A, Kadel S, Moll I, Nagore E, Hemminki K, et al. TERT promoter mutations in familial and sporadic melanoma. *Science.* 2013;339(6122):959-61.
50. Robles-Espinoza CD, Harland M, Ramsay AJ, Aoude LG, Quesada V, Ding Z, Pooley KA, Pritchard AL, Tiffen JC, Petljak M, et al. POT1 loss-of-function variants predispose to familial melanoma. *Nat Genet.* 2014;46(5):478-81.
51. Shi J, Yang XR, Ballew B, Rotunno M, Calista D, Fargnoli MC, Ghiorzo P, Bressac-de Paillerets B, Nagore E, Avril MF, et al. Rare missense variants in POT1 predispose to familial cutaneous malignant melanoma. *Nat Genet.* 2014;46(5):482-6.
52. Aoude LG, Pritchard AL, Robles-Espinoza CD, Wadt K, Harland M, Choi J, Gartside M, Quesada V, Johansson P, Palmer JM, et al. Nonsense mutations in the shelterin complex genes ACD and TERF2IP in familial melanoma. *J Natl Cancer Inst.* 2015;107(2).
53. Bertolotto C, Lesueur F, Giuliano S, Strub T, de Lichy M, Bille K, Dessen P, d'Hayer B, Mohamdi H, Remenieras A, et al. A SUMOylation-defective MITF germline mutation predisposes to melanoma and renal carcinoma. *Nature.* 2011;480(7375):94-8.
54. Yokoyama S, Woods SL, Boyle GM, Aoude LG, MacGregor S, Zismann V, Gartside M, Cust AE, Haq R, Harland M, et al. A novel recurrent mutation in MITF predisposes to familial and sporadic melanoma. *Nature.* 2011;480(7375):99-103.
55. Fernandez LP, Milne RL, Pita G, Floristan U, Sendagorta E, Feito M, Aviles JA, Martin-Gonzalez M, Lazaro P, Benitez J, et al. Pigmentation-related genes and their implication in malignant melanoma susceptibility. *Exp Dermatol.* 2009;18(7):634-42.
56. Planelles D, Nagore E, Moret A, Botella-Estrada R, Vila E, Guillen C, and Montoro JA. HLA class II polymorphisms in Spanish melanoma patients: homozygosity for HLA-DQA1 locus can be a potential melanoma risk factor. *Br J Dermatol.* 2006;154(2):261-6.
57. Pena-Chilet M, Blanquer-Maceiras M, Ibarrola-Villava M, Martinez-Cadenas C, Martin-Gonzalez M, Gomez-Fernandez C, Mayor M, Aviles JA, Lluch A, and Ribas G. Genetic variants in PARP1 (rs3219090) and IRF4 (rs12203592) genes associated with melanoma susceptibility in a Spanish population. *BMC Cancer.* 2013;13(160.
58. Ibarrola-Villava M, Martin-Gonzalez M, Lazaro P, Pizarro A, Lluch A, and Ribas G. Role of glutathione S-transferases in melanoma susceptibility: association with GSTP1 rs1695 polymorphism. *Br J Dermatol.* 2012;166(6):1176-83.

59. Denzer N, Vogt T, and Reichrath J. Vitamin D receptor (VDR) polymorphisms and skin cancer: A systematic review. *Dermatoendocrinol.* 2011;3(3):205-10.
60. Narayanan DL, Saladi RN, and Fox JL. Ultraviolet radiation and skin cancer. *Int J Dermatol.* 2010;49(9):978-86.
61. Elwood JM, Lee JA, Walter SD, Mo T, and Green AE. Relationship of melanoma and other skin cancer mortality to latitude and ultraviolet radiation in the United States and Canada. *Int J Epidemiol.* 1974;3(4):325-32.
62. Moan J, Dahlback A, and Setlow RB. Epidemiological support for an hypothesis for melanoma induction indicating a role for UVA radiation. *Photochem Photobiol.* 1999;70(2):243-7.
63. Holman CD, and Armstrong BK. Cutaneous malignant melanoma and indicators of total accumulated exposure to the sun: an analysis separating histogenetic types. *J Natl Cancer Inst.* 1984;73(1):75-82.
64. Whiteman DC, Whiteman CA, and Green AC. Childhood sun exposure as a risk factor for melanoma: a systematic review of epidemiologic studies. *Cancer Causes Control.* 2001;12(1):69-82.
65. Elwood JM, and Jopson J. Melanoma and sun exposure: an overview of published studies. *Int J Cancer.* 1997;73(2):198-203.
66. Pfahlberg A, Kolmel KF, Gefeller O, and Febim Study G. Timing of excessive ultraviolet radiation and melanoma: epidemiology does not support the existence of a critical period of high susceptibility to solar ultraviolet radiation-induced melanoma. *Br J Dermatol.* 2001;144(3):471-5.
67. Chang YM, Barrett JH, Bishop DT, Armstrong BK, Bataille V, Bergman W, Berwick M, Bracci PM, Elwood JM, Ernstoff MS, et al. Sun exposure and melanoma risk at different latitudes: a pooled analysis of 5700 cases and 7216 controls. *Int J Epidemiol.* 2009;38(3):814-30.
68. Runger TM. Mechanisms of Melanoma Promotion by Ultraviolet Radiation. *J Invest Dermatol.* 2016;136(9):1751-2.
69. Beissert S, and Loser K. Molecular and cellular mechanisms of photocarcinogenesis. *Photochem Photobiol.* 2008;84(1):29-34.
70. Kamenisch Y, Baban TS, Schuller W, von Thaler AK, Sinnberg T, Metzler G, Bauer J, Schittek B, Garbe C, Rocken M, et al. UVA-Irradiation Induces Melanoma Invasion via the Enhanced Warburg Effect. *J Invest Dermatol.* 2016;136(9):1866-75.
71. Waster P, Rosdahl I, Gilmore BF, and Seifert O. Ultraviolet exposure of melanoma cells induces fibroblast activation protein-alpha in fibroblasts: Implications for melanoma invasion. *Int J Oncol.* 2011;39(1):193-202.

72. Leonard MK, Pamidimukkala N, Puts GS, Snyder DE, Slominski AT, and Kaetzel DM. The HGF/SF Mouse Model of UV-Induced Melanoma as an In Vivo Sensor for Metastasis-Regulating Gene. *Int J Mol Sci.* 2017;18(8).
73. Bald T, Quast T, Landsberg J, Rogava M, Glodde N, Lopez-Ramos D, Kohlmeyer J, Riesenberg S, van den Boorn-Konijnenberg D, Homig-Holzel C, et al. Ultraviolet-radiation-induced inflammation promotes angiotropism and metastasis in melanoma. *Nature.* 2014;507(7490):109-13.
74. Krauthammer M, Kong Y, Ha BH, Evans P, Bacchiocchi A, McCusker JP, Cheng E, Davis MJ, Goh G, Choi M, et al. Exome sequencing identifies recurrent somatic RAC1 mutations in melanoma. *Nat Genet.* 2012;44(9):1006-14.
75. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL, et al. Signatures of mutational processes in human cancer. *Nature.* 2013;500(7463):415-21.
76. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature.* 2013;499(7457):214-8.
77. Cancer Genome Atlas Network. Electronic address imo, and Cancer Genome Atlas N. Genomic Classification of Cutaneous Melanoma. *Cell.* 2015;161(7):1681-96.
78. Clark WH, Jr., Reimer RR, Greene M, Ainsworth AM, and Mastrangelo MJ. Origin of familial malignant melanomas from heritable melanocytic lesions. 'The B-K mole syndrome'. *Arch Dermatol.* 1978;114(5):732-8.
79. Elder DE, Goldman LI, Goldman SC, Greene MH, and Clark WH, Jr. Dysplastic nevus syndrome: a phenotypic association of sporadic cutaneous melanoma. *Cancer.* 1980;46(8):1787-94.
80. Duffy K, and Grossman D. The dysplastic nevus: from historical perspective to management in the modern era: part II. Molecular aspects and clinical management. *J Am Acad Dermatol.* 2012;67(1):19 e1-2; quiz 31-2.
81. MacKie RM. Nevi as risk factors for melanoma. *Pediatr Dermatol.* 1992;9(4):340-1.
82. Bergman W, and Fusaro RM. Precursor lesions to melanoma. *Clin Dermatol.* 1992;10(1):21-9.
83. Newton Bishop JA, Bataille V, Pinney E, and Bishop DT. Family studies in melanoma: identification of the atypical mole syndrome (AMS) phenotype. *Melanoma Res.* 1994;4(4):199-206.
84. Ang CG, Kelly JW, Fritschi L, and Dowling JP. Characteristics of familial and non-familial melanoma in Australia. *Melanoma Res.* 1998;8(5):459-64.

85. Hashemi J, Linder S, Platz A, and Hansson J. Melanoma development in relation to non-functional p16/INK4A protein and dysplastic naevus syndrome in Swedish melanoma kindreds. *Melanoma Res.* 1999;9(1):21-30.
86. Landi MT, Calista D, Landi G, Bernucci I, Bertazzi PA, Clark WH, Jr., Goldstein AM, and Tucker MA. Clinical characteristics of 20 Italian melanoma-prone families. *Arch Dermatol.* 1999;135(12):1554-5.
87. Ruiz A, Puig S, Malvehy J, Lazaro C, Lynch M, Gimenez-Arnau AM, Puig L, Sanchez-Conejo J, Estivill X, and Castel T. CDKN2A mutations in Spanish cutaneous malignant melanoma families and patients with multiple melanomas and other neoplasia. *J Med Genet.* 1999;36(6):490-3.
88. Avril MF, Chompret A, Verne-Fourment L, Terrier-Lacombe MJ, Spatz A, Fizazi K, Bressac-de Paillerets B, Demenais F, and Theodore C. Association between germ cell tumours, large numbers of naevi, atypical naevi and melanoma. *Melanoma Res.* 2001;11(2):117-22.
89. Slade J, Marghoob AA, Salopek TG, Rigel DS, Kopf AW, and Bart RS. Atypical mole syndrome: risk factor for cutaneous malignant melanoma and implications for management. *J Am Acad Dermatol.* 1995;32(3):479-94.
90. Crijns MB, Vink J, Van Hees CL, Bergman W, and Vermeer BJ. Dysplastic nevi. Occurrence in first- and second-degree relatives of patients with 'sporadic' dysplastic nevus syndrome. *Arch Dermatol.* 1991;127(9):1346-51.
91. Tucker MA, and Goldstein AM. Melanoma etiology: where are we? *Oncogene.* 2003;22(20):3042-52.
92. Chang YM, Newton-Bishop JA, Bishop DT, Armstrong BK, Bataille V, Bergman W, Berwick M, Bracci PM, Elwood JM, Ernstoff MS, et al. A pooled analysis of melanocytic nevus phenotype and the risk of cutaneous melanoma at different latitudes. *Int J Cancer.* 2009;124(2):420-8.
93. Gandini S, Sera F, Cattaruzza MS, Pasquini P, Abeni D, Boyle P, and Melchi CF. Meta-analysis of risk factors for cutaneous melanoma: I. Common and atypical naevi. *Eur J Cancer.* 2005;41(1):28-44.
94. Bataille V, Grulich A, Sasieni P, Swerdlow A, Newton Bishop J, McCarthy W, Hersey P, and Cuzick J. The association between naevi and melanoma in populations with different levels of sun exposure: a joint case-control study of melanoma in the UK and Australia. *Br J Cancer.* 1998;77(3):505-10.
95. Harrison SL, MacLennan R, Speare R, and Wronski I. Sun exposure and melanocytic naevi in young Australian children. *Lancet.* 1994;344(8936):1529-32.
96. Stierner U. Melanocytes, moles and melanoma--a study on UV effects. *Acta Derm Venereol Suppl (Stockh).* 1991;168(1-31.

97. Falchi M, Bataille V, Hayward NK, Duffy DL, Bishop JA, Pastinen T, Cervino A, Zhao ZZ, Deloukas P, Soranzo N, et al. Genome-wide association study identifies variants at 9p21 and 22q13 associated with development of cutaneous nevi. *Nat Genet.* 2009;41(8):915-9.

98. Bishop JA, Wachsmuth RC, Harland M, Bataille V, Pinney E, Mac KP, Baglietto L, Cuzick J, and Bishop DT. Genotype/phenotype and penetrance studies in melanoma families with germline CDKN2A mutations. *J Invest Dermatol.* 2000;114(1):28-33.

99. Zhu G, Montgomery GW, James MR, Trent JM, Hayward NK, Martin NG, and Duffy DL. A genome-wide scan for naevus count: linkage to CDKN2A and to other chromosome regions. *Eur J Hum Genet.* 2007;15(1):94-102.

100. de Snoo FA, Hottenga JJ, Gillanders EM, Sandkuijl LA, Jones MP, Bergman W, van der Drift C, van Leeuwen I, van Mourik L, Huurne JA, et al. Genome-wide linkage scan for atypical nevi in p16-Leiden melanoma families. *Eur J Hum Genet.* 2008;16(9):1135-41.

101. Halpern AC, Guerry Dt, Elder DE, Trock B, Synnestvedt M, and Humphreys T. Natural history of dysplastic nevi. *J Am Acad Dermatol.* 1993;29(1):51-7.

102. Tsao H, Bevona C, Goggins W, and Quinn T. The transformation rate of moles (melanocytic nevi) into cutaneous melanoma: a population-based estimate. *Arch Dermatol.* 2003;139(3):282-8.

103. Pollock PM, Harper UL, Hansen KS, Yudt LM, Stark M, Robbins CM, Moses TY, Hostetter G, Wagner U, Kakareka J, et al. High frequency of BRAF mutations in nevi. *Nat Genet.* 2003;33(1):19-20.

104. Loewe R, Kittler H, Fischer G, Fae I, Wolff K, and Petzelbauer P. BRAF kinase gene V599E mutation in growing melanocytic lesions. *J Invest Dermatol.* 2004;123(4):733-6.

105. Papp T, Schipper H, Kumar K, Schiffmann D, and Zimmermann R. Mutational analysis of the BRAF gene in human congenital and dysplastic melanocytic naevi. *Melanoma Res.* 2005;15(5):401-7.

106. Zalaudek I, Guelly C, Pellacani G, Hofmann-Wellenhof R, Trajanoski S, Kittler H, Scope A, Marghoob AA, Longo C, Leinweber B, et al. The dermoscopical and histopathological patterns of nevi correlate with the frequency of BRAF mutations. *J Invest Dermatol.* 2011;131(2):542-5.

107. Damsky WE, and Bosenberg M. Melanocytic nevi and melanoma: unraveling a complex relationship. *Oncogene.* 2017;36(42):5771-92.

108. Laud K, Marian C, Avril MF, Barrois M, Chompret A, Goldstein AM, Tucker MA, Clark PA, Peters G, Chaudru V, et al. Comprehensive analysis of CDKN2A (p16INK4A/p14ARF) and CDKN2B genes in 53 melanoma index cases

considered to be at heightened risk of melanoma. *J Med Genet.* 2006;43(1):39-47.

109. Michaloglou C, Vredeveld LC, Mooi WJ, and Peeper DS. BRAF(E600) in benign and malignant human tumours. *Oncogene.* 2008;27(7):877-95.
110. Roh MR, Eliades P, Gupta S, and Tsao H. Genetics of melanocytic nevi. *Pigment Cell Melanoma Res.* 2015;28(6):661-72.
111. Scherer D, and Kumar R. Genetics of pigmentation in skin cancer--a review. *Mutat Res.* 2010;705(2):141-53.
112. Bliss JM, Ford D, Swerdlow AJ, Armstrong BK, Cristofolini M, Elwood JM, Green A, Holly EA, Mack T, MacKie RM, et al. Risk of cutaneous melanoma associated with pigmentation characteristics and freckling: systematic overview of 10 case-control studies. The International Melanoma Analysis Group (IMAGE). *Int J Cancer.* 1995;62(4):367-76.
113. Jablonski NG, and Chaplin G. The evolution of human skin coloration. *J Hum Evol.* 2000;39(1):57-106.
114. Morgan AM, Lo J, and Fisher DE. How does pheomelanin synthesis contribute to melanomagenesis?: Two distinct mechanisms could explain the carcinogenicity of pheomelanin synthesis. *Bioessays.* 2013;35(8):672-6.
115. Ancans J, Tobin DJ, Hoogduijn MJ, Smit NP, Wakamatsu K, and Thody AJ. Melanosomal pH controls rate of melanogenesis, eumelanin/phaeomelanin ratio and melanosome maturation in melanocytes and melanoma cells. *Exp Cell Res.* 2001;268(1):26-35.
116. Rees JL, and Healy E. Melanocortin receptors, red hair, and skin cancer. *J Investig Dermatol Symp Proc.* 1997;2(1):94-8.
117. Rees JL, and Harding RM. Understanding the evolution of human pigmentation: recent contributions from population genetics. *J Invest Dermatol.* 2012;132(3 Pt 2):846-53.
118. Garcia-Borron JC, Abdel-Malek Z, and Jimenez-Cervantes C. MC1R, the cAMP pathway, and the response to solar UV: extending the horizon beyond pigmentation. *Pigment Cell Melanoma Res.* 2014;27(5):699-720.
119. Beaumont KA, Shekar SN, Newton RA, James MR, Stow JL, Duffy DL, and Sturm RA. Receptor function, dominant negative activity and phenotype correlations for MC1R variant alleles. *Hum Mol Genet.* 2007;16(18):2249-60.
120. Doyle JR, Fortin JP, Beinborn M, and Kopin AS. Selected melanocortin 1 receptor single-nucleotide polymorphisms differentially alter multiple signalling pathways. *J Pharmacol Exp Ther.* 2012;342(2):318-26.
121. Schioth HB, Phillips SR, Rudzish R, Birch-Machin MA, Wikberg JE, and Rees JL. Loss of function mutations of the human melanocortin 1 receptor are

common and are associated with red hair. *Biochem Biophys Res Commun.* 1999;260(2):488-91.

122. Valverde P, Healy E, Jackson I, Rees JL, and Thody AJ. Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans. *Nat Genet.* 1995;11(3):328-30.
123. Beaumont KA, Shekar SN, Cook AL, Duffy DL, and Sturm RA. Red hair is the null phenotype of MC1R. *Hum Mutat.* 2008;29(8):E88-94.
124. Raimondi S, Sera F, Gandini S, Iodice S, Caini S, Maisonneuve P, and Fargnoli MC. MC1R variants, melanoma and red hair color phenotype: a meta-analysis. *Int J Cancer.* 2008;122(12):2753-60.
125. Bishop DT, Demenais F, Iles MM, Harland M, Taylor JC, Corda E, Randerson-Moor J, Aitken JF, Avril MF, Azizi E, et al. Genome-wide association study identifies three loci associated with melanoma risk. *Nat Genet.* 2009;41(8):920-5.
126. Pasquali E, Garcia-Borron JC, Fargnoli MC, Gandini S, Maisonneuve P, Bagnardi V, Specchia C, Liu F, Kayser M, Nijsten T, et al. MC1R variants increased the risk of sporadic cutaneous melanoma in darker-pigmented Caucasians: a pooled-analysis from the M-SKIP project. *Int J Cancer.* 2015;136(3):618-31.
127. Robles-Espinoza CD, Roberts ND, Chen S, Leacy FP, Alexandrov LB, Pornputtapong N, Halaban R, Krauthammer M, Cui R, Timothy Bishop D, et al. Germline MC1R status influences somatic mutation burden in melanoma. *Nat Commun.* 2016;7(12064.
128. Luger TA, Scholzen T, Brzoska T, Becher E, Slominski A, and Paus R. Cutaneous immunomodulation and coordination of skin stress responses by alpha-melanocyte-stimulating hormone. *Ann N Y Acad Sci.* 1998;840(381-94.
129. Catania A, Gatti S, Colombo G, and Lipton JM. Targeting melanocortin receptors as a novel strategy to control inflammation. *Pharmacol Rev.* 2004;56(1):1-29.
130. Jarrett SG, Carter KM, Shelton BJ, and D'Orazio JA. The melanocortin signalling cAMP axis accelerates repair and reduces mutagenesis of platinum-induced DNA damage. *Sci Rep.* 2017;7(1):11708.
131. Jarrett SG, Wolf Horrell EM, Christian PA, Vanover JC, Boulanger MC, Zou Y, and D'Orazio JA. PKA-mediated phosphorylation of ATR promotes recruitment of XPA to UV-induced DNA damage. *Mol Cell.* 2014;54(6):999-1011.
132. John A. D'Orazio AM, James Lagrew and W. Brooke Veith. In: Armstrong AW ed. *Advances in Malignant Melanoma*. IntechOpen; 2011.

133. Cambridge EL, McIntyre Z, Clare S, Arends MJ, Goulding D, Isherwood C, Caetano SS, Reviriego CB, Swiatkowska A, Kane L, et al. The AMP-activated protein kinase beta 1 subunit modulates erythrocyte integrity. *Exp Hematol.* 2017;45(64-8 e5.
134. Gudbjartsson DF, Sulem P, Stacey SN, Goldstein AM, Rafnar T, Sigurgeirsson B, Benediktsdottir KR, Thorisdottir K, Ragnarsson R, Sveinsdottir SG, et al. ASIP and TYR pigmentation variants associate with cutaneous melanoma and basal cell carcinoma. *Nat Genet.* 2008;40(7):886-91.
135. Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Magnusson KP, Manolescu A, Karason A, Palsson A, Thorleifsson G, et al. Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat Genet.* 2007;39(12):1443-52.
136. Amos CI, Wang LE, Lee JE, Gershenwald JE, Chen WV, Fang S, Kosoy R, Zhang M, Qureshi AA, Vattathil S, et al. Genome-wide association study identifies novel loci predisposing to cutaneous melanoma. *Hum Mol Genet.* 2011;20(24):5012-23.
137. Hartman ML, and Czyz M. Pro-survival role of MITF in melanoma. *J Invest Dermatol.* 2015;135(2):352-8.
138. Hsiao JJ, and Fisher DE. The roles of microphthalmia-associated transcription factor and pigmentation in melanoma. *Arch Biochem Biophys.* 2014;563(28-34.
139. Hilger RA, Scheulen ME, and Strumberg D. The Ras-Raf-MEK-ERK pathway in the treatment of cancer. *Onkologie.* 2002;25(6):511-8.
140. Kolch W. Meaningful relationships: the regulation of the Ras/Raf/MEK/ERK pathway by protein interactions. *Biochem J.* 2000;351 Pt 2(289-305.
141. Houben R, Becker JC, Kappel A, Terheyden P, Brocker EB, Goetz R, and Rapp UR. Constitutive activation of the Ras-Raf signalling pathway in metastatic melanoma is associated with poor prognosis. *J Carcinog.* 2004;3(6.
142. Edlundh-Rose E, Egyhazi S, Omholt K, Mansson-Brahme E, Platz A, Hansson J, and Lundeberg J. NRAS and BRAF mutations in melanoma tumours in relation to clinical characteristics: a study based on mutation screening by pyrosequencing. *Melanoma Res.* 2006;16(6):471-8.
143. Platz A, Egyhazi S, Ringborg U, and Hansson J. Human cutaneous melanoma; a review of NRAS and BRAF mutation frequencies in relation to histogenetic subclass and body site. *Mol Oncol.* 2008;1(4):395-405.
144. Burd CE, Liu W, Huynh MV, Waqas MA, Gillahan JE, Clark KS, Fu K, Martin BL, Jeck WR, Souroullas GP, et al. Mutation-specific RAS oncogenicity explains NRAS codon 61 selection in melanoma. *Cancer Discov.* 2014;4(12):1418-29.

145. Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, Teague J, Woffendin H, Garnett MJ, Bottomley W, et al. Mutations of the BRAF gene in human cancer. *Nature.* 2002;417(6892):949-54.
146. Dankort D, Curley DP, Cartlidge RA, Nelson B, Karnezis AN, Damsky WE, Jr., You MJ, DePinho RA, McMahon M, and Bosenberg M. Braf(V600E) cooperates with Pten loss to induce metastatic melanoma. *Nat Genet.* 2009;41(5):544-52.
147. Patton EE, Widlund HR, Kutok JL, Kopani KR, Amatruda JF, Murphey RD, Berghmans S, Mayhall EA, Traver D, Fletcher CD, et al. BRAF mutations are sufficient to promote nevi formation and cooperate with p53 in the genesis of melanoma. *Curr Biol.* 2005;15(3):249-54.
148. Yu H, McDaid R, Lee J, Possik P, Li L, Kumar SM, Elder DE, Van Belle P, Gimotty P, Guerra M, et al. The role of BRAF mutation and p53 inactivation during transformation of a subpopulation of primary human melanocytes. *Am J Pathol.* 2009;174(6):2367-77.
149. Yu JS, and Cui W. Proliferation, survival and metabolism: the role of PI3K/AKT/mTOR signalling in pluripotency and cell fate determination. *Development.* 2016;143(17):3050-60.
150. Blume-Jensen P, and Hunter T. Oncogenic kinase signalling. *Nature.* 2001;411(6835):355-65.
151. Hemmings BA, and Restuccia DF. The PI3K-PKB/Akt pathway. *Cold Spring Harb Perspect Biol.* 2015;7(4).
152. Stahl JM, Sharma A, Cheung M, Zimmerman M, Cheng JQ, Bosenberg MW, Kester M, Sandirasegarane L, and Robertson GP. Deregulated Akt3 activity promotes development of malignant melanoma. *Cancer Res.* 2004;64(19):7002-10.
153. Robertson GP. Functional and therapeutic significance of Akt deregulation in malignant melanoma. *Cancer Metastasis Rev.* 2005;24(2):273-85.
154. Dhawan P, Singh AB, Ellis DL, and Richmond A. Constitutive activation of Akt/protein kinase B in melanoma leads to up-regulation of nuclear factor-kappaB and tumour progression. *Cancer Res.* 2002;62(24):7335-42.
155. Bastian BC, LeBoit PE, Hamm H, Brocker EB, and Pinkel D. Chromosomal gains and losses in primary cutaneous melanomas detected by comparative genomic hybridization. *Cancer Res.* 1998;58(10):2170-5.
156. Robertson GP, Furnari FB, Miele ME, Glendening MJ, Welch DR, Fountain JW, Lugo TG, Huang HJ, and Cavenee WK. In vitro loss of heterozygosity targets the PTEN/MMAC1 gene in melanoma. *Proc Natl Acad Sci U S A.* 1998;95(16):9418-23.

157. Stahl JM, Cheung M, Sharma A, Trivedi NR, Shanmugam S, and Robertson GP. Loss of PTEN promotes tumour development in malignant melanoma. *Cancer Res.* 2003;63(11):2881-90.
158. Conde-Perez A, Gros G, Longvert C, Pedersen M, Petit V, Aktary Z, Viros A, Gesbert F, Delmas V, Rambow F, et al. A caveolin-dependent and PI3K/AKT-independent role of PTEN in beta-catenin transcriptional activity. *Nat Commun.* 2015;6(8093.
159. Davies MA, Stemke-Hale K, Tellez C, Calderone TL, Deng W, Prieto VG, Lazar AJ, Gershenwald JE, and Mills GB. A novel AKT3 mutation in melanoma tumours and cell lines. *Br J Cancer.* 2008;99(8):1265-8.
160. Hartman ML, and Czyz M. MITF in melanoma: mechanisms behind its expression and activity. *Cell Mol Life Sci.* 2015;72(7):1249-60.
161. Garraway LA, Widlund HR, Rubin MA, Getz G, Berger AJ, Ramaswamy S, Beroukhim R, Milner DA, Granter SR, Du J, et al. Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature.* 2005;436(7047):117-22.
162. Rivera RS, Nagatsuka H, Gunduz M, Cengiz B, Gunduz E, Siar CH, Tsujigiwa H, Tamamura R, Han KN, and Nagai N. C-kit protein expression correlated with activating mutations in KIT gene in oral mucosal melanoma. *Virchows Arch.* 2008;452(1):27-32.
163. Ashida A, Takata M, Murata H, Kido K, and Saida T. Pathological activation of KIT in metastatic tumours of acral and mucosal melanomas. *Int J Cancer.* 2009;124(4):862-8.
164. Sherr CJ. Principles of tumour suppression. *Cell.* 2004;116(2):235-46.
165. Akslen LA, Monstad SE, Larsen B, Straume O, and Ogreid D. Frequent mutations of the p53 gene in cutaneous melanoma of the nodular type. *Int J Cancer.* 1998;79(1):91-5.
166. Papp T, Jafari M, and Schiffmann D. Lack of p53 mutations and loss of heterozygosity in non-cultured human melanocytic lesions. *J Cancer Res Clin Oncol.* 1996;122(9):541-8.
167. Krauthammer M, Kong Y, Bacchiocchi A, Evans P, Pornputtapong N, Wu C, McCusker JP, Ma S, Cheng E, Straub R, et al. Exome sequencing identifies recurrent mutations in NF1 and RASopathy genes in sun-exposed melanomas. *Nat Genet.* 2015;47(9):996-1002.
168. Paget S. The distribution of secondary growths in cancer of the breast. 1889. *Cancer Metastasis Rev.* 1989;8(2):98-101.
169. Gambichler T, Kempka J, Kampilafkos P, Bechara FG, Altmeyer P, and Stucker M. Clinicopathological characteristics of 270 patients with lentigo maligna and

lentigo maligna melanoma: data from a German skin cancer centre. *Br J Dermatol.* 2014;171(6):1605-7.

170. Wang M, Zhao J, Zhang L, Wei F, Lian Y, Wu Y, Gong Z, Zhang S, Zhou J, Cao K, et al. Role of tumour microenvironment in tumourigenesis. *J Cancer.* 2017;8(5):761-73.
171. McGovern VJ. Spontaneous regression of melanoma. *Pathology.* 1975;7(2):91-9.
172. Park CK, and Kim SK. Clinicopathological significance of intratumoural and peritumoural lymphocytes and lymphocyte score based on the histologic subtypes of cutaneous melanoma. *Oncotarget.* 2017;8(9):14759-69.
173. Saldanha G, Flatman K, Teo KW, and Bamford M. A Novel Numerical Scoring System for Melanoma Tumour-infiltrating Lymphocytes Has Better Prognostic Value Than Standard Scoring. *Am J Surg Pathol.* 2017;41(7):906-14.
174. Donizy P, Kaczorowski M, Halon A, Leskiewicz M, Kozyra C, and Matkowski R. Paucity of tumour-infiltrating lymphocytes is an unfavorable prognosticator and predicts lymph node metastases in cutaneous melanoma patients. *Anticancer Res.* 2015;35(1):351-8.
175. Fortes C, Mastroeni S, Mannooranparampil TJ, Passarelli F, Zappala A, Annessi G, Marino C, Caggiati A, Russo N, and Michelozzi P. Tumour-infiltrating lymphocytes predict cutaneous melanoma survival. *Melanoma Res.* 2015;25(4):306-11.
176. Weiss SA, Han SW, Lui K, Tchack J, Shapiro R, Berman R, Zhong J, Krogsgaard M, Osman I, and Darvishian F. Immunologic heterogeneity of tumour-infiltrating lymphocyte composition in primary melanoma. *Hum Pathol.* 2016;57(116-25.
177. Lee SJ, Lim HJ, Choi YH, Chang YH, Lee WJ, Kim DW, and Yoon GS. The clinical significance of tumour-infiltrating lymphocytes and microscopic satellites in acral melanoma in a korean population. *Ann Dermatol.* 2013;25(1):61-6.
178. Barnes TA, and Amir E. HYPE or HOPE: the prognostic value of infiltrating immune cells in cancer. *Br J Cancer.* 2018;118(2):e5.
179. Eriksson H, Frohm-Nilsson M, Jaras J, Kanter-Lewensohn L, Kjellman P, Mansson-Brahme E, Vassilaki I, and Hansson J. Prognostic factors in localized invasive primary cutaneous malignant melanoma: results of a large population-based study. *Br J Dermatol.* 2015;172(1):175-86.
180. Kakavand H, Vilain RE, Wilmott JS, Burke H, Yearley JH, Thompson JF, Hersey P, Long GV, and Scolyer RA. Tumour PD-L1 expression, immune cell correlates and PD-1+ lymphocytes in sentinel lymph node melanoma metastases. *Mod Pathol.* 2015;28(12):1535-44.

181. Erdag G, Schaefer JT, Smolkin ME, Deacon DH, Shea SM, Dengel LT, Patterson JW, and Slingluff CL, Jr. Immunotype and immunohistologic characteristics of tumour-infiltrating immune cells are associated with clinical outcome in metastatic melanoma. *Cancer Res.* 2012;72(5):1070-80.
182. Garg K, Maurer M, Griss J, Bruggen MC, Wolf IH, Wagner C, Willi N, Mertz KD, and Wagner SN. Tumour-associated B cells in cutaneous primary melanoma and improved clinical outcome. *Hum Pathol.* 2016;54(157-64.
183. Ladanyi A, Kiss J, Mohos A, Somlai B, Liszkay G, Gilde K, Fejos Z, Gaudi I, Dobos J, and Timar J. Prognostic impact of B-cell density in cutaneous melanoma. *Cancer Immunol Immunother.* 2011;60(12):1729-38.
184. Messaoudene M, Perier A, Fregni G, Neves E, Zitvogel L, Cremer I, Chanal J, Sastre-Garau X, Deschamps L, Marinho E, et al. Characterization of the Microenvironment in Positive and Negative Sentinel Lymph Nodes from Melanoma Patients. *PLoS One.* 2015;10(7):e0133363.
185. Knol AC, Nguyen JM, Quereux G, Brocard A, Khammari A, and Dreno B. Prognostic value of tumour-infiltrating Foxp3+ T-cell subpopulations in metastatic melanoma. *Exp Dermatol.* 2011;20(5):430-4.
186. Harlin H, Kuna TV, Peterson AC, Meng Y, and Gajewski TF. Tumour progression despite massive influx of activated CD8(+) T cells in a patient with malignant melanoma ascites. *Cancer Immunol Immunother.* 2006;55(10):1185-97.
187. Cabrera CM. The double role of the endoplasmic reticulum chaperone tapasin in peptide optimization of HLA class I molecules. *Scand J Immunol.* 2007;65(6):487-93.
188. Bernsen MR, Hakansson L, Gustafsson B, Krysander L, Rettrup B, Ruiter D, and Hakansson A. On the biological relevance of MHC class II and B7 expression by tumour cells in melanoma metastases. *Br J Cancer.* 2003;88(3):424-31.
189. Seliger B, Ritz U, Abele R, Bock M, Tampe R, Sutter G, Drexler I, Huber C, and Ferrone S. Immune escape of melanoma: first evidence of structural alterations in two distinct components of the MHC class I antigen processing pathway. *Cancer Res.* 2001;61(24):8647-50.
190. Hussein MR. Dendritic cells and melanoma tumourigenesis: an insight. *Cancer Biol Ther.* 2005;4(5):501-5.
191. Janeway CA Jr TP, Walport M, et al. *Immunobiology: The Immune System in Health and Disease*. New York: Garland Science; 2001.

192. Passarelli A, Mannavola F, Stucci LS, Tucci M, and Silvestris F. Immune system and melanoma biology: a balance between immunosurveillance and immune escape. *Oncotarget.* 2017;8(62):106132-42.
193. Cooper MA, Colonna M, and Yokoyama WM. Hidden talents of natural killers: NK cells in innate and adaptive immunity. *EMBO Rep.* 2009;10(10):1103-10.
194. Mirjacic Martinovic KM, Babovic N, Dzodic RR, Jurisic VB, Tanic NT, and Konjevic GM. Decreased expression of NKG2D, NKp46, DNAM-1 receptors, and intracellular perforin and STAT-1 effector molecules in NK cells and their dim and bright subsets in metastatic melanoma patients. *Melanoma Res.* 2014;24(4):295-304.
195. Tarazona R, Duran E, and Solana R. Natural Killer Cell Recognition of Melanoma: New Clues for a More Effective Immunotherapy. *Front Immunol.* 2015;6(649.
196. Beissert S, Schwarz A, and Schwarz T. Regulatory T cells. *J Invest Dermatol.* 2006;126(1):15-24.
197. Talmadge JE, and Gabrilovich DI. History of myeloid-derived suppressor cells. *Nature reviews Cancer.* 2013;13(10):739-52.
198. Munn DH. Indoleamine 2,3-dioxygenase, tumour-induced tolerance and counter-regulation. *Curr Opin Immunol.* 2006;18(2):220-5.
199. Lakshmikanth T, Burke S, Ali TH, Kimpfler S, Ursini F, Ruggeri L, Capanni M, Umansky V, Paschen A, Sucker A, et al. NCRs and DNAM-1 mediate NK cell recognition and lysis of human and mouse melanoma cell lines in vitro and in vivo. *J Clin Invest.* 2009;119(5):1251-63.
200. Jordan KR, Amaria RN, Ramirez O, Callihan EB, Gao D, Borakove M, Manthey E, Borges VF, and McCarter MD. Myeloid-derived suppressor cells are associated with disease progression and decreased overall survival in advanced-stage melanoma patients. *Cancer Immunol Immunother.* 2013;62(11):1711-22.
201. Madar S, Goldstein I, and Rotter V. 'Cancer associated fibroblasts'--more than meets the eye. *Trends Mol Med.* 2013;19(8):447-53.
202. Kalluri R. The biology and function of fibroblasts in cancer. *Nature reviews Cancer.* 2016;16(9):582-98.
203. Cortez E, Roswall P, and Pietras K. Functional subsets of mesenchymal cell types in the tumour microenvironment. *Semin Cancer Biol.* 2014;25(3-9.
204. Ishii G, Ochiai A, and Neri S. Phenotypic and functional heterogeneity of cancer-associated fibroblast within the tumour microenvironment. *Adv Drug Deliv Rev.* 2016;99(Pt B):186-96.

205. Lakins MA, Ghorani E, Munir H, Martins CP, and Shields JD. Cancer-associated fibroblasts induce antigen-specific deletion of CD8 (+) T Cells to protect tumour cells. *Nat Commun.* 2018;9(1):948.
206. Ziani L, Safta-Saadoun TB, Gourbeix J, Cavalcanti A, Robert C, Favre G, Chouaib S, and Thiery J. Melanoma-associated fibroblasts decrease tumour cell susceptibility to NK cell-mediated killing through matrix-metalloproteinases secretion. *Oncotarget.* 2017;8(12):19780-94.
207. Hirata E, Girotti MR, Viros A, Hooper S, Spencer-Dene B, Matsuda M, Larkin J, Marais R, and Sahai E. Intravital imaging reveals how BRAF inhibition generates drug-tolerant microenvironments with high integrin beta1/FAK signalling. *Cancer Cell.* 2015;27(4):574-88.
208. Dunn GP, Old LJ, and Schreiber RD. The three Es of cancer immunoediting. *Annu Rev Immunol.* 2004;22(329-60.
209. Farkona S, Diamandis EP, and Blasutig IM. Cancer immunotherapy: the beginning of the end of cancer? *BMC Med.* 2016;14(73.
210. Peggs KS, Quezada SA, and Allison JP. Cell intrinsic mechanisms of T-cell inhibition and application to cancer therapy. *Immunol Rev.* 2008;224(141-65.
211. Huang CT, Workman CJ, Flies D, Pan X, Marson AL, Zhou G, Hipkiss EL, Ravi S, Kowalski J, Levitsky HI, et al. Role of LAG-3 in regulatory T cells. *Immunity.* 2004;21(4):503-13.
212. Robert C, Long GV, Brady B, Dutriaux C, Maio M, Mortier L, Hassel JC, Rutkowski P, McNeil C, Kalinka-Warzocha E, et al. Nivolumab in previously untreated melanoma without BRAF mutation. *N Engl J Med.* 2015;372(4):320-30.
213. Taube JM, Young GD, McMiller TL, Chen S, Salas JT, Pritchard TS, Xu H, Meeker AK, Fan J, Cheadle C, et al. Differential Expression of Immune-Regulatory Genes Associated with PD-L1 Display in Melanoma: Implications for PD-1 Pathway Blockade. *Clin Cancer Res.* 2015;21(17):3969-76.
214. Johnson DB, Estrada MV, Salgado R, Sanchez V, Doxie DB, Opalenik SR, Vilgelm AE, Feld E, Johnson AS, Greenplate AR, et al. Melanoma-specific MHC-II expression represents a tumour-autonomous phenotype and predicts response to anti-PD-1/PD-L1 therapy. *Nat Commun.* 2016;7(10582.
215. Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, Walsh LA, Postow MA, Wong P, Ho TS, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N Engl J Med.* 2014;371(23):2189-99.
216. Van Allen EM, Miao D, Schilling B, Shukla SA, Blank C, Zimmer L, Sucker A, Hillen U, Foppen MHG, Goldinger SM, et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science.* 2015;350(6257):207-11.

217. Riaz N, Havel JJ, Kendall SM, Makarov V, Walsh LA, Desrichard A, Weinhold N, and Chan TA. Recurrent SERPINB3 and SERPINB4 mutations in patients who respond to anti-CTLA4 immunotherapy. *Nat Genet.* 2016;48(11):1327-9.
218. Hugo W, Zaretsky JM, Sun L, Song C, Moreno BH, Hu-Lieskovan S, Berent-Maoz B, Pang J, Chmielowski B, Cherry G, et al. Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell.* 2017;168(3):542.
219. Roh W, Chen PL, Reuben A, Spencer CN, Prieto PA, Miller JP, Gopalakrishnan V, Wang F, Cooper ZA, Reddy SM, et al. Integrated molecular analysis of tumour biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance. *Sci Transl Med.* 2017;9(379).
220. Inoue H, Park JH, Kiyotani K, Zewde M, Miyashita A, Jinnin M, Kiniwa Y, Okuyama R, Tanaka R, Fujisawa Y, et al. Intratumoural expression levels of PD-L1, GZMA, and HLA-A along with oligoclonal T cell expansion associate with response to nivolumab in metastatic melanoma. *Oncoimmunology.* 2016;5(9):e1204507.
221. Johnson DB, Frampton GM, Rioth MJ, Yusko E, Xu Y, Guo X, Ennis RC, Fabrizio D, Chalmers ZR, Greenbowe J, et al. Targeted Next Generation Sequencing Identifies Markers of Response to PD-1 Blockade. *Cancer Immunol Res.* 2016;4(11):959-67.
222. Hogan SA, Courtier A, Cheng PF, Jaberg-Bentele NF, Goldinger SM, Manuel M, Perez S, Plantier N, Mouret JF, Nguyen-Kim TDL, et al. Peripheral Blood TCR Repertoire Profiling May Facilitate Patient Stratification for Immunotherapy against Melanoma. *Cancer Immunol Res.* 2019;7(1):77-85.
223. Bouillon R, and Suda T. Vitamin D: calcium and bone homeostasis during evolution. *Bonekey Rep.* 2014;3(480.
224. Holick MF, MacLaughlin JA, Clark MB, Holick SA, Potts JT, Jr., Anderson RR, Blank IH, Parrish JA, and Elias P. Photosynthesis of previtamin D3 in human skin and the physiologic consequences. *Science.* 1980;210(4466):203-5.
225. Webb AR, DeCosta BR, and Holick MF. Sunlight regulates the cutaneous production of vitamin D3 by causing its photodegradation. *J Clin Endocrinol Metab.* 1989;68(5):882-7.
226. Klingberg E, Olerod G, Konar J, Petzold M, and Hammarsten O. Seasonal variations in serum 25-hydroxy vitamin D levels in a Swedish cohort. *Endocrine.* 2015;49(3):800-8.
227. Field S, Davies J, Bishop DT, and Newton-Bishop JA. Vitamin D and melanoma. *Dermatoendocrinol.* 2013;5(1):121-9.

228. Bikle DD. Vitamin D metabolism, mechanism of action, and clinical applications. *Chem Biol.* 2014;21(3):319-29.
229. Deeb KK, Trump DL, and Johnson CS. Vitamin D signalling pathways in cancer: potential for anticancer therapeutics. *Nature reviews Cancer.* 2007;7(9):684-700.
230. Haussler MR, Jurutka PW, Mizwicki M, and Norman AW. Vitamin D receptor (VDR)-mediated actions of 1alpha,25(OH)(2)vitamin D(3): genomic and non-genomic mechanisms. *Best practice & research Clinical endocrinology & metabolism.* 2011;25(4):543-59.
231. Altucci L, Leibowitz MD, Ogilvie KM, de Lera AR, and Gronemeyer H. RAR and RXR modulation in cancer and metabolic disease. *Nat Rev Drug Discov.* 2007;6(10):793-810.
232. Haussler MR, Haussler CA, Jurutka PW, Thompson PD, Hsieh JC, Remus LS, Selznick SH, and Whitfield GK. The vitamin D hormone and its nuclear receptor: molecular actions and disease states. *J Endocrinol.* 1997;154 Suppl(S57-73.
233. Mangelsdorf DJ, Thummel C, Beato M, Herrlich P, Schutz G, Umesono K, Blumberg B, Kastner P, Mark M, Chambon P, et al. The nuclear receptor superfamily: the second decade. *Cell.* 1995;83(6):835-9.
234. Baker AR, McDonnell DP, Hughes M, Crisp TM, Mangelsdorf DJ, Haussler MR, Pike JW, Shine J, and O'Malley BW. Cloning and expression of full-length cDNA encoding human vitamin D receptor. *Proc Natl Acad Sci U S A.* 1988;85(10):3294-8.
235. Wan LY, Zhang YQ, Chen MD, Liu CB, and Wu JF. Relationship of structure and function of DNA-binding domain in vitamin D receptor. *Molecules.* 2015;20(7):12389-99.
236. Kraichely DM, and MacDonald PN. Transcriptional activation through the vitamin D receptor in osteoblasts. *Frontiers in bioscience : a journal and virtual library.* 1998;3(d821-33.
237. Umesono K, Murakami KK, Thompson CC, and Evans RM. Direct repeats as selective response elements for the thyroid hormone, retinoic acid, and vitamin D3 receptors. *Cell.* 1991;65(7):1255-66.
238. Kim MS, Fujiki R, Murayama A, Kitagawa H, Yamaoka K, Yamamoto Y, Mihara M, Takeyama K, and Kato S. 1Alpha,25(OH)2D3-induced transrepression by vitamin D receptor through E-box-type elements in the human parathyroid hormone gene promoter. *Molecular endocrinology.* 2007;21(2):334-42.
239. Ozono K, Liao J, Kerner SA, Scott RA, and Pike JW. The vitamin D-responsive element in the human osteocalcin gene. Association with a nuclear proto-

oncogene enhancer. *The Journal of biological chemistry.* 1990;265(35):21881-8.

240. Kerner SA, Scott RA, and Pike JW. Sequence elements in the human osteocalcin gene confer basal activation and inducible response to hormonal vitamin D3. *Proc Natl Acad Sci U S A.* 1989;86(12):4455-9.
241. Spencer TE, Jenster G, Burcin MM, Allis CD, Zhou J, Mizzen CA, McKenna NJ, Onate SA, Tsai SY, Tsai MJ, et al. Steroid receptor coactivator-1 is a histone acetyltransferase. *Nature.* 1997;389(6647):194-8.
242. Kraichely DM, Collins JJ, 3rd, DeLisle RK, and MacDonald PN. The autonomous transactivation domain in helix H3 of the vitamin D receptor is required for transactivation and coactivator interaction. *The Journal of biological chemistry.* 1999;274(20):14352-8.
243. Kraichely DM, Nakai YD, and MacDonald PN. Identification of an autonomous transactivation domain in helix H3 of the vitamin D receptor. *Journal of cellular biochemistry.* 1999;75(1):82-92.
244. Rachez C, Suldan Z, Ward J, Chang CP, Burakov D, Erdjument-Bromage H, Tempst P, and Freedman LP. A novel protein complex that interacts with the vitamin D3 receptor in a ligand-dependent manner and enhances VDR transactivation in a cell-free system. *Genes & development.* 1998;12(12):1787-800.
245. Peleg S, and Nguyen CV. The importance of nuclear import in protection of the vitamin D receptor from polyubiquitination and proteasome-mediated degradation. *Journal of cellular biochemistry.* 2010;110(4):926-34.
246. Van Cromphaut SJ, Dewerchin M, Hoenderop JG, Stockmans I, Van Herck E, Kato S, Bindels RJ, Collen D, Carmeliet P, Bouillon R, et al. Duodenal calcium absorption in vitamin D receptor-knockout mice: functional and molecular aspects. *Proc Natl Acad Sci U S A.* 2001;98(23):13324-9.
247. Portale AA, Halloran BP, and Morris RC, Jr. Physiologic regulation of the serum concentration of 1,25-dihydroxyvitamin D by phosphorus in normal men. *J Clin Invest.* 1989;83(5):1494-9.
248. Lips P. Vitamin D deficiency and secondary hyperparathyroidism in the elderly: consequences for bone loss and fractures and therapeutic implications. *Endocr Rev.* 2001;22(4):477-501.
249. Wang Y, Zhu J, and DeLuca HF. Where is the vitamin D receptor? *Arch Biochem Biophys.* 2012;523(1):123-33.
250. Holick MF. Vitamin D: A millenium perspective. *Journal of cellular biochemistry.* 2003;88(2):296-307.

251. Norman AW. Minireview: vitamin D receptor: new assignments for an already busy receptor. *Endocrinology.* 2006;147(12):5542-8.
252. Carlberg C. Genome-wide (over)view on the actions of vitamin D. *Front Physiol.* 2014;5(167.
253. C W. On the use and administration of cod-liver oil in pulmonary consumption. *London Journal of Medicine.* 1849;1:1–18(
254. Ginde AA, Mansbach JM, and Camargo CA, Jr. Association between serum 25-hydroxyvitamin D level and upper respiratory tract infection in the Third National Health and Nutrition Examination Survey. *Arch Intern Med.* 2009;169(4):384-90.
255. Laaksi I, Ruohola JP, Tuohimaa P, Auvinen A, Haataja R, Pihlajamaki H, and Ylikomi T. An association of serum vitamin D concentrations < 40 nmol/L with acute respiratory tract infection in young Finnish men. *Am J Clin Nutr.* 2007;86(3):714-7.
256. Cannell JJ, Vieth R, Umhau JC, Holick MF, Grant WB, Madronich S, Garland CF, and Giovannucci E. Epidemic influenza and vitamin D. *Epidemiol Infect.* 2006;134(6):1129-40.
257. Rodriguez M, Daniels B, Gunawardene S, and Robbins GK. High frequency of vitamin D deficiency in ambulatory HIV-Positive patients. *AIDS Res Hum Retroviruses.* 2009;25(1):9-14.
258. Munger KL, Levin LI, Hollis BW, Howard NS, and Ascherio A. Serum 25-hydroxyvitamin D levels and risk of multiple sclerosis. *JAMA.* 2006;296(23):2832-8.
259. Littorin B, Blom P, Scholin A, Arnqvist HJ, Blohme G, Bolinder J, Ekbom-Schnell A, Eriksson JW, Gudbjornsdottir S, Nystrom L, et al. Lower levels of plasma 25-hydroxyvitamin D among young adults at diagnosis of autoimmune type 1 diabetes compared with control subjects: results from the nationwide Diabetes Incidence Study in Sweden (DISS). *Diabetologia.* 2006;49(12):2847-52.
260. Merlino LA, Curtis J, Mikuls TR, Cerhan JR, Criswell LA, Saag KG, and Iowa Women's Health S. Vitamin D intake is inversely associated with rheumatoid arthritis: results from the Iowa Women's Health Study. *Arthritis Rheum.* 2004;50(1):72-7.
261. Kamen D, and Aranow C. Vitamin D in systemic lupus erythematosus. *Curr Opin Rheumatol.* 2008;20(5):532-7.
262. Bhalla AK, Amento EP, Clemens TL, Holick MF, and Krane SM. Specific high-affinity receptors for 1,25-dihydroxyvitamin D3 in human peripheral blood mononuclear cells: presence in monocytes and induction in T lymphocytes following activation. *J Clin Endocrinol Metab.* 1983;57(6):1308-10.

263. Liu PT, Stenger S, Li H, Wenzel L, Tan BH, Krutzik SR, Ochoa MT, Schauber J, Wu K, Meinken C, et al. Toll-like receptor triggering of a vitamin D-mediated human antimicrobial response. *Science.* 2006;311(5768):1770-3.
264. Camp RL, Dolled-Filhart M, and Rimm DL. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clin Cancer Res.* 2004;10(21):7252-9.
265. Wang TT, Tavera-Mendoza LE, Laperriere D, Libby E, MacLeod NB, Nagai Y, Bourdeau V, Konstorum A, Lallemant B, Zhang R, et al. Large-scale in silico and microarray-based identification of direct 1,25-dihydroxyvitamin D3 target genes. *Molecular endocrinology.* 2005;19(11):2685-95.
266. Bhalla AK, Amento EP, Serog B, and Glimcher LH. 1,25-Dihydroxyvitamin D3 inhibits antigen-induced T cell activation. *Journal of immunology.* 1984;133(4):1748-54.
267. Nunn JD, Katz DR, Barker S, Fraher LJ, Hewison M, Hendy GN, and O'Riordan JL. Regulation of human tonsillar T-cell proliferation by the active metabolite of vitamin D3. *Immunology.* 1986;59(4):479-84.
268. Barrat FJ, Cua DJ, Boonstra A, Richards DF, Crain C, Savelkoul HF, de Waal-Malefyt R, Coffman RL, Hawrylowicz CM, and O'Garra A. In vitro generation of interleukin 10-producing regulatory CD4(+) T cells is induced by immunosuppressive drugs and inhibited by T helper type 1 (Th1)- and Th2-inducing cytokines. *The Journal of experimental medicine.* 2002;195(5):603-16.
269. Heine G, Niesner U, Chang HD, Steinmeyer A, Zugel U, Zuberbier T, Radbruch A, and Worm M. 1,25-dihydroxyvitamin D(3) promotes IL-10 production in human B cells. *European journal of immunology.* 2008;38(8):2210-8.
270. Shirakawa AK, Nagakubo D, Hieshima K, Nakayama T, Jin Z, and Yoshie O. 1,25-dihydroxyvitamin D3 induces CCR10 expression in terminally differentiating human B cells. *Journal of immunology.* 2008;180(5):2786-95.
271. Autier P, Boniol M, Pizot C, and Mullie P. Vitamin D status and ill health--author's reply. *Lancet Diabetes Endocrinol.* 2014;2(4):275-6.
272. To SH. Statistics How To. https://www.statisticshowto.datasciencecentral.com/reverse-causality/.
273. Amer M, and 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(2):226-30.
274. Mondul AM, Weinstein SJ, Layne TM, and Albanes D. Vitamin D and Cancer Risk and Mortality: State of the Science, Gaps, and Challenges. *Epidemiol Rev.* 2017;39(1):28-48.

275. Gandini S, Boniol M, Haukka J, Byrnes G, Cox B, Sneyd MJ, Mullie P, and Autier P. Meta-analysis of observational studies of serum 25-hydroxyvitamin D levels and colorectal, breast and prostate cancer and colorectal adenoma. *Int J Cancer.* 2011;128(6):1414-24.
276. Lee JE, Li H, Chan AT, Hollis BW, Lee IM, Stampfer MJ, Wu K, Giovannucci E, and Ma J. Circulating levels of vitamin D and colon and rectal cancer: the Physicians' Health Study and a meta-analysis of prospective studies. *Cancer Prev Res (Phila).* 2011;4(5):735-43.
277. Ma Y, Zhang P, Wang F, Yang J, Liu Z, and Qin H. Association between vitamin D and risk of colorectal cancer: a systematic review of prospective studies. *J Clin Oncol.* 2011;29(28):3775-82.
278. Weinstein SJ, Yu K, Horst RL, Ashby J, Virtamo J, and Albanes D. Serum 25-hydroxyvitamin D and risks of colon and rectal cancer in Finnish men. *Am J Epidemiol.* 2011;173(5):499-508.
279. Chandler PD, Buring JE, Manson JE, Giovannucci EL, Moorthy MV, Zhang S, Lee IM, and Lin JH. Circulating Vitamin D Levels and Risk of Colorectal Cancer in Women. *Cancer Prev Res (Phila).* 2015;8(8):675-82.
280. Shao T, Klein P, and Grossbard ML. Vitamin D and breast cancer. *Oncologist.* 2012;17(1):36-45.
281. Yin L, Grandi N, Raum E, Haug U, Arndt V, and Brenner H. Meta-analysis: serum vitamin D and breast cancer risk. *Eur J Cancer.* 2010;46(12):2196-205.
282. Trump DL, and Aragon-Ching JB. Vitamin D in prostate cancer. *Asian J Androl.* 2018;20(3):244-52.
283. Zhang H, Zhang H, Wen X, Zhang Y, Wei X, and Liu T. Vitamin D Deficiency and Increased Risk of Bladder Carcinoma: A Meta-Analysis. *Cell Physiol Biochem.* 2015;37(5):1686-92.
284. Zhao Y, Chen C, Pan W, Gao M, He W, Mao R, Lin T, and Huang J. Comparative efficacy of vitamin D status in reducing the risk of bladder cancer: A systematic review and network meta-analysis. *Nutrition.* 2016;32(5):515-23.
285. Chen GC, Zhang ZL, Wan Z, Wang L, Weber P, Eggersdorfer M, Qin LQ, and Zhang W. Circulating 25-hydroxyvitamin D and risk of lung cancer: a dose-response meta-analysis. *Cancer Causes Control.* 2015;26(12):1719-28.
286. Zhang L, Wang S, Che X, and Li X. Vitamin D and lung cancer risk: a comprehensive review and meta-analysis. *Cell Physiol Biochem.* 2015;36(1):299-305.
287. Tang JY, Parimi N, Wu A, Boscardin WJ, Shikany JM, Chren MM, Cummings SR, Epstein EH, Jr., Bauer DC, and Osteoporotic Fractures in Men Study G.

Inverse association between serum 25(OH) vitamin D levels and non-melanoma skin cancer in elderly men. *Cancer Causes Control.* 2010;21(3):387-91.

288. Newton-Bishop JA, Beswick S, Randerson-Moor J, Chang YM, Affleck P, Elliott F, Chan M, Leake S, Karpavicius B, Haynes S, et al. Serum 25-hydroxyvitamin D3 levels are associated with breslow thickness at presentation and survival from melanoma. *J Clin Oncol.* 2009;27(32):5439-44.
289. Timerman D, McEnery-Stonelake M, Joyce CJ, Nambudiri VE, Hodi FS, Claus EB, Ibrahim N, and Lin JY. Vitamin D deficiency is associated with a worse prognosis in metastatic melanoma. *Oncotarget.* 2017;8(4):6873-82.
290. Gambichler T, Bindsteiner M, Hoxtermann S, and Kreuter A. Serum 25-hydroxyvitamin D serum levels in a large German cohort of patients with melanoma. *Br J Dermatol.* 2013;168(3):625-8.
291. Fang S, Sui D, Wang Y, Liu H, Chiang YJ, Ross MI, Gershenwald JE, Cormier JN, Royal RE, Lucci A, et al. Association of Vitamin D Levels With Outcome in Patients With Melanoma After Adjustment For C-Reactive Protein. *J Clin Oncol.* 2016;34(15):1741-7.
292. Wyatt C, Lucas RM, Hurst C, and Kimlin MG. Vitamin D deficiency at melanoma diagnosis is associated with higher Breslow thickness. *PLoS One.* 2015;10(5):e0126394.
293. Colston K, Colston MJ, and Feldman D. 1,25-dihydroxyvitamin D3 and malignant melanoma: the presence of receptors and inhibition of cell growth in culture. *Endocrinology.* 1981;108(3):1083-6.
294. Hager G, Formanek M, Gedlicka C, Thurnher D, Knerer B, and Kornfehl J. 1,25(OH)2 vitamin D3 induces elevated expression of the cell cycle-regulating genes P21 and P27 in squamous carcinoma cell lines of the head and neck. *Acta Otolaryngol.* 2001;121(1):103-9.
295. Koike M, Elstner E, Campbell MJ, Asou H, Uskokovic M, Tsuruoka N, and Koeffler HP. 19-nor-hexafluoride analogue of vitamin D3: a novel class of potent inhibitors of proliferation of human breast cell lines. *Cancer Res.* 1997;57(20):4545-50.
296. Kumagai T, O'Kelly J, Said JW, and Koeffler HP. Vitamin D2 analog 19-nor-1,25-dihydroxyvitamin D2: antitumour activity against leukemia, myeloma, and colon cancer cells. *J Natl Cancer Inst.* 2003;95(12):896-905.
297. Liu M, Lee MH, Cohen M, Bommakanti M, and Freedman LP. Transcriptional activation of the Cdk inhibitor p21 by vitamin D3 leads to the induced differentiation of the myelomonocytic cell line U937. *Genes & development.* 1996;10(2):142-53.

298. Saramaki A, Banwell CM, Campbell MJ, and Carlberg C. Regulation of the human p21(waf1/cip1) gene promoter via multiple binding sites for p53 and the vitamin D3 receptor. *Nucleic Acids Res.* 2006;34(2):543-54.

299. Yang X, Young LH, and Voigt JM. Expression of a vitamin D-regulated gene (VDUP-1) in untreated- and MNU-treated rat mammary tissue. *Breast Cancer Res Treat.* 1998;48(1):33-44.

300. Junn E, Han SH, Im JY, Yang Y, Cho EW, Um HD, Kim DK, Lee KW, Han PL, Rhee SG, et al. Vitamin D3 up-regulated protein 1 mediates oxidative stress via suppressing the thioredoxin function. *Journal of immunology.* 2000;164(12):6287-95.

301. Byrne BM, and Welsh J. Altered thioredoxin subcellular localization and redox status in MCF-7 cells following 1,25-dihydroxyvitamin D3 treatment. *J Steroid Biochem Mol Biol.* 2005;97(1-2):57-64.

302. Ricca C, Aillon A, Bergandi L, Alotto D, Castagnoli C, and Silvagno F. Vitamin D Receptor Is Necessary for Mitochondrial Function and Cell Health. *Int J Mol Sci.* 2018;19(6).

303. Pendas-Franco N, Garcia JM, Pena C, Valle N, Palmer HG, Heinaniemi M, Carlberg C, Jimenez B, Bonilla F, Munoz A, et al. DICKKOPF-4 is induced by TCF/beta-catenin and upregulated in human colon cancer, promotes tumour cell invasion and angiogenesis and is repressed by 1alpha,25-dihydroxyvitamin D3. *Oncogene.* 2008;27(32):4467-77.

304. Larriba MJ, Valle N, Palmer HG, Ordonez-Moran P, Alvarez-Diaz S, Becker KF, Gamallo C, de Herreros AG, Gonzalez-Sancho JM, and Munoz A. The inhibition of Wnt/beta-catenin signalling by 1alpha,25-dihydroxyvitamin D3 is abrogated by Snail1 in human colon cancer cells. *Endocr Relat Cancer.* 2007;14(1):141-51.

305. Palmer HG, Larriba MJ, Garcia JM, Ordonez-Moran P, Pena C, Peiro S, Puig I, Rodriguez R, de la Fuente R, Bernad A, et al. The transcription factor SNAIL represses vitamin D receptor expression and responsiveness in human colon cancer. *Nat Med.* 2004;10(9):917-9.

306. Stambolsky P, Tabach Y, Fontemaggi G, Weisz L, Maor-Aloni R, Siegfried Z, Shiff I, Kogan I, Shay M, Kalo E, et al. Modulation of the vitamin D3 response by cancer-associated mutant p53. *Cancer Cell.* 2010;17(3):273-85.

307. Jones PA, and Baylin SB. The epigenomics of cancer. *Cell.* 2007;128(4):683-92.

308. Lopes N, Carvalho J, Duraes C, Sousa B, Gomes M, Costa JL, Oliveira C, Paredes J, and Schmitt F. 1Alpha,25-dihydroxyvitamin D3 induces de novo E-

cadherin expression in triple-negative breast cancer cells by CDH1-promoter demethylation. *Anticancer Res.* 2012;32(1):249-57.

309. Reichrath J, Rech M, Moeini M, Meese E, Tilgen W, and 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(1):48-55.
310. Slominski AT, Janjetovic Z, Kim TK, Wright AC, Grese LN, Riney SJ, Nguyen MN, and Tuckey RC. Novel vitamin D hydroxyderivatives inhibit melanoma growth and show differential effects on normal melanocytes. *Anticancer Res.* 2012;32(9):3733-42.
311. Randerson-Moor JA, Taylor JC, Elliott F, Chang YM, Beswick S, Kukalizch K, Affleck P, Leake S, Haynes S, Karpavicius B, et al. Vitamin D receptor gene polymorphisms, serum 25-hydroxyvitamin D levels, and melanoma: UK case-control comparisons and a meta-analysis of published VDR data. *Eur J Cancer.* 2009;45(18):3271-81.
312. Gershenwald JE, Scolyer RA, Hess KR, Sondak VK, Long GV, Ross MI, Lazar AJ, Faries MB, Kirkwood JM, McArthur GA, et al. Melanoma staging: Evidence-based changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA Cancer J Clin.* 2017;67(6):472-92.
313. Balch CM, Gershenwald JE, Soong SJ, Thompson JF, Atkins MB, Byrd DR, Buzaid AC, Cochran AJ, Coit DG, Ding S, et al. Final version of 2009 AJCC melanoma staging and classification. *J Clin Oncol.* 2009;27(36):6199-206.
314. Du P, Kibbe WA, and Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics.* 2008;24(13):1547-8.
315. Lauss M, Visne I, Kriegner A, Ringner M, Jonsson G, and Hoglund M. Monitoring of technical variation in quantitative high-throughput datasets. *Cancer Inform.* 2013;12(193-201.
316. Nsengimana J, Laye J, Filia A, O'Shea S, Muralidhar S, Pozniak J, Droop A, Chan M, Walker C, Parkinson L, et al. beta-Catenin-mediated immune evasion pathway frequently operates in primary cutaneous melanomas. *J Clin Invest.* 2018;128(5):2048-63.
317. Scheinin I, Sie D, Bengtsson H, van de Wiel MA, Olshen AB, van Thuijl HF, van Essen HF, Eijk PP, Rustenburg F, Meijer GA, et al. DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. *Genome Res.* 2014;24(12):2022-32.
318. Venkatraman ES, and Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics.* 2007;23(6):657-63.

319. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, and 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(4):R41.
320. StataCorp. College Station, TX: StataCorp LP.; 2015.
321. Wu G, and Stein L. A network module-based method for identifying cancer prognostic signatures. *Genome Biol.* 2012;13(12):R112.
322. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, and Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498-504.
323. Wu G, Feng X, and Stein L. A human functional protein interaction network and its application to cancer data analysis. *Genome Biol.* 2010;11(5):R53.
324. Jones G, Strugnell SA, and DeLuca HF. Current understanding of the molecular actions of vitamin D. *Physiological reviews.* 1998;78(4):1193-231.
325. Carlberg C, Bendik I, Wyss A, Meier E, Sturzenbecker LJ, Grippo JF, and Hunziker W. Two nuclear signalling pathways for vitamin D. *Nature.* 1993;361(6413):657-60.
326. Murayama A, Kim MS, Yanagisawa J, Takeyama K, and Kato S. Transrepression by a liganded nuclear receptor via a bHLH activator through co-regulator switching. *EMBO J.* 2004;23(7):1598-608.
327. Fujiki R, Kim MS, Sasaki Y, Yoshimura K, Kitagawa H, and Kato S. Ligand-induced transrepression by VDR through association of WSTF with acetylated histones. *EMBO J.* 2005;24(22):3881-94.
328. Chen KS, and DeLuca HF. Cloning of the human 1 alpha,25-dihydroxyvitamin D-3 24-hydroxylase gene promoter and identification of two vitamin D-responsive elements. *Biochim Biophys Acta.* 1995;1263(1):1-9.
329. Wali RK, Baum CL, Sitrin MD, and Brasitus TA. 1,25(OH)2 vitamin D3 stimulates membrane phosphoinositide turnover, activates protein kinase C, and increases cytosolic calcium in rat colonic epithelium. *J Clin Invest.* 1990;85(4):1296-303.
330. Morelli S, Buitrago C, Boland R, and de Boland AR. The stimulation of MAP kinase by 1,25(OH)(2)-vitamin D(3) in skeletal muscle cells is mediated by protein kinase C and calcium. *Mol Cell Endocrinol.* 2001;173(1-2):41-52.
331. Hsieh JC, Jurutka PW, Galligan MA, Terpening CM, Haussler CA, Samuels DS, Shimizu Y, Shimizu N, and Haussler MR. Human vitamin D receptor is selectively phosphorylated by protein kinase C on serine 51, a residue crucial to its trans-activation function. *Proc Natl Acad Sci U S A.* 1991;88(20):9315-9.

332. Haussler MR, Whitfield GK, Haussler CA, Hsieh JC, Thompson PD, Selznick SH, Dominguez CE, and Jurutka PW. The nuclear vitamin D receptor: biological and molecular regulatory properties revealed. *J Bone Miner Res.* 1998;13(3):325-49.
333. Wikvall K. Cytochrome P450 enzymes in the bioactivation of vitamin D to its hormonal form (review). *Int J Mol Med.* 2001;7(2):201-9.
334. Carmeliet G, Dermauw V, and Bouillon R. Vitamin D signalling in calcium and bone homeostasis: a delicate balance. *Best practice & research Clinical endocrinology & metabolism.* 2015;29(4):621-31.
335. Peleg S, Ismail A, Uskokovic MR, and Avnur Z. Evidence for tissue- and cell-type selective activation of the vitamin D receptor by Ro-26-9228, a noncalcemic analog of vitamin D3. *Journal of cellular biochemistry.* 2003;88(2):267-73.
336. Bikle DD. Vitamin D regulated keratinocyte differentiation. *Journal of cellular biochemistry.* 2004;92(3):436-44.
337. Gombart AF, Borregaard N, and Koeffler HP. Human cathelicidin antimicrobial peptide (CAMP) gene is a direct target of the vitamin D receptor and is strongly up-regulated in myeloid cells by 1,25-dihydroxyvitamin D3. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology.* 2005;19(9):1067-77.
338. Schauber J, Dorschner RA, Coda AB, Buchau AS, Liu PT, Kiken D, Helfrich YR, Kang S, Elalieh HZ, Steinmeyer A, et al. Injury enhances TLR2 function and antimicrobial peptide expression through a vitamin D-dependent mechanism. *J Clin Invest.* 2007;117(3):803-11.
339. Matsumoto K, Hashimoto K, Nishida Y, Hashiro M, and Yoshikawa K. Growth-inhibitory effects of 1,25-dihydroxyvitamin D3 on normal human keratinocytes cultured in serum-free medium. *Biochem Biophys Res Commun.* 1990;166(2):916-23.
340. Garland CF, Garland FC, Gorham ED, Lipkin M, Newmark H, Mohr SB, and Holick MF. The role of vitamin D in cancer prevention. *Am J Public Health.* 2006;96(2):252-61.
341. Palacios C, and Gonzalez L. Is vitamin D deficiency a major global public health problem? *J Steroid Biochem Mol Biol.* 2014;144 Pt A(138-45.
342. Baggerly CA, Cuomo RE, French CB, Garland CF, Gorham ED, Grant WB, Heaney RP, Holick MF, Hollis BW, McDonnell SL, et al. Sunlight and Vitamin D: Necessary for Public Health. *J Am Coll Nutr.* 2015;34(4):359-65.
343. Huang SJ, Wang XH, Liu ZD, Cao WL, Han Y, Ma AG, and Xu SF. Vitamin D deficiency and the risk of tuberculosis: a meta-analysis. *Drug Des Devel Ther.* 2017;11(91-102.

344. Kamen DL, Cooper GS, Bouali H, Shaftman SR, Hollis BW, and Gilkeson GS. Vitamin D deficiency in systemic lupus erythematosus. *Autoimmun Rev.* 2006;5(2):114-7.
345. Hypponen E, Laara E, Reunanen A, Jarvelin MR, and Virtanen SM. Intake of vitamin D and risk of type 1 diabetes: a birth-cohort study. *Lancet.* 2001;358(9292):1500-3.
346. Laverny G, and Metzger D. *Vitamin D (Fourth Edition)*. 2018.
347. Hsieh JC, Sisk JM, Jurutka PW, Haussler CA, Slater SA, Haussler MR, and Thompson CC. Physical and functional interaction between the vitamin D receptor and hairless corepressor, two proteins required for hair cycling. *The Journal of biological chemistry.* 2003;278(40):38665-74.
348. Skorija K, Cox M, Sisk JM, Dowd DR, MacDonald PN, Thompson CC, and Demay MB. Ligand-independent actions of the vitamin D receptor maintain hair follicle homeostasis. *Molecular endocrinology.* 2005;19(4):855-62.
349. Trivedi T, Zheng Y, Fournier PGJ, Murthy S, John S, Schillo S, Dunstan CR, Mohammad KS, Zhou H, Seibel MJ, et al. The vitamin D receptor is involved in the regulation of human breast cancer cell growth via a ligand-independent function in cytoplasm. *Oncotarget.* 2017;8(16):26687-701.
350. Guzey M, Luo J, and Getzenberg RH. Vitamin D3 modulated gene expression patterns in human primary normal and cancer prostate cells. *Journal of cellular biochemistry.* 2004;93(2):271-85.
351. Swami S, Raghavachari N, Muller UR, Bao YP, and Feldman D. Vitamin D growth inhibition of breast cancer cells: gene expression patterns assessed by cDNA microarray. *Breast Cancer Res Treat.* 2003;80(1):49-62.
352. Suzuki T, Tazoe H, Taguchi K, Koyama Y, Ichikawa H, Hayakawa S, Munakata H, and Isemura M. DNA microarray analysis of changes in gene expression induced by 1,25-dihydroxyvitamin D3 in human promyelocytic leukemia HL-60 cells. *Biomed Res.* 2006;27(3):99-109.
353. Wood RJ, Tchack L, Angelo G, Pratt RE, and Sonna LA. DNA microarray analysis of vitamin D-induced gene expression in a human colon carcinoma cell line. *Physiol Genomics.* 2004;17(2):122-9.
354. Jensen SS, Madsen MW, Lukas J, Binderup L, and Bartek J. Inhibitory effects of 1alpha,25-dihydroxyvitamin D(3) on the G(1)-S phase-controlling machinery. *Molecular endocrinology.* 2001;15(8):1370-80.
355. Yanagisawa J, Yanagi Y, Masuhiro Y, Suzawa M, Watanabe M, Kashiwagi K, Toriyabe T, Kawabata M, Miyazono K, and Kato S. Convergence of transforming growth factor-beta and vitamin D signalling pathways on SMAD transcriptional coactivators. *Science.* 1999;283(5406):1317-21.

356. Aguilera O, Pena C, Garcia JM, Larriba MJ, Ordonez-Moran P, Navarro D, Barbachano A, Lopez de Silanes I, Ballestar E, Fraga MF, et al. The Wnt antagonist DICKKOPF-1 gene is induced by 1alpha,25-dihydroxyvitamin D3 associated to the differentiation of human colon cancer cells. *Carcinogenesis.* 2007;28(9):1877-84.

357. Tong WM, Kallay E, Hofer H, Hulla W, Manhardt T, Peterlik M, and Cross HS. Growth regulation of human colon cancer cells by epidermal growth factor and 1,25-dihydroxyvitamin D3 is mediated by mutual modulation of receptor expression. *Eur J Cancer.* 1998;34(13):2119-25.

358. Vink-van Wijngaarden T, Pols HA, Buurman CJ, Birkenhager JC, and van Leeuwen JP. Inhibition of insulin- and insulin-like growth factor-I-stimulated growth of human breast cancer cells by 1,25-dihydroxyvitamin D3 and the vitamin D3 analogue EB1089. *Eur J Cancer.* 1996;32A(5):842-8.

359. Kallay E, Pietschmann P, Toyokuni S, Bajna E, Hahn P, Mazzucco K, Bieglmayer C, Kato S, and Cross HS. Characterization of a vitamin D receptor knockout mouse as a model of colorectal hyperproliferation and DNA damage. *Carcinogenesis.* 2001;22(9):1429-35.

360. Majewski S, Skopinska M, Marczak M, Szmurlo A, Bollag W, and Jablonska S. Vitamin D3 is a potent inhibitor of tumour cell-induced angiogenesis. *J Investig Dermatol Symp Proc.* 1996;1(1):97-101.

361. Flynn G, Chung I, Yu WD, Romano M, Modzelewski RA, Johnson CS, and Trump DL. Calcitriol (1,25-dihydroxycholecalciferol) selectively inhibits proliferation of freshly isolated tumour-derived endothelial cells and induces apoptosis. *Oncology.* 2006;70(6):447-57.

362. Nakagawa K, Kawaura A, Kato S, Takeda E, and Okano T. Metastatic growth of lung cancer cells is extremely reduced in Vitamin D receptor knockout mice. *J Steroid Biochem Mol Biol.* 2004;89-90(1-5):545-7.

363. Gonzalez-Sancho JM, Alvarez-Dolado M, and Munoz A. 1,25-Dihydroxyvitamin D3 inhibits tenascin-C expression in mammary epithelial cells. *FEBS letters.* 1998;426(2):225-8.

364. Spina CS, Tangpricha V, Uskokovic M, Adorinic L, Maehr H, and Holick MF. Vitamin D and cancer. *Anticancer Res.* 2006;26(4A):2515-24.

365. Slominski AT, Kim TK, Janjetovic Z, Tuckey RC, Bieniek R, Yue J, Li W, Chen J, Nguyen MN, Tang EK, et al. 20-Hydroxyvitamin D2 is a noncalcemic analog of vitamin D with potent antiproliferative and prodifferentiation activities in normal and malignant cells. *American journal of physiology Cell physiology.* 2011;300(3):C526-41.

366. Janjetovic Z, Brozyna AA, Tuckey RC, Kim TK, Nguyen MN, Jozwicki W, Pfeffer SR, Pfeffer LM, and Slominski AT. High basal NF-kappaB activity in nonpigmented melanoma cells is associated with an enhanced sensitivity to vitamin D3 derivatives. *Br J Cancer.* 2011;105(12):1874-84.
367. Skobowiat C, Oak AS, Kim TK, Yang CH, Pfeffer LM, Tuckey RC, and Slominski AT. Noncalcemic 20-hydroxyvitamin D3 inhibits human melanoma growth in in vitro and in vivo models. *Oncotarget.* 2017;8(6):9823-34.
368. Brozyna AA, Jozwicki W, and Slominski AT. Decreased VDR expression in cutaneous melanomas as marker of tumour progression: new data and analyses. *Anticancer Res.* 2014;34(6):2735-43.
369. Lewis JG, and Elder PA. Serum 25-OH vitamin D2 and D3 are stable under exaggerated conditions. *Clin Chem.* 2008;54(11):1931-2.
370. Lideikaite A, Mozuraitiene J, and Letautiene S. Analysis of prognostic factors for melanoma patients. *Acta Med Litu.* 2017;24(1):25-34.
371. Bikle DD. Vitamin D and the skin: Physiology and pathophysiology. *Rev Endocr Metab Disord.* 2012;13(1):3-19.
372. Nsengimana J, Laye J, Filia A, Walker C, Jewell R, Van den Oord JJ, Wolter P, Patel P, Sucker A, Schadendorf D, 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(13):11683-93.
373. Angelova M, Charoentong P, Hackl H, Fischer ML, Snajder R, Krogsdam AM, Waldner MJ, Bindea G, Mlecnik B, Galon J, et al. Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumour escape mechanisms and novel targets for immunotherapy. *Genome Biol.* 2015;16(64.
374. Tuoresmaki P, Vaisanen S, Neme A, Heikkinen S, and Carlberg C. Patterns of genome-wide VDR locations. *PLoS One.* 2014;9(4):e96105.
375. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, and Bejerano G. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol.* 2010;28(5):495-501.
376. Najita JS, Swetter SM, Geller AC, Gershenwald JE, Zelen M, and Lee SJ. Sex Differences in Age at Primary Melanoma Diagnosis in a Population-Based Analysis (US Surveillance, Epidemiology, and End Results, 2005-2011). *J Invest Dermatol.* 2016;136(9):1894-7.
377. Garbe C, Buttner P, Bertz J, Burg G, d'Hoedt B, Drepper H, Guggenmoos-Holzmann I, Lechner W, Lippold A, Orfanos CE, et al. Primary cutaneous melanoma. Identification of prognostic groups and estimation of individual prognosis for 5093 patients. *Cancer.* 1995;75(10):2484-91.

378. SACN TSACoN. *SACN recommendations on vitamin D*. 2016.
379. Mocellin S, and Nitti D. Vitamin D receptor polymorphisms and the risk of cutaneous melanoma: a systematic review and meta-analysis. *Cancer.* 2008;113(9):2398-407.
380. Sandilands A, Sutherland C, Irvine AD, and McLean WH. Filaggrin in the frontline: role in skin barrier function and disease. *J Cell Sci.* 2009;122(Pt 9):1285-94.
381. Lauss M, Nsengimana J, Staaf J, Newton-Bishop J, and Jonsson G. Consensus of Melanoma Gene Expression Subtypes Converges on Biological Entities. *J Invest Dermatol.* 2016;136(12):2502-5.
382. Heikkinen S, Vaisanen S, Pehkonen P, Seuter S, Benes V, and Carlberg C. Nuclear hormone 1alpha,25-dihydroxyvitamin D3 elicits a genome-wide shift in the locations of VDR chromatin occupancy. *Nucleic Acids Res.* 2011;39(21):9181-93.
383. Ramagopalan SV, Heger A, Berlanga AJ, Maugeri NJ, Lincoln MR, Burrell A, Handunnetthi L, Handel AE, Disanto G, Orton SM, et al. A ChIP-seq defined genome-wide map of vitamin D receptor binding: associations with disease and evolution. *Genome Res.* 2010;20(10):1352-60.
384. Ding N, Yu RT, Subramaniam N, Sherman MH, Wilson C, Rao R, Leblanc M, Coulter S, He M, Scott C, et al. A vitamin D receptor/SMAD genomic circuit gates hepatic fibrotic response. *Cell.* 2013;153(3):601-13.
385. Meyer MB, Goetsch PD, and Pike JW. VDR/RXR and TCF4/beta-catenin cistromes in colonic cells of colorectal tumour origin: impact on c-FOS and c-MYC gene expression. *Molecular endocrinology.* 2012;26(1):37-51.
386. Daily K, Patel VR, Rigor P, Xie X, and Baldi P. MotifMap: integrative genome-wide maps of regulatory motif sites for model species. *BMC Bioinformatics.* 2011;12(495.
387. Xie X, Rigor P, and Baldi P. MotifMap: a human genome-wide map of candidate regulatory motif sites. *Bioinformatics.* 2009;25(2):167-74.
388. Cancer Genome Atlas N. Genomic Classification of Cutaneous Melanoma. *Cell.* 2015;161(7):1681-96.
389. Jonsson G, Busch C, Knappskog S, Geisler J, Miletic H, Ringner M, Lillehaug JR, Borg A, and Lonning PE. Gene expression profiling-based identification of molecular subtypes in stage IV melanomas with different clinical outcome. *Clin Cancer Res.* 2010;16(13):3356-67.
390. Bennesch MA, and Picard D. Minireview: Tipping the balance: ligand-independent activation of steroid receptors. *Molecular endocrinology.* 2015;29(3):349-63.

391. Lagishetty V, Liu NQ, and Hewison M. Vitamin D metabolism and innate immunity. *Mol Cell Endocrinol.* 2011;347(1-2):97-105.
392. Kearns MD, and Tangpricha V. The role of vitamin D in tuberculosis. *J Clin Transl Endocrinol.* 2014;1(4):167-9.
393. Chen L, Eapen MS, and Zosky GR. Vitamin D both facilitates and attenuates the cellular response to lipopolysaccharide. *Sci Rep.* 2017;7(45172.
394. Akbas EM, Gungor A, Ozcicek A, Akbas N, Askin S, and Polat M. Vitamin D and inflammation: evaluation with neutrophil-to-lymphocyte ratio and platelet-to-lymphocyte ratio. *Arch Med Sci.* 2016;12(4):721-7.
395. Souto Filho JTD, de Andrade AS, Ribeiro FM, Alves PAS, and Simonini VRF. Impact of vitamin D deficiency on increased blood eosinophil counts. *Hematol Oncol Stem Cell Ther.* 2018;11(1):25-9.
396. Long MD, and Campbell MJ. Pan-cancer analyses of the nuclear receptor superfamily. *Nucl Receptor Res.* 2015;2(
397. Mallbris L, Edstrom DW, Sundblad L, Granath F, and Stahle M. UVB upregulates the antimicrobial protein hCAP18 mRNA in human skin. *J Invest Dermatol.* 2005;125(5):1072-4.
398. Newton-Bishop JA, Chang YM, Elliott F, Chan M, Leake S, Karpavicius B, Haynes S, Fitzgibbon E, Kukalizch K, Randerson-Moor J, 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(5):732-41.
399. Saccone D, Asani F, and Bornman L. Regulation of the vitamin D receptor gene by environment, genetics and epigenetics. *Gene.* 2015;561(2):171-80.
400. Essa S, Reichrath S, Mahlknecht U, Montenarh M, Vogt T, and Reichrath J. Signature of VDR miRNAs and epigenetic modulation of vitamin D signalling in melanoma cell lines. *Anticancer Res.* 2012;32(1):383-9.
401. SnapGene.
402. Li Z, Michael IP, Zhou D, Nagy A, and Rini JM. Simple piggyBac transposon-based mammalian cell expression system for inducible protein production. *Proc Natl Acad Sci U S A.* 2013;110(13):5004-9.
403. van der Weyden L, Arends MJ, Campbell AD, Bald T, Wardle-Jones H, Griggs N, Velasco-Herrera MD, Tuting T, Sansom OJ, Karp NA, et al. Genome-wide in vivo screen identifies novel host regulators of metastatic colonization. *Nature.* 2017;541(7636):233-6.
404. Speak AO, Swiatkowska A, Karp NA, Arends MJ, Adams DJ, and van der Weyden L. A high-throughput in vivo screening method in the mouse for

identifying regulators of metastatic colonization. *Nat Protoc.* 2017;12(12):2465-77.

405. Zhong Z, Ethen NJ, and Williams BO. WNT signalling in bone development and homeostasis. *Wiley Interdiscip Rev Dev Biol.* 2014;3(6):489-500.
406. Cianferotti L, and Demay MB. VDR-mediated inhibition of DKK1 and SFRP2 suppresses adipogenic differentiation of murine bone marrow stromal cells. *Journal of cellular biochemistry.* 2007;101(1):80-8.
407. Katoh M. WNT/PCP signalling pathway and human cancer (review). *Oncol Rep.* 2005;14(6):1583-8.
408. Weeraratna AT, Becker D, Carr KM, Duray PH, Rosenblatt KP, Yang S, Chen Y, Bittner M, Strausberg RL, Riggins GJ, et al. Generation and analysis of melanoma SAGE libraries: SAGE advice on the melanoma transcriptome. *Oncogene.* 2004;23(12):2264-74.
409. Aviles-Izquierdo JA, and Lazaro-Ochaita P. Histological ulceration as a prognostic factor in cutaneous melanoma: a study of 423 cases in Spain. *Clin Transl Oncol.* 2012;14(3):237-40.
410. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, Meyer L, Gress DM, Byrd DR, and Winchester DP. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA Cancer J Clin.* 2017;67(2):93-9.
411. Spatz A, Cook MG, Elder DE, Piepkorn M, Ruiter DJ, and Barnhill RL. Interobserver reproducibility of ulceration assessment in primary cutaneous melanomas. *Eur J Cancer.* 2003;39(13):1861-5.
412. Hantschke M, Bastian BC, and LeBoit PE. Consumption of the epidermis: a diagnostic criterion for the differential diagnosis of melanoma and Spitz nevus. *Am J Surg Pathol.* 2004;28(12):1621-5.
413. Braun-Falco M, Friedrichson E, and Ring J. Subepidermal cleft formation as a diagnostic marker for cutaneous malignant melanoma. *Hum Pathol.* 2005;36(4):412-5.
414. Bonnelykke-Behrndtz ML, Schmidt H, Christensen IJ, Damsgaard TE, Moller HJ, Bastholt L, Norgaard PH, and Steiniche T. Prognostic stratification of ulcerated melanoma: not only the extent matters. *American journal of clinical pathology.* 2014;142(6):845-56.
415. In 't Hout FE, Haydu LE, Murali R, Bonenkamp JJ, Thompson JF, and Scolyer RA. Prognostic importance of the extent of ulceration in patients with clinically localized cutaneous melanoma. *Ann Surg.* 2012;255(6):1165-70.

416. Storr SJ, Safuan S, Mitra A, Elliott F, Walker C, Vasko MJ, Ho B, Cook M, Mohammed RA, Patel PM, et al. Objective assessment of blood and lymphatic vessel invasion and association with macrophage infiltration in cutaneous melanoma. *Mod Pathol.* 2012;25(4):493-504.
417. Jewell R, Elliott F, Laye J, Nsengimana J, Davies J, Walker C, Conway C, Mitra A, Harland M, Cook MG, et al. The clinicopathological and gene expression patterns associated with ulceration of primary melanoma. *Pigment Cell Melanoma Res.* 2015;28(1):94-104.
418. von Schuckmann LA, Smith D, Hughes MCB, Malt M, van der Pols JC, Khosrotehrani K, Smithers BM, and Green AC. Associations of Statins and Diabetes with Diagnosis of Ulcerated Cutaneous Melanoma. *J Invest Dermatol.* 2017;137(12):2599-605.
419. Eggermont AM, Suciu S, Testori A, Kruit WH, Marsden J, Punt CJ, Santinami M, Sales F, Schadendorf D, Patel P, et al. Ulceration and stage are predictive of interferon efficacy in melanoma: results of the phase III adjuvant trials EORTC 18952 and EORTC 18991. *Eur J Cancer.* 2012;48(2):218-25.
420. Koelblinger P, Emberger M, Drach M, Cheng PF, Lang R, Levesque MP, Bauer JW, and Dummer R. Increased tumour cell PD-L1 expression, macrophage and dendritic cell infiltration characterise the tumour microenvironment of ulcerated primary melanomas. *J Eur Acad Dermatol Venereol.* 2018.
421. Rakosy Z, Ecsedi S, Toth R, Vizkeleti L, Hernandez-Vargas H, Lazar V, Emri G, Szatmari I, Herceg Z, Adany R, et al. Integrative genomics identifies gene signature associated with melanoma ulceration. *PLoS One.* 2013;8(1):e54958.
422. Olmeda D, Cerezo-Wallis D, Riveiro-Falkenbach E, Pennacchi PC, Contreras-Alcalde M, Ibarz N, Cifdaloz M, Catena X, Calvo TG, Canon E, et al. Whole-body imaging of lymphovascular niches identifies pre-metastatic roles of midkine. *Nature.* 2017;546(7660):676-80.
423. Bhardwaj V. *Science behind Non-Specific Science; For Molecular Biologist & Bio-technologist.* Notion press; 2014.
424. Schena M, Shalon D, Davis RW, and Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science.* 1995;270(5235):467-70.
425. Chibon F. Cancer gene expression signatures - the rise and fall? *Eur J Cancer.* 2013;49(8):2000-9.
426. Simon R, Radmacher MD, Dobbin K, and McShane LM. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *J Natl Cancer Inst.* 2003;95(1):14-8.

427. Kim IJ, Kang HC, and Park JG. Microarray applications in cancer research. *Cancer Res Treat.* 2004;36(4):207-13.
428. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science.* 1999;286(5439):531-7.
429. Evers DL, Fowler CB, Cunningham BR, Mason JT, and O'Leary TJ. The effect of formaldehyde fixation on RNA: optimization of formaldehyde adduct removal. *J Mol Diagn.* 2011;13(3):282-8.
430. Fan JB, Yeakley JM, Bibikova M, Chudin E, Wickham E, Chen J, Doucet D, Rigault P, Zhang B, Shen R, et al. A versatile assay for high-throughput gene expression profiling on universal array matrices. *Genome Res.* 2004;14(5):878-85.
431. Weinstein SJ, Mondul AM, Yu K, Layne TM, Abnet CC, Freedman ND, Stolzenberg-Solomon RZ, Lim U, Gail MH, and Albanes D. Circulating 25-hydroxyvitamin D up to 3 decades prior to diagnosis in relation to overall and organ-specific cancer survival. *Eur J Epidemiol.* 2018;33(11):1087-99.