Biomarkers of prognosis and response to vitamin D in cutaneous melanoma

Anastasia Filia

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

The University of Leeds
School of Medicine

September 2013

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.

Chapter 6 includes work from the publication: FILIA, A., ELLIOTT, F., WIND, T., FIELD, S., DAVIES, J., KUKALIZCH, K., RANDERSON-MOOR, J., HARLAND, M., BISHOP, D. T., BANKS, R. E. & NEWTON-BISHOP, J. A. 2013. Plasma osteopontin concentrations in patients with cutaneous melanoma. *Oncol Rep,* 30**,** 1575-80. I have performed the statistical analysis and prepared the manuscript for publication with the help and advice of Faye Elliott. Tobias Wind has performed the ELISA experiments. Sinead Field has helped me stage the patients based on the pathology reports. John Davies has performed sample selection and Juliette Randerson-Moor, Mark Harland and Kairen Kukalizch have processed the plasma samples. This work was supervised by Julia Newton-Bishop, Tim Bishop and Rosamonde Banks, who contributed to the manuscript preparation.

“For the things we have to learn before we can do, we learn by doing.”

Aristotle (384 – 322 BC)

Acknowledgements

This research has been carried out by a team which has included Christy Walker, Clarissa Nolan, Susan Leake, Birute Karpavicius, Tricia Mack, Paul King, Sue Haynes, Elaine Fitzgibbon, Kate Gamble, Saila Waseem, Sandra Tovey, Christy Walker and Paul Affleck, Rosalyn Jewell, Samira lobo, Filomena Esteves, Jonathan Laye, Minttu Polso, Mark Harland, Juliette Randerson-Moor, Kairen Kukalizch, Faye Elliott, Jeremie Nsengimana, Joanne Gascoyne, John Davies, Sinead Field and May Chan from the Section of Epidemiology and Biostatistics. From outside the Section, team members include Phil Chambers, Helen Snowden and Claire Taylor from the Genomics Facility, Sally Harrison and Christopher Watson from the Next Generation Sequencing facility, Alastair Droop and Helene Thygesen from the Bioinformatics and Biostatistics Team, Rosamonde Banks and Tobias Wind from the Clinical and Biomedical Proteomics Group, Mike Shires from the Histology Facility all based in the Cancer Research UK Centre, Leeds Institute of Cancer and Pathology. External collaborators were Jorg Reichrath, Heike Palm, Alexandra Stark from the Department of Dermatology, Saarland University, Homburg, Germany and finally the service providers ServiceXS from Leiden, Netherlands, Gen-probe, Manchester, Source BioScience, Nottingham and the Patterson Institute at the University of Manchester.

My own contributions, fully and explicitly indicated in the thesis, have been:

- Sampling of tumour tissue and extraction of nucleic acids from both tumours and cell lines for work described in Chapters 2, 3, 4 and 5.
- Embedding of cells in agar described in Chapter 2.
- Performing cell culture work described in Chapters 2 and 5.
- Library preparation for NGS studies described in Chapters 2, 3 and 4.
- Planning and undertaking methodological studies to assess new technologies described in Chapters 2 and 3.
- Performing statistical analysis between osteopontin protein levels and clinico-pathological data (Chapter 6), between measures of sample characteristics (such as DNA yield) and tumour characteristics (Chapter 4). Performing statistical analysis of the data generated from the crystal violet assay and the data generated to predict VDR target genes (Chapter 5).
- Pre-processing of the HT WG-DASL gene expression data.

- Performing pathway analysis using MetaCore™ pathway database and data analysis tool in Chapters 4 and 5.

The other members of the group and their contributions have been as follows:

- May Chan, Clarissa Nolan, Susan Leake, Birute Karpavicius, Tricia Mack, Paul King, Sue Haynes, Elaine Fitzgibbon, Kate Gamble, Saila Waseem, Sandra Tovey, Christy Walker and Paul Affleck have collected and managed data for the Leeds Melanoma Cohort study. Sandra Tovey has traced tissue samples from the Leeds Melanoma Cohort (Chapters 3 and 4). Christy Walker has also assisted in project-management of the study described in Chapter 4 and traced samples from the Chemotherapy study;
- Jonathan Laye, Filomena Esteves and Minttu Polso have sectioned and sampled tumour tissues. Jonathan Laye and Minttu Polso have also extracted nucleic acids from tumour samples. Jonathan Laye has provided constant support and advice during my PhD. He has also helped to trace samples eligible for the study described in Chapter 4 that had been used in various other assays;
- Rosalyn Jewell has extracted nucleic acids from samples in studies described in Chapters 3 and 4. Rosalyn has also helped me get started in the lab;
- Sinead Field and Joanne Gascoyne have staged the patients in the Leeds Melanoma Cohort based on the pathology reports;
- Rosamonde Banks has provided advice for the study design and data analysis described in Chapter 6;
- Tobias Wind performed the ELISA assays and analysed the data from the validation of this assay;
- Juliette Randerson-Moor, Mark Harland and Kairen Kukalizch have processed plasma samples from the Leeds Melanoma Cohort study. Juliette Randerson-Moor has helped me get started in the lab and provided great support during my PhD;
- John Davies has performed sample selection for the studies presented in Chapters 4 and 6;
- Faye Elliott and Jeremie Nsengimana have provided statistical advice. Jeremie Nsengimana has helped in HT WG-DASL gene expression data pre-processing;

- Sally Harrison and Christopher Watson have performed quality control of pooled libraries and next generation sequencing experiments;
- Phil Chambers performed pyrosequencing analyses;
- Claire Taylor has performed the cell line authentication experiments;
- Mike Shires has embedded cells in paraffin-wax blocks;
- Helen Snowden performed library preparation from most of the samples presented in Chapter 4. Helen has provided advice to optimise the library preparation protocols;
- Alexandra Stark has performed immunocytochemistry experiments.
- Heike Palm and Jorg Reichrath have provided advice regarding the design of the cell culture work described in Chapter 5. Heike Palm has extracted DNA from the cell lines used in their lab. Jorg Reichrath has supervised my work while in Homburg;
- Alastair Droop has performed NGS data-pre-processing. I have worked closely with Alastair and he has taught me the basic principles of NGS data analysis. He has actively been involved in all the discussions regarding NGS data analysis. Alastair has also performed pre-processing and statistical analysis of the Affymetrix U133 plus 2 gene expression data. He also performed the *in silico* experiment described in Chapter 5;
- Helene Thygesen has provided statistical advice and guidance and has performed NGS statistical analysis and NGS-DASL data integration while I was intimately involved in the process;
- ServiceXS performed the HT WG-DASL gene expression microarray experiments;
- Source BioScience performed library preparation using the Nextera protocol and the Illumina Truseq™ protocol on the Agilent Bravo robot;
- The Patterson Institute at the University Of Manchester performed the Affymetrix U133 plus 2 gene expression array experiments;
- Gen-probe extracted nucleic acids from tumour cores (Chapter 4).

I hugely appreciate the hard work, great support and sound advice from all the team members described above. Without their work and support, the work presented in this thesis would not have been possible.

I would like to thank my supervisors Professor Julia Newton-Bishop, Professor Tim Bishop and Dr Mark Harland for their support and guidance during my PhD. I am grateful for the opportunities they gave me to develop new skills and knowledge.

I would like to thank my family for their love and constant support throughout my study in the UK despite the distance and Panos for his love and encouragement especially while I was working in Germany and during the stressful last month before submission.

During my research period I have been funded by a Cancer Research UK PhD studentship (C588/A10721). Further support has been provided by Cancer Research UK programme and project grants (Programme Award C588/A10589 and Project Grants C8216/A6129 and C8216/A8168) and an Association for International Cancer Research (AICR) grant (12-0023).

Abstract

This thesis describes an evaluation of next generation sequencing (NGS) as a tool to study melanoma genomics and identify prognostic biomarkers using formalin-fixed paraffin-embedded (FFPE) primary tumours. I report the evaluation of different library preparation protocols as a pre-requisite for whole-genome DNA NGS. I show that libraries can be generated using small amounts of DNA extracted from primary melanomas (92% success rate). NGS data from 75 samples are presented. To ensure the accuracy of the data and in order to develop a robust protocol for copy number analysis using small FFPE tumours the quality of the data was assessed and potential biases were explored. I observed copy number changes previously reported in melanoma which provide some support of the validity of the data when small DNA input is used. Although, this is work under development, overall the data show that NGS using small FFPE tumours is likely to provide insight into melanoma biology.

I report a study where osteopontin was tested as a potential prognostic biomarker in the plasma of patients with melanoma. The analysis suggests that measurement of plasma osteopontin is unlikely to serve as a useful blood prognostic marker in early-stage disease patients, but it remains a possible marker of relapse in patients under follow up as part of a panel of such markers.

The Leeds group have reported that low vitamin D levels at diagnosis are associated with thicker tumours and poorer outcome. I report an *in vitro* study where melanoma cells were cultured with and without vitamin D. Whole-genome gene expression data were generated using RNA from two vitamin D sensitive melanoma cell lines. The differentially expressed genes in response to vitamin D supported the anti-tumour effects of vitamin D previously reported in cancer models, providing some evidence for vitamin D as a potential adjuvant therapy.

Table of Contents

List of Tables

List of Figures

Abbreviations

(a)CGH	(Array) comparative genomic hybridisation
$1\alpha,25(OH)_2D_3$	1-alpha,25-dihydroxyvitamin D_3
AJCC	American Joint Committee in Cancer
ALM	Acral lentiginous melanoma
ANOVA	Analysis of variance
ATP	Adenosine triphosphate
BMI	Body mass index
bp (suffix)	Base pair
BRAF-i	BRAF-inhibitors
BSA	Bovine serum albumin
BWA	Burrows-Wheeler transform
cDNA	Complementary deoxyribonucleic acid
CGP	Cancer Genome Project
ChIP	Chromatin Immunoprecipitation
CI	Confidence interval
CLIMA	Cell Line Integrated Molecular Authentication
CNA	Copy number alteration
CNV	Copy number variation
cRNA	Complementary RNA
CSD	Chronic sun-induced damage
Ct	Cycle threshold
CV	Coefficient of variation
DASL	cDNA-mediated annealing, selection, extension and ligation
dATP	Deoxyadenosine triphosphate
DFS	Disease-free survival
dH_2O	Distilled water
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DOC	Depth of coverage
ds-cDNA	Double stranded complementary DNA
DSMZ	German Collection of Microorganisms and Cell Culture
DSS	Disease-specific survival
EB	Elution buffer
EDTA	Ethylenediaminetetraacetic acid

ELISA	Enzyme-linked Immunosorbent assay
ENCODE	Encyclopaedia of DNA Elements
EORTC	European Organisation for Research and Treatment of Cancer
FC	Fold change
FCS	Foetal calf serum
FDA	Food and drug administration
FDR	False discovery rate
FFPE	Formalin-fixed paraffin embedded
FISH	Fluorescence in situ hybridisation
g (suffix)	Force of gravity
g (suffix)	Gram
G1	Gap 1 phase of the cell cycle
GO	Gene ontology
H&E	Haematoxylin and Eosin
hg18 (hg19)	Human reference genome version 18 (version 19)
Hi-Di	Highly deionized
HR	Hazard ratio
ICC	Immunocytochemistry
IFN	Interferon
IHC	Immunohistochemistry
kb (suffix)	Kilobase
KEGG	Kyoto encyclopaedia of genes and genomes
LCM	Laser capture microdissection
LDH	Lactate dehydrogenase
LIMMA	Linear models for microarray data
LM	Lentigo maligna melanoma
LREC	Local Research Ethics Committee
µg (suffix)	Microgram
µl (suffix)	Microliter
µM (suffix)	Micromolar
M (suffix)	Molar
MAPK	Mitogen-activated protein kinase
Mb (suffix)	Megabase
mg (suffix)	Milligram
ml (suffix)	Millilitre

MLPA	Multiplex ligation-dependent probe amplification
MM	Metastatic melanoma
mM (suffix)	Millimolar
MM probe	Mismatch probe
MREC	Multicentre Research Ethics Committee
mRNA	Messenger ribonucleic acid
N (suffix)	Normal
NF-κB	Nuclear factor kappa B
ng (suffix)	Nanogram
NGS	Next generation sequencing
NK	Natural killer
NM	Nodular melanoma
nm (suffix)	Nanometre
nM (suffix)	Nanomolar
NSCLC	Non-small cell lung cancer
NUSE	Normalised Unscaled Standard Errors
^{o}C (suffix)	Degrees Celsius
oligo(dT)	Oligodeoxythymidylic acid residues(dT)
OR	Odds ratio
OS	Overall survival
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDA	Pancreatic ductal adenocarcinoma
pH	Measure of acidity or alkalinity
PI3K	Phosphotidylinositol-3-kinase
PIAG	Patient information advisory group
PKA	Protein kinase A
PKC	Protein kinase C
PLCγ	Phospholipase C gamma
pM (suffix)	Picomolar
PM probe	Perfect match
PPAR	Peroxisone Proliferator-Activated Receptor
PWM	Position weight matrix
QC	Quality control
qPCR	Quantitative PCR

qRT-PCR	Real Time quantitative Reverse Transcription PCR
RCC	Renal cell carcinoma
RFS	Relapse-free survival
RGP	Radial growth phase
RIN	RNA integrity number
RLE	Relative log expression
RMA	Robust Multi-Chip Analysis
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
RSN	Robust spline normalisation
RT	Reverse transcriptase
RXR	Retinoid X receptor
S	S phase of the cell cycle
SIBLINGs	Small integrin-binding ligand N-linked glycoproteins
SMM	Superficial spreading melanoma
SNB	Sentinel node biopsy
SNP	Single nucleotide polymorphisms
STR	Short tandem repeats
TE	Tris-EDTA
TGF-beta	Transforming growth factor beta
TMA	Tissue Micro-Array
TNM	Tumour, Node, Metastasis
Tris-Cl	Tris(hydroxymethyl)amino methane
UK	United Kingdom
USA	United States of America
UV(B)	Ultraviolet radiation (B)
VDR	Vitamin D receptor
VDRE	Vitamin D response element
VGP	Vertical growth phase
WG-DASL HT	HT Human Whole-Genome DASL assay

Part I
Introduction and Methods

Chapter 1
Introduction

1.1 The aims of the work reported in this thesis

The major aims are:

- To investigate the use of formalin-fixed paraffin-embedded (FFPE) primary melanoma tissue for genomic analyses, towards the aims of:
    - identifying novel independent prognostic factors;
    - better understanding the biology of melanoma;
- To assess, in particular, the feasibility of using next generation sequencing (NGS) technology to analyse FFPE tissue;
- To explore use of osteopontin blood levels as a potential prognostic biomarker for patients with melanoma;
- To explore the biological effects of vitamin D in melanoma using an *in vitro* model.

1.2 Aims of this chapter

This chapter aims:

- To describe what is known about cutaneous melanoma and common genetic alterations in melanoma tumours;
- To describe current prognostic factors used in the American Joint Committee on Cancer (AJCC) staging system and their limitations;
- To describe the clinico-pathological subtypes of melanoma and a new proposed classification system based on frequent genetic events in melanomas arising in sites of differential sun exposure;
- To provide an overview of tumour and serological prognostic markers;
- To describe current treatment options in melanoma;
- To provide an overview of what is known about vitamin D and melanoma;
- To introduce the following chapters of this thesis.

1.3 Cutaneous melanoma

The malignant transformation of melanocytes, pigment-producing cells, is responsible for melanoma genesis (Chin, 2003). Melanoma was first described as a disease entity in 1806 by Rene Laennec, in his presentation to the Faculté de Médicine in Paris, but references to “black cancer” and “fatal black tumours with metastasis” are traced back to the writings of Hippocrates in the fifth century B.C (Chin *et al*., 1998).

Melanoma can arise wherever pigment cells are found, but is most common in the skin where the greater part of pigment cell population lies. Melanomas may also arise however in non-skin sites including the mucosae (within the sinuses, mouth, nose or vagina) and in the pigmented cell surface of the eye and the internal surface of the eyelids (conjunctival, iris, ciliary body, choroidal). Melanomas arising from the iris, choroid and ciliary body are known as uveal melanomas. The genomic changes reported in uveal melanoma are different to melanomas of the skin and mucosae and I will not discuss uveal melanoma in this thesis. When I use the term henceforth I am referring to cutaneous melanoma unless stated otherwise.

Both genetic and environmental risk factors contribute to the formation of melanoma (Chin, 2003). As originally described by Clark *et al*, primary melanomas progress through two well-defined phases; the radial growth phase (RGP) followed by the vertical growth phase (VGP). RGP is characterised by a flat lesion expanding horizontally in the epidermis and VGP is characterised by invasion into the dermis and other subcutaneous tissues. Finally, if these primary melanomas spread in the skin or other parts of the body, they lead to metastatic disease (Clark *et al*., 1984, Miller and Mihm, 2006, Gray-Schopfer *et al*., 2007).

1.3.1 Incidence

Cutaneous melanoma is the 5th most common cancer in the UK (Cancer Research UK Statistical Information Team, 2010a). The observed increasing incidence rates in Great Britain since the mid-1970s might be in part due to early detection but most is likely to be due to changes in sun-related behaviour such as an increase in holidays in sunny places over time (Cancer Research UK Statistical Information Team, 2010a).

1.4 Melanoma susceptibility

The strongest environmental risk factor for melanoma is ultraviolet radiation (Gandini *et al*., 2005b, Parkin *et al*., 2011, Newton-Bishop *et al*., 2011). The type of sun exposure which is causal is now established to be intermittent and associated with sunburn (Chang *et al*., 2009, Newton-Bishop *et al*., 2011).

That cutaneous melanoma is rare outside the pale skinned populations reflects the view that genetics is a strong determinant or risk. Family history is a significant risk factor (Gandini *et al*., 2005c, Law *et al*., 2012) especially where multiple cases occur in the family (Newton Bishop and Bishop, 2005). The phenotypes associated with melanoma risk such as sun-sensitive skin phenotype, red hair, blue eyes and presence of multiple melanocytic naevi (moles) are all genetically determined, with sun exposure having a relatively small effect on naevus number (Olsen *et al*., 2010, Gandini *et al*., 2005a, Newton-Bishop *et al*., 2010).

The first high penetrance melanoma susceptibility gene (*CDKN2A*) was identified in 1994 by Kamb *et al* (Kamb *et al*., 1994), and less common inherited mutations were subsequently reported in *CDK4* (Soufir *et al*., 1998), and most recently in *TERT* (Horn *et al*., 2013).

Genome-wide association studies have been used to identify lower penetrance susceptibility genes and associations between polymorphisms of pigmentation and naevus-related genes (*MC1R*, *ASIP*, *TYR*) and increased risk of melanoma have been reported (Barrett *et al*., 2011, Iles *et al*., 2013). Thus the genetic basis of the melanoma-related phenotypes is better understood now than before (Bishop *et al*., 2009, Chatzinasiou *et al*., 2011). The more recent genome-wide association studies have highlighted the importance of cell cycle and DNA repair pathways as polymorphisms in genes such as *CASP8* and *CCND1* were reported to be associated with increased melanoma risk (Barrett *et al*., 2011, Law *et al*., 2012).

1.5 Melanoma tumour development

A number of key genetic and epigenetic events in melanomas have been identified through genome characterisation of melanoma compared to melanocytes. Indeed, melanomas have been shown to have very large numbers of mutations supposed to result from exposure to ultraviolet light (Hodis *et al*., 2012). Only some of these events contribute to the development of the tumour and are known as ‘driver

events'. The remaining mutations are denoted as 'passengers' having little or no phenotypic effect. In 2009 Meyle *et al* proposed a model of melanoma development (Figure 1-1) based on evidence from the genome analysis of melanoma (Meyle and Guldberg, 2009). The proposed model describes a linear series of events where a normal melanocyte becomes a benign naevus and then a melanoma (Meyle and Guldberg, 2009). However, it is known that melanoma might occur *de novo* with no pre-existence of a benign naevus. The altered key pathways thought to contribute to melanoma development from a pre-existing naevus are discussed below.

1.5.1 Activation of mitogen-activated protein kinase (MAPK) pathway

The most deregulated pathway in melanoma is the mitogen-activated protein kinase (MAPK) pathway which disrupts the cell cycle and induces proliferation (Figure 1-2) (Chin, 2003). *BRAF* mutations occur in 40% of primary melanomas with the V600E substitution be the most common mutation (Davies *et al*., 2002, Colombino *et al*., 2012). *BRAF* is a proto-oncogene protein kinase and mutations are also found in melanocytic naevi which supports the view that this is an early event in the formation of melanoma arising from a naevus (Pollock *et al*., 2003, Poynter *et al*., 2006). When *BRAF* is not mutated, other genes in the pathway (*NRAS*, *KIT*, *GNAQ*) are often found to be mutated (Meyle and Guldberg, 2009). *BRAF* and *NRAS* mutations are found in melanoma of the skin (Platz *et al*., 2008) while *KIT* mutations in melanomas arising in acral sites (palms, soles, under the nail bed) (Curtin *et al*., 2006) and *GNAQ* in uveal melanomas (Van Raamsdonk *et al*., 2009).

Thus, the driver mutations identified to date in cutaneous melanoma, are predominantly within the MAPK signalling pathway. The dramatic responses to systemic therapies which target mutant *BRAF* (BRAF-i) (Chapman *et al*., 2011), confirm that *BRAF* at least is a driver, even though in the majority of patients, relapse subsequently occurs.

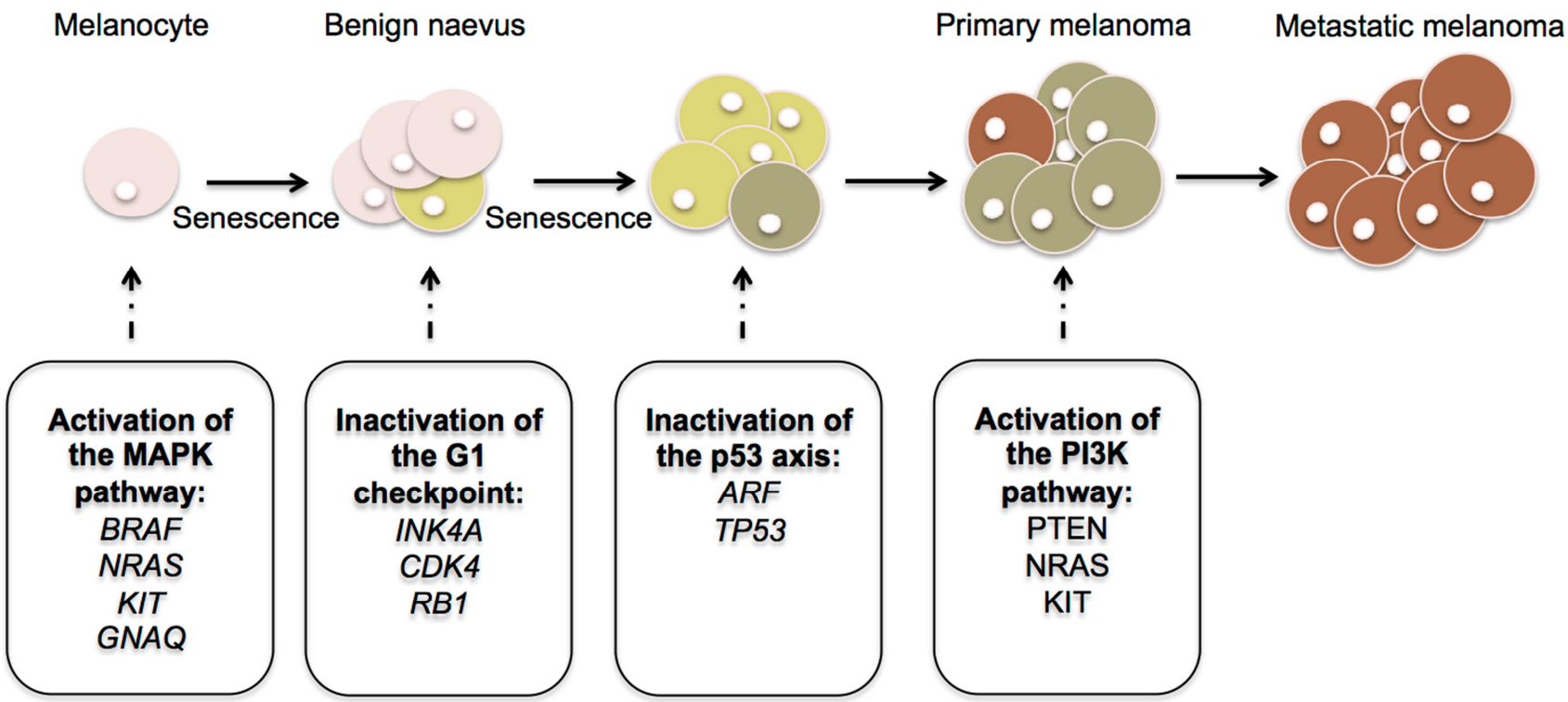


Figure 1-1: A proposed model of melanoma development.

The acquisition of genetic changes drives progression from melanocytes to benign naevus and then primary and metastatic melanoma (more details in text). Adapted from (Meyle and Guldberg, 2009). Abbreviations used: MAPK, mitogen-activated protein kinase; G1, gap 1 phase of the cell cycle; PI3K, phosphatidylinositol-3-kinase.

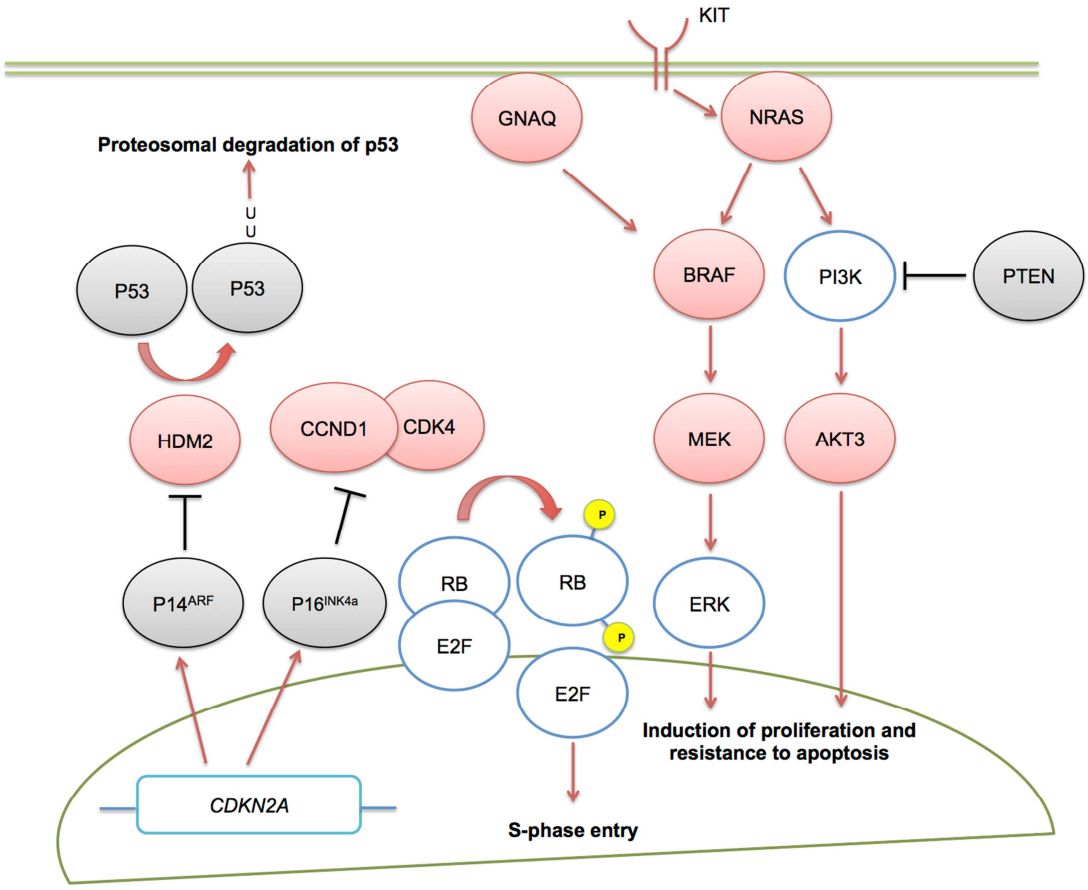


Figure 1-2: Deregulated pathways in melanoma.

Red represents gain of function (e.g. BRAF) and grey loss of function (e.g. p53) due to somatic alterations. Loss of $p16^{INK4A}$ (coded by CDKN2A) leads to release of E2F transcription factors stimulating entry into the cell cycle. Inactivation of $p14^{ARF}$ leads to HDM2-mediated inactivation of p53 through proteosomal degradation. The MAPK pathway can be activated by mutations in *BRAF* or *NRAS*. In the absence of *BRAF* and *NRAS* mutations, ERK can still be constitutively active. NRAS mutations can also activate the PI3K pathway. The PI3K pathway may also be activated by loss of the tumour suppressor PTEN, or by gene amplification of *AKT3*. Adapted from (Gray-Schopfer *et al.*, 2007, Tsao *et al.*, 2012).

1.5.2 Inactivation of $p16^{INK4a}$-RB and $p14^{ARF}$-p53 senescence barrier

When melanocytes acquire *BRAF* mutations they proliferate and then exit the cell cycle, entering a state known as senescence (a benign naevus has been formed) (Gray-Schopfer *et al*., 2006, Michaloglou *et al*., 2005). A *BRAF* mutated naevus evolves into a tumour when the cells exit senescence through inactivation of $p16^{INK4a}$ (Gray-Schopfer *et al*., 2006, Michaloglou *et al*., 2005). Normally, $p16^{INK4a}$, encoded by the *CDKN2A* locus, inhibits cyclin-dependent kinases CDK4/6, which maintains the retinoblastoma protein (RB) in an inactive hypophosphorylated state and the cell does not enter the S phase of the cell cycle (Figure 1-2) (Meyle and Guldberg, 2009). In melanomas $p16^{INK4a}$ is usually inactivated through deletions, mutations and promoter hypermethylation of *CDKN2A* (Chin, 2003, Meyle and Guldberg, 2009). Less frequently, *CDK4* mutations are found in melanomas. CDK4 mutant proteins cannot bind $p16^{INK4a}$ and therefore RB protein remains in its active phosphorylated form (Meyle and Guldberg, 2009, Bartkova *et al*., 1996).

An alternative senescence barrier is observed when melanocytic cells lack $p16^{INK4a}$ through up-regulation of the tumour suppressor p53 (Meyle and Guldberg, 2009, Sviderskaya *et al*., 2003). $p14^{ARF}$ induces p53 and therefore inactivation of the $p14^{ARF}$-p53 barrier is considered essential for the melanocytic cells to exit senescence and evolve into a tumour (Meyle and Guldberg, 2009). *TP53* mutations are found in 5-25% of melanomas (Meyle and Guldberg, 2009). $P14^{ARF}$ is encoded through the *CDKN2A* locus through an alternative reading frame to $p16^{INK4a}$ and therefore is often inactivated when *CDKN2A* is mutated (Figure 1-2) (Meyle and Guldberg, 2009, Tsao *et al*., 2012). $p14^{ARF}$ inactivation acts as an alternative way to abrogate the function of activated p53 through activation of HDM2 which targets p53 for proteosomal degradation (Figure 1-2) (Meyle and Guldberg, 2009, Tsao *et al*., 2012).

1.5.3 Activation of the phosphatidylinositol-3-kinase (PI3K) pathway

PTEN negatively regulates the PI3K pathway, which is important for the regulation of cell growth and survival (Figure 1-2) (Meyle and Guldberg, 2009). *PTEN* inactivation through copy number loss or mutations leads to activation of this pathway (Meyle and Guldberg, 2009, Tsao *et al*., 2012). Reports suggest that *PTEN* inactivation is a rare event in primary melanomas, but common in melanoma metastases, suggesting the importance of PI3K pathway activation in the late

stages of melanoma development (Meyle and Guldberg, 2009, Wu *et al*., 2003). Constitutive activation of this pathway occurs less frequently through copy number gain in *AKT3* and mutations in *NRAS, KIT* and *PIK3CA* (Meyle and Guldberg, 2009, Tsao *et al*., 2012). *PTEN* loss is usually observed in *BRAF* mutated tumours and rarely in *NRAS* mutated tumours and therefore suggesting that *BRAF* and *PTEN* cooperate in melanoma progression (Tsao *et al*., 2012) (Jonsson *et al*., 2007).

In summary then, since the identification of *BRAF* as commonly mutated in melanoma, there has been an explosion in knowledge about the genomic landscape of melanoma. Furthermore this has resulted for the first time in the development of targeted therapies having an effect on survival.

1.6 Clinico-pathological subtypes of cutaneous melanoma

There are 4 main clinico-pathological subtypes of melanoma: superficial spreading melanoma (SMM); lentigo maligna melanoma (LMM); nodular melanoma (NM); and acral lentiginous melanoma (ALM). This classification system was developed based on work by Clark *et al* in the 1970s (Clark *et al*., 1969, Clark *et al*., 1984) and it is still used in the clinic. The system incorporates histological features, anatomical site of the tumour and degree of sun damage to distinguish between the four subtypes.

- SSM is the most common subtype and appears particularly on sites with intermittent sun exposure. SSM usually occurs in young people, mostly on the legs in women and on the trunk in men (Haneke and Bastian, 2006). Histologically, it is characterised by RGP with or without VGP and the cells typically form a pagetoid pattern (Haneke and Bastian, 2006). A “pagetoid” pattern refers to upward spread of atypical cells within the top layer of the skin, the epidermis. Clinically, SSM is presented as a flat lesion (plaque) with irregular pigmentation and margins (Haneke and Bastian, 2006).
- NM is the second most common subtype of melanoma which appears not to have a precursor RGP, presenting exclusively in the VGP. NM usually appears in older people compared to SSM. NM may occur in various sites with the trunk, head and neck and lower legs being the most common (intermittent sun-exposed sites) (Bergman *et al*., 2006). Histologically, it is characterised by atypical cells proliferating within the dermis (Bergman *et al*., 2006). Clinically, NM is presented as a papule or nodule usually symmetric, typically with abnormal pigmentation. A failure to diagnose this sort of melanoma early is

common resulting from an apparently more rapidly growth rate, and although such tumours may be deeply pigmented, they may also look very red. Nodular melanomas are more likely to be ulcerated (Bergman *et al*., 2006).

- LMM is a subtype which appears particularly on chronic sun-exposed sites in the RGP with or without VGP. LMM usually occurs in elderly people with the head and neck being the commonest sites in both sexes (Heenan *et al*., 2006). Histologically it is characterised by small naevoid or epithelioid cells usually extending down the walls of hair follicles and sweat ducts when in VGP (Heenan *et al*., 2006). Clinically, LMM appears as a plaque or a nodule with irregular pigmentation and well-defined irregular borders with evidence of severe solar skin damage (Heenan *et al*., 2006).
- ALM is the rarest subtype which is defined based on the anatomic location of the tumour and a particular histological pattern, which occurs in palms, soles and subungual sites. ALM shows similarities to LMM histologically. In the RGP, lesions are characterised by variable size cells often dendritic and in the VGP, by small epithelioid often spindled (Tokura *et al*., 2006). Atypical melanocytes can extend along the sweat glands into the dermis (VGP) which shows nests of tumours (Tokura *et al*., 2006). Clinically, lesions in the RGP appear as a plaque with irregular borders and in the VGP as a nodule which may be accompanied by the presence of ulceration (Tokura *et al*., 2006).

Rarer subtypes including desmoplastic, spitzoid, minimal deviation and malignant blue naevus account for the 5% of the rest of melanomas (Elder and Murphy, 2010).

These well-described clinico-pathologically determined subtypes of melanomas arising in the skin have been established to be biologically different to each other to a certain extent. Molecular changes might provide a more accurate classification system, which will not only reflect the biological differences between the subtypes but form the basis of predictive tests used for the selection of patients for clinical trials of targeted therapies (Kabbarah and Chin, 2005). In 2005, Curtin *et al* proposed such a classification system based on the genetic changes identified in melanomas in relation to the anatomical site of the tumour, and patterns of sun exposure (Curtin *et al*., 2005). These are broad groupings reflecting a tendency to see differences rather than absolutes.

The four biologically relevant-subtypes proposed by Curtin *et al* (Curtin *et al*., 2005), and more recently reviewed by Whiteman *et al* (Whiteman *et al*., 2011), are:

- Melanomas arising on chronically sun-exposed skin (face) which usually appear in older people. These melanomas often do not arise from a pre-existing naevus. At the molecular level, these melanomas have lower prevalence of mutations in *BRAF* than in melanomas arising on minimally or non-sun exposed sites, with 20-30% showing mutations in *NRAS* or *KIT (Curtin et al., 2005, Curtin et al., 2006)*;
- Melanomas arising on intermittently sun exposed skin (trunk, arms, legs) which represent the most common type of melanoma in European populations. These melanomas are more likely to be associated with either a pre-existing naevus or generally high number of naevi. Molecularly, these melanomas are more frequently associated with a high prevalence of *BRAF* mutations (about 70%) (Curtin *et al*., 2005);
- Melanomas arising on minimally sun-exposed skin (soles, palms) for which there are few epidemiological data on causation, but are thought not to be related to sun exposure. About 20-30% of these melanomas show mutations in *KIT* (Curtin *et al*., 2006, Whiteman *et al*., 2011) and 40% show *CCND1/CDK4* amplifications (Sauter *et al*., 2002), otherwise they are characterised by frequent copy number gains and losses;
- Melanomas arising on sun-protected sites (mucosal membranes) are characterised by as frequent copy number alterations as melanomas arising on the acral sites (minimally sun-exposed skin). However, the regions affected by copy number gains or losses are different to those found on minimally sun-exposed melanomas. In melanomas arising on sun-protected sites, *CCND1* amplifications are less frequent and *CDK4* amplifications are more prevalent compared to melanomas arising on minimally sun-exposed sites (Curtin *et al*., 2005).

Histopathologists have long recognised that there are marked differences in the appearance of melanomas both in terms of architecture of the tumour and the nature of the cells. As the genomics of melanoma are explored then there is an emerging pattern of genomic changes which reflects the histological subtypes of the tumours to a certain degree. These patterns give support to the view that melanoma represents a wide range of different tumours, but are recognised to be only the beginning of the process. My work was designed to add to the information on primary tumour genomics to allow the recognition of better prognostic and predictive biomarkers and to better understand the biology.

1.7 Prognosis in melanoma

Survival from melanoma, as with other cancers, is improving over time, which is attributed to faster and earlier diagnosis and improvements in treatment (Cancer Research UK Statistical Information Team, 2010b). In England survival rates have increased over the last 30 years (Cancer Research UK Statistical Information Team, 2010b). Patients with thin melanomas (≤1mm) have excellent prognosis with 90-95% surviving for at least 5 years, and 83-88% for at least 10 years (Table 1-1) (Balch *et al*., 2009). However, for patients with larger tumours or with metastasis survival rates are much lower (Table 1-1) (Balch *et al*., 2009).

Currently, clinical and histopathological prognostic markers, which predict outcome for melanoma patients, are included in the final version of the American Joint Committee on Cancer (AJCC) staging system. This system uses primary tumour characteristics (Breslow thickness, mitotic rate per mm^2, tumour ulceration), lymph node metastasis, site of distant metastasis, and serum lactate dehydrogenase (LDH) level, as criteria for the tumour-node-metastasis (TNM) classification and the group staging (Table 1-1) (Balch *et al*., 2009). These criteria reflect the most significant independent prognostic factors, as described and analysed in 30,946 patients with stages I, II and III melanoma and 7,972 with stage IV melanoma.

Table 1-1: Current AJCC staging system (7th edition) for cutaneous melanoma.

Survival rates for individual TNM stage are based on the 6th edition of the AJCC staging system where Clark's level of invasion determine T1a and T1b stages and has now been replaced with mitotic rate (Balch *et al*., 2001a). Adapted from (Balch *et al*., 2009). Abbreviations: TNM (Tumour, Node, Metastasis); LDH (lactate dehydrogenase).

Group staging	TNM classification	Thickness (mm)	Ulceration & mitotic rate (per mm^2)	Positive nodes	Nodal size	Distant metastasis	5-year survival	10-year survival
IA	T1a	≤1	No & mitosis<1	0	-	-	95.3	87.9
IB	T1b	≤1	Yes or mitosis ≥1	0	-	-	90.9	83.1
	T2a	1.01-2.0	No	0	-	-	89.0	79.2
IIA	T2b	1.01-2.0	Yes	0	-	-	77.4	64.4
	T3a	2.01-4.0	No	0	-	-	78.7	63.8
IIB	T3b	2.01-4.0	Yes	0	-	-	63.0	50.8
	T4a	>4.0	No	0	-	-	67.4	53.9
IIC	T4b	>4.0	Yes	0	-	-	45.1	32.3
IIIA	N1a	Any	No	1	Micro	-	69.5	63.0
	N2a	Any	No	2-3	Micro	-	63.3	56.9
IIIB	N1a	Any	Yes	1	Micro	-	52.8	37.8
	N2a	Any	Yes	2-3	Micro	-	49.6	35.9
	N1b	Any	No	1	Macro	-	59.0	47.7
	N2b	Any	No	2-3	-	-	46.3	39.2

Group staging	TNM classification	Thickness (mm)	Ulceration & mitotic rate (per mm^2)	Positive nodes	Nodal size	Distant metastasis	5-year survival	10-year survival
IIIB cont.	N2c	Any	No	In transit metastases/satellites without metastatic nodes				
IIIC	N1b	Any	Yes	1	Macro	-	29.0	24.4
	N2b	Any	Yes	2-3	Macro	-	24.0	15.0
	N2c	Any	Yes	In transit metastases/satellites without metastatic nodes	-	-	26.7	18.4
	N3	Any	Any	≥4 nodes, or matted nodes, or in transit metastases/satellites with metastatic nodes	Micro/ Macro	-		
IV	M1a	Any	Any	Any	Any	Distant skin, subcutaneous or nodal	18.8	15.7
	M1b	Any	Any	Any	Any	Lung	6.7	2.5
	M1c	Any	Any	Any	Any	Other visceral or any distant metastases with elevated LDH level	9.5	6.0

1.7.1 Primary tumour characteristics (Stage I/II)

1.7.1.1 Breslow thickness

Breslow identified back in 1970 that the thickness of the tumour predicts its ability to metastasise (Breslow, 1970). Breslow thickness is assessed by a pathologist as the vertical thickness of the tumour measured in millimetres at the thickest part of the lesion, from the granular layer of the epidermis to the deepest part of the tumour. It is the strongest prognostic factor in localised disease with thicker tumours indicating worse prognosis and therefore it has been incorporated into the AJCC staging system (Balch *et al*., 2009, Balch *et al*., 2001b). Breslow thickness also predicts survival for patients with nodal disease (stage III) (Balch *et al*., 2010, Balch *et al*., 2009).

1.7.1.2 Tumour ulceration

Ulceration of a primary melanoma is determined histologically and defined as the absence of intact epidermis overlying a major part of the lesion (Balch *et al*., 2001a). Ulceration is an independent prognostic factor for both localised and stage III melanomas (Balch *et al*., 2009, Balch *et al*., 2010). Breslow thickness remains the strongest predictor of survival in localised disease, but the presence of ulceration strongly influences survival. Thus, survival rates for patients with an ulcerated tumour are lower compared to patients with a non-ulcerated tumour in the same thickness group, and remarkably similar compared to patients with a non-ulcerated tumour in the next thickness group (Table 1-1) (Balch *et al*., 2009).

1.7.1.3 Mitotic rate

Mitotic rate is defined as the number of mitoses per mm^2 in the most mitotically active part of the lesion. Mitotic rate reflects tumour proliferation and is a strong prognostic factor in localised melanoma (Balch *et al*., 2009, Thompson *et al*., 2011). It has replaced the level of invasion as described by Clark (Clark *et al*., 1969) in the 7^{th} edition of AJCC staging system with mitotic rate >1mm^2 and/or presence of ulceration to determine T1b subcategory (Balch *et al*., 2009, Thompson *et al*., 2011).

1.7.2 Metastatic tumour characteristics

There is evidence that a primary tumour is actually a heterogeneous cell population, with a minority of cells being able to metastasise through the lymphatic and haematogenous systems and form colonies in distant organs (Hanahan and Weinberg, 2011). Metastasis depends on tumour characteristics, such as loss of cell adhesion, greater motility and invasiveness, but also on tumour/host interactions which involve the immune system and tumour microenvironment (Hanahan and Weinberg, 2011).

1.7.2.1 Nodal metastasis (Stage III)

Typically, the first site of melanoma metastasis is the regional nodal basin, the sentinel node, which is the first tumour-draining node in the lymphatic chain (Leong *et al*., 2006). If clinical examination fails to detect lymph node involvement a sentinel node biopsy is used to identify metastatic deposits. This procedure involves injection of blue dye and radioactive tracer into the skin where melanoma was located and identification of the sentinel node (which takes up the isotope and appears as a “hot” node), which is then removed and assessed histologically using melanocyte lineage markers, such as S100 (Marsden *et al*., 2010). If a patient has a positive sentinel node biopsy, then patients undergo complete lymphadenectomy to remove all regional nodes and the number of nodes involved is evaluated histologically (Marsden *et al*., 2010). The number of tumour-involved nodes, and whether the tumour deposits in these nodes are microscopic (diagnosed after sentinel node biopsy and/or complete lymphadenectomy) or macroscopic (diagnosed clinically), are predictors of outcome, with larger tumour burden associated with worse prognosis (Balch *et al*., 2009, Balch *et al*., 2010). Primary Breslow thickness and ulceration remain significant prognostic factors for stage III disease (Balch *et al*., 2009). In transit and satellite lesions are defined as intralymphatic metastases and, combined with nodal metastases, are associated with worse prognosis when compared to the absence of nodal metastases (Table 1-1) (Balch *et al*., 2009). It has been reported that the size of nodal metastasis is also of prognostic significance, with small tumour burden (<0.1mm) associated with a similar prognosis to patients with a negative sentinel node biopsy (van Akkooi *et al*., 2008). Also, location of tumour deposits in the nodes influences outcome, as tumour deposits in the subcapsular sinus are associated with better outcome compared to deposits within the parenchyma of the node (Dewar *et al*., 2004). The location and size of nodal metastasis is not currently involved in the AJCC staging

system, but this might change in the future with more sentinel node biopsies taking place and accurate pathological reporting.

1.7.2.2 Systemic metastasis (Stage IV)

A proportion of melanoma tumour cells acquire the ability to spread systemically, commonly to skin, lymph nodes, liver, lung and brain. The site of distant metastasis has prognostic significance, such that skin, subcutaneous and lymph node sites are associated with better survival rates than metastases at visceral sites, while lung confer slightly better prognosis than other visceral sites (Balch *et al*., 2009). These differences in outcome might depict biological differences between tumours growing at different sites but they are as yet poorly understood.

Elevated serum lactate dehydrogenase (LDH) levels are associated with worse prognosis in stage IV disease irrespective of the site of metastasis (Balch *et al*., 2009). LDH is thought to be involved in the cellular production of energy during the hypoxic conditions of the tumour microenvironment (Agarwala *et al*., 2009). Elevated LDH serum levels reflect cellular apoptosis when part of the tumour outgrows its blood supply and thus correlate with tumour burden (Deichmann *et al*., 1999, Agarwala *et al*., 2009).

1.7.3 Prognostic factors not included in the current staging system

In addition to the current staging criteria, both host and tumour related factors have been related to outcome. Host factors, such as age, gender and anatomical site of the tumour have shown independent prognostic significance as reviewed by Payette *et al* (Payette *et al*., 2009). In 2001, when the AJCC staging guidelines were updated for the 6th edition, it was reported that increasing age, male gender and tumours arising on the trunk, head, neck, palms, soles and under the nails were associated with worse prognosis, even for patients with stage III disease (Balch *et al*., 2001b). Despite their exclusion from the staging guidelines, these factors are taken into account in several studies reporting multivariable models to predict outcome from melanoma (Elder and Murphy, 2010).

Tumour characteristics such as clinico-pathological subtype, vascular invasion, tumour-infiltrating lymphocytes and tumour regression have also emerged as potential prognostic factors, reviewed by Ross and Payette *et al* (Ross, 2006, Payette *et al*., 2009), but they are not part of the current staging guidelines.

1.7.4 Limitations of the current staging system

Despite the efforts placed to accurately identify prognostic factors for patients with melanoma, the variance in survival within stage is still quite large and cannot be explained using the current prognostic factors (Ross, 2006). There is a proportion of low-risk tumours which progress to advanced disease (Slingluff *et al*., 1988) and which are supposed to be biologically different and potentially therefore identifiable using genomics. There is an increasing need to identify new prognostic biomarkers to enable better stratification of risk of metastasis for patients with melanoma but also enable early treatment intervention for those at high risk.

1.8 New prognostic biomarkers

Increasing numbers of new prognostic biomarkers for melanoma are now emerging but their clinical utility in large-scale studies is still unproven. Biomarker development includes 5 phases: preclinical exploratory studies; assay development and validation; retrospective longitudinal clinical repository studies; prospective screening studies; and randomized control trials (Pepe *et al*., 2001). An ideal biomarker should be measured easily, reliably and with low cost, by a sensitive and specific assay (Kulasingam and Diamandis, 2008). The following sections outline used and under investigation tumour and blood biomarkers in melanoma.

1.9 Tumour biomarkers

The histopathological features of primary melanoma, such as Breslow thickness, ulceration and mitotic rate are robust prognostic indicators, and therefore likely correlate well with key biological changes in tumours. Independent prognostic biomarkers, both germline and tumour, would however be valuable biomarkers, and would likely give key biological insights into the determinants of relapse (Bosserhoff, 2006).

Many studies have attempted to identify new tumour prognostic biomarkers. However, none have managed to improve risk stratification (Gould Rothberg *et al*., 2009). Techniques such as immunohistochemistry (IHC) for protein expression, microarrays for gene expression profiling and detection of copy number changes, and Sanger sequencing for mutation screening have been used. Most studies have

used small sample numbers and lack multivariable analyses, therefore validation studies need to be done to assess the prognostic significance of markers identified (Gogas *et al*., 2009, Gould Rothberg *et al*., 2009). Emerging technologies, such as next generation sequencing (NGS), are promising tools for the validation of previous studies and identification of novel biomarkers.

NGS technologies offer the advantage of high-throughput, massively-parallel sequencing of the genome, transcriptome and epigenome (Meyerson *et al*., 2010). All NGS technologies utilise the sequential addition of nucleotides in immobilised DNA templates on glass slides (flow cells). After years of evolution, NGS systems since their first release in 2005-2006 show better performance in terms of sequence length (read length), accuracy, applications and informatics infrastructure (Liu *et al*., 2012). These technologies have revolutionised cancer research allowing fast, high-depth sequencing of cancer cells (Ding *et al*., 2010). A number of different analytical tools have been developed to allow a variety of biological questions to be addressed. At the same time the analysis of the large amount of data generated (>600GB per run of the Illumina HiSeq system) has been a major challenge in understanding the complex cancer genome (Ding *et al*., 2010).

In this thesis, in Chapters 3 and 4, I report attempts to develop a next generation sequencing assay to identify copy number changes using formalin-fixed paraffin-embedded (FFPE) specimens as a means of generation of prognostic biomarkers in patients with melanoma.

1.9.1 Copy number alterations

DNA copy number changes (losses or gains) are an important part of genetic variation which may involve very large proportions of the cancer genome. Copy number changes are defined as being larger than 1kb in size (Shlien and Malkin, 2009). Germline DNA copy number changes are usually referred as copy number variations (CNVs) while somatic DNA copy number changes are referred as copy number alterations (CNAs). If a copy number change in a specific region is shown to be associated with a given cancer or prognosis for example, the genes located in that region are candidates for playing a role in tumour development and/or progression (Shlien and Malkin, 2009). As was discussed above however for driver mutations, genes identified may be “passengers” rather than playing a causal role. An example of a causal deletion is genomic DNA copy number loss in the chromosome arm 9p affecting the *CDKN2A* gene, as was discussed above (Kamb *et al*., 1994).

A large number of studies have been conducted to identify somatic CNAs in melanoma as a means to identify key biological changes (Table 1-2). In the past, karyotype analysis was the only whole-genome method to identify such changes. More recently, single nucleotide polymorphism (SNP) arrays, array comparative genomic hybridisation (aCGH) and NGS are used. Most whole-genome studies conducted so far have used tumour cell lines or frozen tumours but only a few used FFPE tissue (Table 1-2). Melanoma tumours are small and they are usually preserved in formalin rather than cryopreserved, so availability of frozen tissue is limited. In Table 1-2 I report CNAs identified in fresh-frozen and FFPE melanoma tumours, and human melanoma cell lines, through whole-genome analysis, and the reported associations with histological features or mutation status in a proportion of the studies. Frequent CNAs identified by whole-genome studies include losses in chromosomes 9 and 10 and gains in chromosomes 6, 7 and 11 (Table 1-2). An example of a recurrent focal copy number loss has been reported in 9p21 region where the *CDKN2A* gene is located (Lin *et al*., 2008, Gast *et al*., 2010). Studies using candidate approaches such as fluorescence in situ hybridisation (FISH) and multiplex ligation-dependent probe amplification (MLPA) assays have reported supportive information that the 9p21 region is frequently altered in melanoma tumour samples (Takata *et al*., 2005, Rakosy *et al*., 2008).

A number of studies have also reported CNAs in different histological subtypes (Table 1-2). Bastian *et al* first identified different somatic copy number changes in acral melanomas compared to SSM (Bastian *et al*., 2000). Two further studies supported the view that melanoma is a heterogeneous disease with different genetic alterations identified in melanomas arising in sun-exposed sites versus those in sun-protected sites (Table 1-2) (Bastian *et al*., 2003, Curtin *et al*., 2005). In addition, a few studies have provided evidence that different CNAs are seen in melanomas with different mutations (Table 1-2). For example, it has been shown that *BRAF* mutation is associated with copy number loss in the 10q region where the *PTEN* gene is located (Jonsson *et al*., 2007, Gast *et al*., 2010).

More recently, whole-exome NGS studies using cell lines or frozen tumours were conducted to identify recurrent mutations in melanoma but also reported the observed CNAs (Hodis *et al*., 2012, Krauthammer *et al*., 2012). For example, both studies reported losses in 9p and 10q and gain in 11q in the samples tested confirming previous studies (Table 1-2).

Table 1-2: Whole-genome studies reporting copy number alterations in melanoma tumour samples and human melanoma cell lines.

Studies reporting frequent alterations in distinct histological subtypes and in *BRAF/NRAS/PTEN* mutated tumours or ulceration are also included. Chromosomal changes which were validated using a different method are highlighted in bold. Studies focusing on the identification of the heterogeneity or clonal evolution of melanoma are not included. Abbreviations used: MM (metastatic melanoma); FISH (fluorescence in situ hybridisation); CGH (comparative genomic hybridisation); SNP (single nucleotide polymorphism); TMA (tissue microarray); LMM (lentigo maligna melanoma); SSM (superficial spreading melanoma); CSD (chronic sun-induced damaged).

Samples	Methods	Copy number losses	Copy number gains	Reference
2 primaries & 15 MM	Karyotype	1p, 6p	7q	(Balaban *et al.*, 1984)
8 MM	Karyotype	1p	1q, 6, 7	(Limon *et al.*, 1988)
11 primaries & 43 MM	Karyotype	10	7	(Parmiter *et al.*, 1988)
1 primary & 10 MM	Karyotype	1, 6, 9	7	(Cowan *et al.*, 1988)
158 MM	Karyotype	In all cases (>20%): 6, 10, 1, 7, 9 In 49 near-diploid cases (>20%): 9, 10, 19, 22	In 49 near-diploid cases (>20%): 7	(Thompson *et al.*, 1995)
172 tumours (Data derived from a chromosome data bank)	Karyotype	(>15%): 1p21-36, 3, 4p35, 6q12-27, 8p10-23, 9p22-24, 10, 11p15, 11q13, 11q22-25, 12q24, 13p11-13, 13p34, 14p10-13, 15, 16	(>15%): 1q10-44, 6p21-25, 7, 8, 20	(Mertens *et al.*, 1997)
3 MM cell lines	CGH FISH (validation)	In the highly metastatic line: **4**, 9p21.3, 10p		(Adam *et al.*, 2000)

Samples	Methods	Copy number losses	Copy number gains	Reference
1 primary short-term culture	CGH Karyotype	4q, 5q, 8p (in CGH only), 9p, 10, 11 In karyotype only: 12p, 12q, 16q, 19	5p, 8q, 13q (karyotype only)	(Wiltshire *et al.*, 2001)
46 primary and MM tumours (short-term cultures)	CGH FISH (validation)	(>30%): 4q, 6q, 9p, 10p/q, 11q	(>30%): 6p, 7p/q, 17q, 20p/q (>20%): **3q26** (*hTERC*), **5p15.33** (*hTERT*), 6p, 7p, 7q, 17q, 20p, 20q	(Pirker *et al.*, 2003)
68 FFPE tumours (19 *BRAF* mutated)	aCGH FISH (validation)		**7q** in 24 tumours 9 of which were *BRAF* mutated	(Maldonado *et al.*, 2003)
NCI60 cell lines (8 melanoma lines)/ 62 tumours (qPCR)/179 primary and metastasis TMAs (FISH)	SNP array qPCR (validation) FISH (validation)		**3p13-14** (*MITF*) in 70.5% of lines Associated with *MITF* increased expression, metastasis & decreased overall survival (tumour samples)	(Garraway *et al.*, 2005)
30 microdissected tumours (15 AMs, 15 SSMs)	CGH FISH (validation) *Tumour to normal ratio>1.5 including a whole chromosome not considered a gain	In AMs (>40%): 6q In both AMs and SSMs (>40%): 9p, 10q	In AMs (>20%): **5p15**, **7p, 11q13**, **22q11-13**	(Bastian *et al.*, 2000)
132 FFPE primaries	CGH	(>20%): 9p, 9q, 10q , 10p, 6q, 11q Frequent in acral vs non-acral melanomas: 6q23.1, 15q13, 16q24 Frequent in LLM vs SSM: 13q21.1, 17pter Frequent in non-CSD vs CSD melanomas: 10pter, 10q22.2 Frequent in CSD vs non-CSD melanomas: 17pter	(>20%): 6p, 1q, 7p, 8q, 17q, 20q Frequent in acral vs non-acral melanomas: 5pter, 12q14, 4q11, 11q13.3 Frequent in LLM vs SSM: 15q21.1, 17qter Frequent in CSD vs non-CSD melanomas: 15q15	(Bastian, 2003)

Samples	Methods	Copy number losses	Copy number gains	Reference
2 frozen tumours (acral primary & matched MM)	SNP array WG-NGS aCGH FISH (validation)	(In both primary and metastasis): 9p	(In both the primary and metastasis): **4q12** (*KIT*), 11q13, 11q14, 17p11, 20q11	(Turajlic *et al*., 2012)
17 fresh primaries	aCGH	4q, 6q*, 7q, 9q, 10q* 11p, 14q, 15p, 17q, 20q, 21p (frequent in ulcerated tumours) *Losses were associated with gene expression in these locations	3p, 6p, 7p/q, 8q, 11p/q, 15q, 16p, 17q, 18q, 22q (frequent in ulcerated tumours)	(Rakosy *et al*., 2013)
18 FFPE primaries	aCGH	9p & 10q (independent of prognosis), frequent focal losses (in patients with bad prognosis vs those with good)	Frequent focal gains (in patients with bad prognosis vs those with good)	(Hirsch *et al*., 2013)
102 FFPE primaries	aCGH		4q12 (*c-KIT*): (>20% in acral/mucosal/chronically sun-exposed tumours but not observed in non-chronically sun exposed tumours) Associated with increased c-KIT protein levels	(Curtin *et al*., 2006)
9 MM cell lines 49 lines (PCR)	CGH M-FISH Digital karyotyping PCR (validation)	(>30%): 6q, 9p, 10q, **Xp21.2 (*DMD*)*** ***Loss was associated with decreased *DMD* gene & protein expression	(>30%): 1q, 6p, 7p/q, 20p/q, 22q, Xp/q	(Korner *et al*., 2007)
8 primary and MM cell lines	SNP array		7q (in 6 lines of which 5 were *BRAF* mutated)	(Spittle *et al*., 2007)
93 FFPE tumours	aCGH publicly available data	16q24.3 (*MC1R*) in 12.9% of samples (infrequent)	16q24.3 (*MC1R*) in 7.5% of samples (infrequent)	(Kim *et al*., 2008)

Samples	Methods	Copy number losses	Copy number gains	Reference
83 tumours and cell lines (aCGH) 38 TMA cores from primary and metastatic tumours (FISH)	aCGH data available (Maser *et al*., 2007) FISH (validation)		**5p13** (*GOLPH3*) (3.4% of samples on aCGH and 31.6% on TMA cores) Associated with increased *GOLPH3* gene expression	(Scott *et al*., 2009)
14 MM tumour cell lines	SNP array qPCR (validation)		**7q31-34** (mir-182 cluster, *cMET*, *BRAF*) (53.8%) Associated with increased mir-182 expression	(Segura *et al*., 2009)
52 patient-derived primary and MM cell lines/ 43 frozen MM tumours	aCGH data (Greshock *et al*., 2009)		(>20%) : 6p21.32* (*MHC* Class I & II genes) *associated with *MHC* Class I & II gene expression in VGP cell lines only	(Degenhardt *et al*., 2010)
18 MM cell lines 22 FFPE MM (FISH)	SNP array qPCR (validation)		(>40%): **5p13.2*** (*SKP2*) *associated with increased SKP2 protein expression	(Rose *et al*., 2011)
5 frozen MM (mucosal)	Whole-genome NGS SNP array		1q, 8q	(Furney *et al*., 2013)
150 melanomas (101 short-term cultures and cell lines) (SNP array) 42 TMAs (FISH)	SNP array data ((Beroukhim *et al*., 2010)) FISH (validation)		(>40%): **1p12*** (*PHGDH*) *Gain associated with increased PHGDH protein expression	(Locasale *et al*., 2011)

Samples	Methods	Copy number losses	Copy number gains	Reference
101 MM short-term cultures (SNP array)/ 70 FFPE primaries (aCGH)	SNP array aCGH (validation dataset (Curtin *et al.*, 2005))	Most frequent: **9p21** (*CDKN2A*), **10q23*** (*PTEN, CUL2, KLF6*) *Association with *PTEN, CUL2, KLF6* gene expression General pattern of losses validated in aCGH tumour data	Most frequent: **7q34*** (*BRAF*), 3p13 (*MITF*), **11q13** (*CCND1*) *Associated with *BRAF* mutations General pattern of gains validated in aCGH tumour data (3p13 gain not frequent in tumour data)	(Lin *et al.*, 2008)
47 cell lines (SNP array) 130 FFPE primaries (aCGH)	SNP array aCGH (validation dataset (Curtin *et al.*, 2005))		(>20% in tumours but not frequent in lines): 7p22 (*Epo*), 19p13 (*EpoR*) Both associated with acral/mucosal melanomas	(Kumar *et al.*, 2012)
60 patient-derived MM cell lines	SNP array	9p21* (*CDKN2A*), 10q23** (*PTEN*), 17p*** (*TP53*) *associated with lack of expression and was independent of *BRAF*/*NRAS* mutation status **associated with *BRAF* mutation ***associated with *NRAS* mutations	7q* (*BRAF*), 11q** (*CCND1*), 12q** (*CDK4*), *associated with *BRAF* mutation **associated with wild-type *BRAF*/*NRAS*	(Gast *et al.*, 2010)
47 cell lines (1 primary & 46 MM)	aCGH	(>50%): 4p/q, 6q, 8p, 9p*, 10p/q***, 11q** *associated with *PTEN* mutation **associated with *BRAF* mutation ***associated with both *BRAF* and *PTEN* mutations	(>50%): 1q, 7p/q*, 8q, 17q, 20q *associated with *BRAF* mutation	(Jonsson *et al.*, 2007)
52 patient-derived primary and MM tumour cell lines/ 43 frozen metastatic tumours	aCGH	11q24-25*, 16q21-qter** *Associated with *NRAS* mutation **Associated with *BRAF* mutation	5p15, 7q32-33 Both associated with *BRAF* mutation	(Greshock *et al.*, 2009)

Samples	Methods	Copy number losses	Copy number gains	Reference
76 primary and MM cell lines	SNP array	(>40%): 6q, 9p/q, 10p/q, 11q, 17q Frequent focal copy number losses (>50%) included 9p21 (*CDKN2A)*, 10q23 (*PTEN)*	(>20%): 2p, 6p, 7p/q, 19q, 20p, 20q, and 22q (>12%): 3p13 (*MITF*)	(Stark and Hayward, 2007)
85 frozen tumours (23 from sun-protected sites)	aCGH FISH (validation)		(6%):4q12 (*KIT*), (9%): **11q14** (*GAB2*)* Both associated with tumours in sun-protected sites *Independent of *BRAF*/*NRAS*/*KIT* mutations *Associated with increased protein expression	(Chernoff *et al*., 2009)
126 FFPE primaries	aCGH	3q*, 4q* 8p*, 10**, 11p* *Associated with mucosal melanomas **Associated with acral/mucosal and tumours on skin without sun-induced damage	1q*, 6p**, 11q13**, 17q*** *Associated with acral tumours **Associated with mucosal & acral tumours Associated with mucosal melanoma	(Curtin *et al*., 2005)
121 frozen tumour/normal blood pairs (15 primaries, 30 MM, 76 short-term cultures from MM)	Whole-exome NGS	Most frequent: 4q22.1, 4q.35.1, 5p11.2, 9p21.3 (*CDKN2A*), 10q23.31 (*PTEN*)	Most frequent: 2q21.2, 7q.34, 5p15.33 (*TERT*), 7q34, 11q13.2 (*CCND1*),	(Hodis *et al*., 2012)
68 primary and MM cell lines/31 primaries and MM frozen tumours	Whole-exome NGS	(>10%): 9p21, 10p12, 10q11	(>10%): 1q31, 5p13, 11q14	(Krauthammer *et al*., 2012)

The findings reported in Table 1-2 provide support that there are distinct molecular pathways in subsets of melanomas driven by different genetic alterations. Relating such genetic alterations to prognosis might improve prognostication in melanoma and help in the development of new targeted therapies. In the next section, I will review the studies addressing the prognostic significance of CNAs in melanoma.

1.9.1.1 Copy number alterations and prognosis

Only a small number of studies have explored the prognostic significance of CNAs in melanoma. Some studies have reported differences in CNAs between primary and metastatic samples, relapsers and non-relapsers, while a couple associated CNAs with relapse-free survival (RFS) or overall survival (OS). These studies often lacked validation in independent datasets or multivariable analysis which is important to identify if the reported marker adds to the current staging system (McShane *et al*., 2005). It has been reported that chromosomes 1 and 7 are often gained in metastatic tissue (Udart *et al*., 2001, Lee *et al*., 2001) while *a* more recent study reported that *CCND1* gain (located in 11q13) and *MYC* gain (located in 8q34) were associated with worse overall survival in 107 primary melanomas (Gerami *et al*., 2011). The *EGFR* gene, which is located in 7p12, was reported to be amplified in relapsers within 5 years of diagnosis compared to non-relapsers within 5 years (Rakosy *et al*., 2007). However, a different study reported that *EGFR* amplification was not associated with RFS or OS (Boone *et al*., 2011). Chromosome gain in 3p14 where microphthalmia-associated transcription factor (*MITF*) is located has also been reported to be associated with disease-free survival (DFS) in 104 metastatic tumours (Ugurel *et al*., 2007) and worse OS in 160 metastatic tumours (Garraway *et al*., 2005). All studies reporting CNAs associated with melanoma prognosis are listed in Table 1-3.

Most studies have reported the use of candidate molecular cytogenetic assays, such as FISH, and studies using whole genome approaches have been limited (Table 1-3). In Chapter 4 I describe the use of whole-genome NGS sequencing using FFPE melanoma tumours which could be used for the identification of somatic genetic alterations associated with prognosis.

Table 1-3: Copy number alterations and melanoma prognosis.

Abbreviations used: FISH (fluorescence in situ hybridisation); aCGH (array comparative genomic hybridisation); SNP (single nucleotide polymorphism), MLPA (multiplex ligation-dependent probe amplification); RFS (relapse-free survival); OS (overall survival); DFS (disease-free survival); DSS (disease-specific survival); HD (homozygous deletion).

Samples	Methods	Outcome measure	Result	Reference
97 FFPE primaries	FISH (6q25, 6q23, 11q13, 9p21, 8q24, 20q13)	Relapse status	11q13 (*CCND1*) & 8q34 (*MYC*) gains were associated with tumours who metastasized even in multivariable analysis	(Gerami *et al.*, 2011)
104 frozen MM	qRT-PCR (*MITF*)	DSS (univariable analysis)	*MITF* gain (>3 copies) associated with DSS	(Ugurel *et al.*, 2007)
20 FFPE primaries	Metaphase CGH	OS (univariable analysis)	1q or 6p gains were associated with worse OS	(Namiki *et al.*, 2005)
144 FFPE primaries	FISH (6p25, 6q23, 11q13)	DFS/DSS	Gain in all three loci was associated with worse survival rates (remained independent predictors in multivariable analysis)	(North *et al.*, 2011)
20 FFPE primaries	aCGH	Good prognosis (median survival 14.5 years) vs bad prognosis group (median survival 4 years)	Chromothripsis & focal amplifications were associated with the bad prognosis group	(Hirsch *et al.*, 2013)
160 TMA MM	FISH	OS (5-year follow up)	*MITF* gain associated with worse prognosis (not significant when analysis adjusted for Breslow (missing for some data))	(Garraway *et al.*, 2005)

Samples	Methods	Outcome measure	Result	Reference
75 FFPE primaries	MLPA (3 probes for *CDKN2A*, 1 for *MTAP*)	Relapse status (univariable analysis)	*CDKN2A* loss associated with relapse status	(Conway *et al*., 2010)
81 fresh primaries	Interphase FISH-7q12 (*EGFR*)	Relapse status Survival status (within 5 years)	*EGFR* gain was associated with relapse status but not associated with patients who survived within 5 years versus those who did not	(Rakosy *et al*., 2007)
94 FFPE primaries	FISH	DFS/OS	*EGFR* gain was not associated with DFS/OS however associated with thicker tumours & positive SNB	(Boone *et al*., 2011)
18 frozen primaries vs 41 frozen MM	FISH	Progression	Chromosome 7 gain was associated with progression	(Udart *et al*., 2001)
21 FFPE melanomas in situ vs 21 invasive vs 21MM	Colorimetric ISH-1q12 microsatellite III DNA	Progression	Chromosome 1 gain was associated with progression	(Lee *et al*., 2001)
4 pairs of primary and MM tumours and 1 pair of a patient-derived cell line	SNP array	Progression	This study provides evidence for a non-linear model for melanoma progression through identification of HD found in primaries and not in matched MM	(Swoboda *et al*., 2011)

1.9.2 Gene expression profiling

With the development of DNA microarrays, several studies have been performed to identify gene expression signatures of prognostic significance in melanoma. A number of them showed that overexpression of genes involved in DNA repair pathways or cell cycle/division correlated with tumour progression and poorer outcome (Jaeger *et al*., 2007, Kauffmann *et al*., 2008, Winnepenninckx *et al*., 2006, Ryu *et al*., 2007, Jonsson *et al*., 2010), and overexpression of genes involved in cell adhesion and invasion correlated with tumour progression (Busam *et al*., 2005, Haqq *et al*., 2005, Jaeger *et al*., 2007, Mandruzzato *et al*., 2006, Ryu *et al*., 2007, Talantov *et al*., 2005). Other studies showed that overexpression of immune-related genes are associated with better outcome (Jonsson *et al*., 2010, Mandruzzato *et al*., 2006). Most of these studies have used small numbers of frozen tumours or cell lines. The introduction of the gene expression profiling technology using FFPE samples made it possible to study the biology of melanoma. The Leeds group has used the cDNA-mediated Annealing, Selection, extension, and Ligation (DASL) microarray cancer chip (502 genes), which is specifically designed for FFPE tissue in a large dataset, and showed that increased expression of osteopontin and DNA repair genes (*RAD51*, *RAD52*, *TOP2A*) is associated with worse relapse-free survival and overall survival (Conway *et al*., 2009, Jewell *et al*., 2010). In Chapter 4, I report the use of the whole genome DASL assay and describe an attempt to integrate copy number with gene expression data as a means to corroborate the results derived from the copy number data.

As technology develops, RNA-sequencing where cDNA (synthesized from RNA) is sequenced using NGS is likely to become the lead technology to identify altered gene expression levels associated with prognosis as it offers the advantage to detect alternative splice variants and transcripts from gene fusion events compared to traditional microarray approaches (Meyerson *et al*., 2010). In Chapter 2, I describe the successful sample preparation (cDNA libraries) from matched frozen and FFPE cell line melanoma samples as a means to evaluate the feasibility to use FFPE samples for RNA-sequencing.

1.10 Biomarkers in serum and plasma

As discussed in section 1.7.2.2, serum LDH is elevated in advanced stage patients and is the only serum marker which has been included in the AJCC staging system (Balch *et al*., 2009). Serum S100B and melanoma inhibiting activity (MIA), both melanocyte lineage/differentiation antigens produced by melanoma cells have been implemented in the clinic to monitor disease progression in many European countries but not in the UK (Gogas *et al*., 2009, Palmer *et al*., 2011). S100B was reported to be associated with worse overall survival in a meta-analysis of 22 studies, however there were issues related to different assays and different cut-off values in those different studies (Mocellin *et al*., 2008). A study comparing S100B, LDH, MIA and albumin has revealed that S100B was superior to the other markers tested to detect newly diagnosed metastases while MIA provided no additional benefit over the other markers (Krahn *et al*., 2001). The use of S100B and MIA have been studied as early indicators of tumour progression or relapse, but neither showed prognostic significance in early stage melanoma patients (Stahlecker *et al*., 2000, Acland *et al*., 2002). A number of other potential serum biomarkers have been tested but none has been found to be superior to LDH and S100B or add prognostic information in early stage disease patients (Gogas *et al*., 2009, Palmer *et al*., 2011). In the future, serum proteomic profiling may serve as a means to identify new prognostic biomarkers in melanoma. Serum Amyloid A (expressed in response to inflammatory stimuli) came up as a potential prognostic marker from such a study in stage I-III disease (Findeisen *et al*., 2009). Recently, osteopontin was measured in plasma samples from melanoma patients to identify its prognostic value (Kluger *et al*., 2011, Maier *et al*., 2012). These studies are reviewed in Chapter 6, where I also report a study designed to identify if plasma osteopontin concentrations are of prognostic value in stage I-III patients from the Leeds Melanoma Cohort, which was recently published (Filia *et al*., 2013).

1.11 Treatment of melanoma

Patients with no evidence of metastasis (stage I-II) undergo surgical removal of the primary lesion. For patients with evidence of nodal disease (stage III) after surgical removal of the primary lesion complete lymphadenectomy is performed. For patients with thin tumours (stage I-IIA) prognosis is excellent (Balch *et al*., 2009). However patients with thick tumours and nodal disease (stage IIB-III) have a high

risk of relapse (5-year survival rates 24-67.4%) (Table 1-1) and they may be counselled about adjuvant treatment within a clinical trial. Adjuvant treatment by definition is designed to reduce the risk of relapse. For patients with stage IV disease the 5-year survival rate is much lower (6.7-18.8%) (Table 1-1) and chemotherapy and radiotherapy treatment has been used with very few long-term survivors (Lorigan *et al*., 2008). More recently, more effective drugs have been tested in stage IV patients including BRAF inhibitors, and immunotherapies such as antibodies to co-inhibitory molecules such as CTLA4 and PDL-1 (Eggermont and Robert, 2012).

1.12 Adjuvant therapies

Most of the adjuvant therapies have been trialled in melanoma patients with not very promising results (Molife and Hancock, 2002, Eggermont *et al*., 2009) except for IFN-α and Ipilimumab (anti-CTLA4) (Davar *et al*., 2012, Eggermont and Robert, 2011). Anti- PDL-1 treatment is currently being evaluated.

Interferons (IFN) inhibit proliferation, but also enhance immune recognition of tumour cells (Davar *et al*., 2012). IFN-α is an FDA-approved drug which is used in the US and in Europe but not in the UK as the toxicities are perceived to outweigh the small survival benefit reported in meta-analyses (Mocellin *et al*., 2010).

Expression of the cytotoxic T-lymphocyte-associated antigen 4 (CTLA4) causes inhibition of T cell activation and proliferation (Eggermont and Robert, 2011). In the adjuvant setting, a phase-II clinical trial reported survival benefit in resected stage III/IV patients (Sarnaik *et al*., 2011) while an EORTC phase-III clinical trial is on-going to further assess the use of Ipilimumab in high-risk stage III patients (Eggermont and Robert, 2011).

Identification of new adjuvant therapies is needed and also, biomarkers that could predict benefit from them. Vitamin D is a potential new adjuvant therapy on the basis of data published that low vitamin D levels at diagnosis are associated with thicker tumours and poorer prognosis even if data are adjusted for thickness (Newton-Bishop *et al*., 2009) and predictive biomarkers could identify patients that could benefit from it. In the next sections I give an overview of what is known about the role of vitamin D in melanoma to justify why it is at least a potential candidate as an adjuvant therapy.

1.12.1 Vitamin D

Vitamin D is a fat-soluble hormone which is essential to human health; it plays a role in calcium and phosphate intestinal absorption and transport and bone homeostasis (Lieben and Carmeliet, 2013). Vitamin D is synthesized in the skin in response to UVB irradiation (sun exposure) (Holick, 2003).

The importance of sunlight and resultant vitamin D in human health was well illustrated when vitamin D deficiency was recognised to result in rickets, a disease which affects bone development in children (Holick, 2004). Rickets was very common among children during the 19th century which resulted in vitamin D food fortification. But after World War II the absence of monitoring of the fortification process led to hypercalcemia (vitamin D intoxication) in children and as a result vitamin D fortification of dairy products was banned in many European countries (Holick, 2004). In recent years, there has been an increase in cases of rickets especially in dark-skinned people who live in less sunny countries which has fuelled the interest to study the effect of vitamin D in bone health and in health in general (Field and Newton-Bishop, 2011).

Epidemiological studies, reviewed by Holick and Field and Newton-Bishop (Holick, 2004, Field and Newton-Bishop, 2011), have reported evidence that higher vitamin D levels are associated with the risk of cardiovascular disease, autoimmune diseases, and cancers. Experimental *in vitro* and *in vivo* data, reviewed by Mullin and Dobs and Deeb *et al* (Mullin and Dobs, 2007, Deeb *et al*., 2007), support the view that vitamin D has also anti-tumour effects through regulation of proliferation, angiogenesis and induction of apoptosis. These effects have been studied in many cancers but data are limited in melanoma.

In this section I will discuss the role of vitamin D and epidemiological studies designed to identify an association between vitamin D and melanoma.

1.12.1.1 Metabolism

The active form of vitamin D ($1\alpha,25(OH)_2D_3$) comes from the diet but mostly is synthesized in the skin in response to sun exposure. The vitamin D_2 and vitamin D_3 forms also come into blood circulation from the intestinal absorption of food (predominantly oily fish but also eggs and fortified foods). The vitamin D_3 form is also synthesised though conversion of 7-dehydrocholesterol to pre-vitamin D_3 followed by conversion to vitamin D_3 in the skin when exposed to sunlight (Holick, 2003). Vitamin D_3 is bound to vitamin D binding protein and transported to the liver

where vitamin D_3 is converted to 25(OH)D_3 through 25-hydroxylase (coded by *CYP27A1*) (Deeb *et al.*, 2007). 25(OH)D_3 is then transported to the kidney, and the active form of vitamin D (1α,25$(OH)_2D_3$) is released by 1α-hydroxylase (coded by *CYP27B1*) (Deeb *et al.*, 2007). The negative regulation of synthesis of the active form is mediated though 24-hydroxylase (coded by *CYP24A1*) which converts 25(OH)D_3 and 1α,25$(OH)_2D_3$ to their inactive forms and through the repression of parathyroid hormone by 1α,25$(OH)_2D_3$, which normally promotes the synthesis of the active form (Deeb *et al.*, 2007) (Figure 1-3).

1.12.1.2 Dual role of 1α,25$(OH)_2D_3$

The genomic effect of 1α,25$(OH)_2D_3$ is mediated through binding to the heterodimer vitamin D receptor (VDR)-retinoid X receptor (RXR) complex. The 1α,25$(OH)_2D_3$-VDR-RXR complex binds to vitamin D response elements (VDREs), sequences located in the promoter of target genes to promote their transcriptional activation or repression (Figure 1-4) (Campbell *et al.*, 2010).

1α,25$(OH)_2D_3$ has also a rapid, non-genomic effect through regulation of calcium channels and activation of cytoplasmic kinases such as MAPK (Campbell *et al.*, 2010, Haussler *et al.*, 2011). The activation of pathways through the rapid mechanisms of 1α,25$(OH)_2D_3$ has been postulated to cross-interact with other pathways to affect gene transcription (Figure 1-4) (Deeb *et al.*, 2007, Haussler *et al.*, 2011).

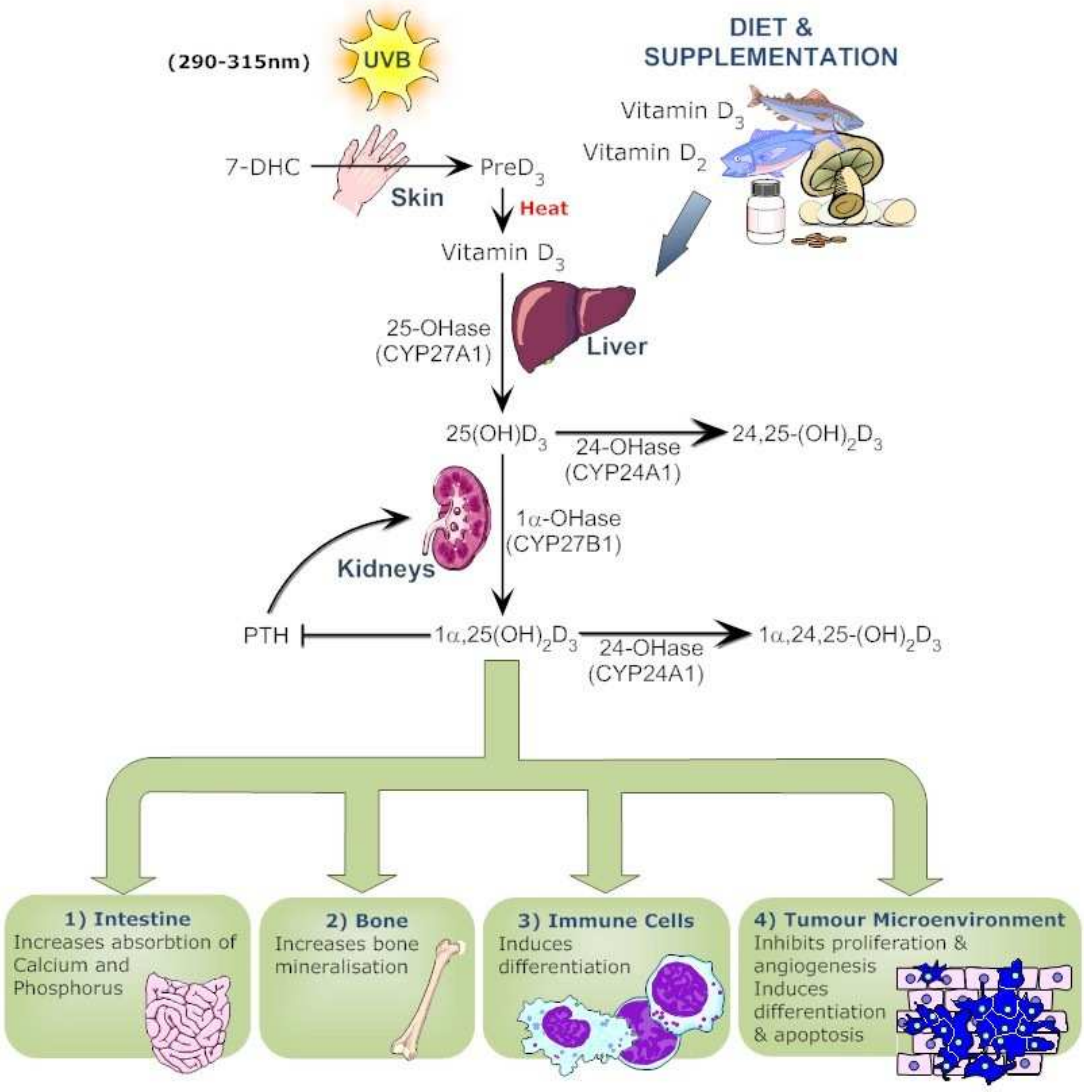


Figure 1-3: Vitamin D metabolism and biological effects.

The vitamin D_3 form is synthesised in the skin when exposed to sunlight. The vitamin D_2 and vitamin D_3 forms come into blood circulation from the intestinal absorption of food. Vitamin D_3 is transported to the liver where vitamin D_3 is converted to $25(OH)D_3$ through 25-hydroxylase. $25(OH)D_3$ is then transported to the kidney, and the active form of vitamin D ($1\alpha,25(OH)_2D_3$) is released by 1α-hydroxylase. The negative regulation of synthesis of the active form is mediated though 24-hydroxylase which converts $25(OH)D_3$ and $1\alpha,25(OH)_2D_3$ to their inactive forms and through the repression of parathyroid hormone by $1\alpha,25(OH)_2D_3$, which normally promotes the synthesis of the active of form. The main effects of $1\alpha,25(OH)_2D_3$ on different tissues are shown. Adapted from (Deeb *et al.*, 2007, Field and Newton-Bishop, 2011).

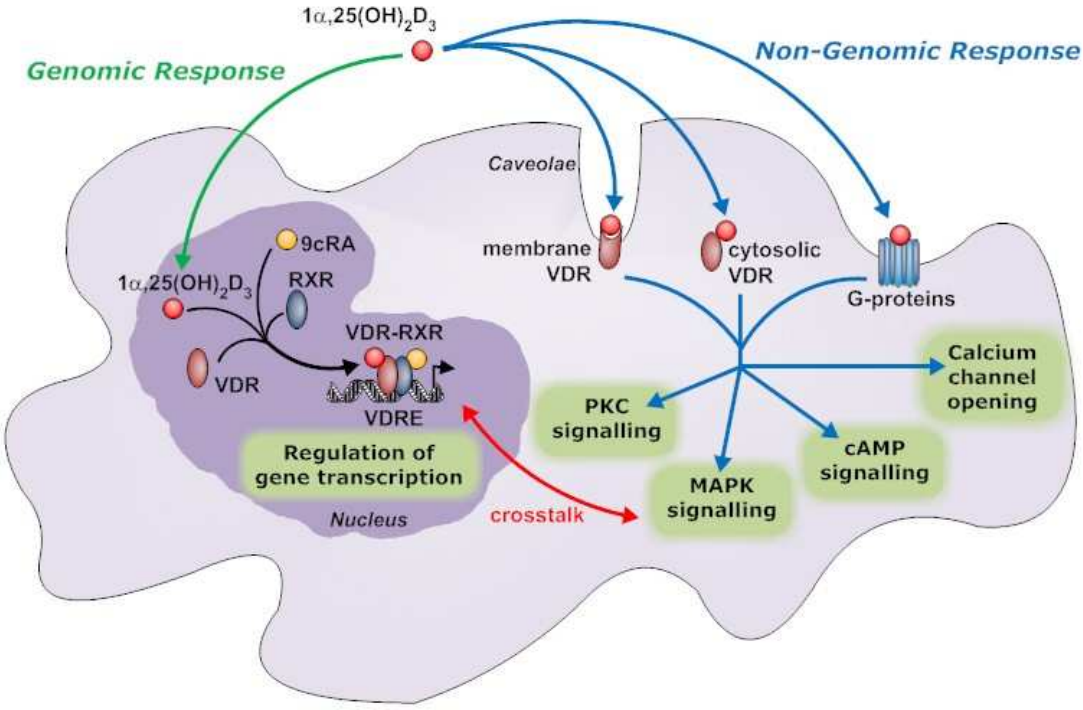


Figure 1-4: Transcriptional regulation in response to vitamin D_3.

The genomic effect of 1α,25$(OH)_2D_3$ is mediated through binding to the heterodimer VDR-RXR complex. The 1α,25$(OH)_2D_3$-VDR-RXR complex binds to vitamin D response elements (VDREs), sequences located in the promoter of target genes to regulate transcription. The non-genomic effect of 1α,25$(OH)_2D_3$ is postulated to be mediated through an alternative membrane or cytosolic VDR or through binding to G-coupled proteins which promote regulation of calcium channels and activation of cytoplasmic kinases such as MAPK. The activation of pathways through the non-genomic mechanisms of 1α,25$(OH)_2D_3$ may cross-interact with the classical genomic pathway to affect gene transcription.

1.12.1.3 The epidemiology of vitamin D and cancer

There has been an interest in the role of vitamin D and cancer risk for many years, in particular following an observation of lower cancer risks in individuals living in areas of the USA with higher sun exposure than those living in areas with lower exposure (Garland and Garland, 1980). That vitamin D has been shown to be anti-proliferative *in vitro* fuelled this interest.

Vitamin D levels are measured as serum $25(OH)D_2$ or $25(OH)D_3$ by most laboratories as the active form, $1\alpha,25(OH)_2D_3$, has a short half-life (<4 hours). The relationship between serum levels and cancer risk or survival in different cancers has been extensively studied but data in melanoma are limited. A meta-analysis of seven studies performed by Chen *et al* reported an association of lower serum $25(OH)D_3$ levels and increased risk of breast cancer (Chen *et al*., 2010). An association was also reported between higher serum $25(OH)D_3$ levels and decreased risk of colorectal cancer (Jenab *et al*., 2010) and better prognosis (Mezawa *et al*., 2010).

1.12.1.4 Vitamin D and susceptibility to melanoma

There is no doubt that sun exposure is associated with melanoma susceptibility, although this relationship has proven complex (Gandini *et al*., 2005b). Intermittent sun exposure related to sunburns but not cumulative sun exposure has been reported as a risk factor for melanoma (Newton-Bishop and Randerson-Moor, 2013, Field and Newton-Bishop, 2011). It has also been reported that occupational sun exposure and regular weekend sun exposure have even shown a protective effect (Gandini *et al*., 2005b). The observation of a possible protective effect of regular sun exposure suggests that regular moderate sun exposure might be protective because of increased levels of vitamin D or photoadaptation, related to a decreased risk of sunburns (Newton-Bishop and Randerson-Moor, 2013, Field and Newton-Bishop, 2011). The reported association between risk and genetic polymorphisms, such as the VDR SNPs, have also shown some support for a relationship between vitamin D levels and risk (Randerson-Moor *et al*., 2009, Newton-Bishop and Randerson-Moor, 2013). Further support for that relationship is provided by biological studies showing that vitamin D has anti-proliferative effects on melanoma cells *in vitro* (Table 1-4). The anti-proliferative and other anti-tumour effects of vitamin D on melanoma and other cancer cells are further discussed in section 1.12.2.

1.12.1.5 Vitamin D and outcome from melanoma

An association between higher serum vitamin D levels and better survival from melanoma and thinner tumours exists as reported by the Leeds group (Newton-Bishop *et al*., 2009). This association was supported by a subsequent small study which showed that vitamin D levels are protective in stage IV melanoma patients (Nurnberg *et al*., 2009). The lack of large cohort studies with stored serum samples means that such associations have not been validated to date. The concern around these data is that thinner, healthier and more active people have higher vitamin D levels suggesting that high vitamin D levels might depict healthier life styles generally (Field and Newton-Bishop, 2011), and that in the 2009 Leeds study higher vitamin D levels might be acting as marker of healthier lifestyles rather than being causally related to melanoma survival. Biological studies have been performed however and demonstrate that vitamin D has anti-tumour effects *in vitro* and *in vivo* but data on melanoma models are limited (Deeb *et al*., 2007). As further evidence from such studies become available clinical trials could be set up to define the right vitamin D dose and identify the effect that vitamin D supplementation could have as an adjuvant therapy for melanoma. To date there are only two clinical trials currently recruiting patients. Understanding the anti-tumour effects of $1\alpha,25(OH)_2D_3$ in melanoma could also lead to the identification of predictive biomarkers that would show likely benefits to vitamin D supplementation.

1.12.2 Anti-tumour effects of $1\alpha,25(OH)_2D_3$

The anti-tumour effects of the active form of vitamin D $1\alpha,25(OH)_2D_3$ or other non-calcaemic analogues have been reported in different cancers in *in vitro* and *in vivo* models (reviewed by (Deeb *et al*., 2007, Fleet *et al*., 2012). It is reported that $1\alpha,25(OH)_2D_3$ plays a role in proliferation, apoptosis, angiogenesis, autophagy and immune response pathways. The diverse role of $1\alpha,25(OH)_2D_3$ in melanoma still needs further elucidation as studies have been limited to its anti-proliferative role (summarized in Table 1-4).

Gene expression experiments after $1\alpha,25(OH)_2D_3$ treatment have been performed in different cancers to confirm the anti-tumour effects but also identify the mechanisms of these effects. These studies have identified a number of $1\alpha,25(OH)_2D_3$ target genes (reviewed by (Deeb *et al*., 2007, Fleet *et al*., 2012). Some of the reported $1\alpha,25(OH)_2D_3$ target genes contained VDREs in their promoters but others did not (Deeb *et al*., 2007, Osborne and Hutchinson, 2002), suggesting a VDR independent action. Supportive evidence of a VDR independent

action was shown in that 1α,25$(OH)_2D_3$ continued to have an anti-proliferative effect when VDR was knocked down in breast cancer cells (Costa *et al*., 2009).

Table 1-4: Summary of *in vitro* and *in vivo* experiments to assess potential anti-tumour effects of 1α,25$(OH)_2D_3$ or other vitamin D-analogue treated melanoma cells compared to untreated cells.

Cell line/xenograft	Effect	Reference
MM96	Growth inhibition	(Frampton *et al*., 1983)
RPMI 7932 xenograft	Unaffected growth	(Eisman *et al*., 1987)
COLO 239F xenograft	Growth inhibition	(Eisman *et al*., 1987)
SkMel28	Growth inhibition	(Evans *et al*., 1996, Reichrath *et al*., 2007)
RPMI7951	Growth inhibition	(Evans *et al*., 1996)
WM1341	Unaffected growth/small induction of apoptosis	(Danielsson *et al*., 1998, Danielsson *et al*., 1999)
MeWo	Growth inhibition/no induction of apoptosis	(Danielsson *et al*., 1998, Danielsson *et al*., 1999, Seifert *et al*., 2004, Reichrath *et al*., 2007)
ME18	Growth inhibition	(Gruber and Anuszewska, 2002)
SkMel5	Unaffected growth	(Seifert *et al*., 2004, Reichrath *et al*., 2007)
SM	Growth inhibition	(Reichrath *et al*., 2007)
IGR	Unaffected growth	(Reichrath *et al*., 2007)
MelJuso	Unaffected growth	(Reichrath *et al*., 2007)
SkMel188	Growth inhibition	(Slominski *et al*., 2011, Li *et al*., 2010, Janjetovic *et al*., 2011)
A375	Growth inhibition	(Ishibashi *et al*., 2012)

1.12.3 Gene expression studies in response to vitamin D in melanoma cells

Functional and gene expression studies in response to 1α,25$(OH)_2D_3$ have been relatively limited in melanoma and here I give an overview of these studies. Seifert *et al* have demonstrated that gene expression *of VDR* and *CYP24A1,* coding for the vitamin D degradation protein*,* is increased in response to 1α,25$(OH)_2D_3$ in vitamin D-sensitive cells but not in vitamin-D resistant cells (Seifert *et al*., 2004), indicating a VDR-mediated effect. Evidence that Peroxisome Proliferator-Activated Receptor

(PPAR) signalling decreases the metastatic potential of melanoma cells (Grabacka *et al*., 2006), and the fact that both VDR and PPAR form a heterodimer with RXR, has led to the investigation of PPAR signalling when cells are treated with 1α,25$(OH)_2D_3$ (Sertznig *et al*., 2009b). Sertznig *et al* showed that *VDR* and *PPARα* gene expression is increased after 1α,25$(OH)_2D_3$ treatment in vitamin D-sensitive cells compared to the resistant cells (Sertznig *et al*., 2009b) suggesting a potential activation of both the VDR and PPAR signalling pathways in response to vitamin D. A further study by the same group showed an increase in expression of the 1α,25$(OH)_2D_3$-induced *CYP24A1* gene in MeWo vitamin D sensitive cells when treated with both 1α,25$(OH)_2D_3$ and PPAR ligand DHA compared to vehicle-treated cells. This induction of *CYP24A1* was significantly lower compared to cells treated with 1α,25$(OH)_2D_3$ alone (Sertznig *et al*., 2009a). They suggested a potential cross-talk between the VDR and PPAR signalling pathways but not at the level of cell proliferation, as similar levels of reduction in cell proliferation were seen in response to 1α,25$(OH)_2D_3$ alone, or in combination with PPAR ligand DHA (Sertznig *et al*., 2009a). Further studies need to be done to investigate the potential cross-talk between these two signalling pathways. Sertznig *et al* also showed that 1α,25$(OH)_2D_3$ might have a pro-apoptotic action when RNA and protein expression of *CLU* (Clusterin), a pro-apoptotic gene, was increased in response to 1α,25$(OH)_2D_3$ in vitamin D sensitive cells compared to resistant cells (Shannan *et al*., 2006). Another study suggested an anti-tumour effect of 1α,25$(OH)_2D_3$ through repression of cytokine IL-8 at transcriptional level, mediated through NF-κB in melanoma cell line A3 (Harant *et al*., 1997). In 1988, a study showed that mRNA and protein levels of HLA-class II *DRa* are downregulated after 1α,25$(OH)_2D_3$ treatment, showing a role for 1α,25$(OH)_2D_3$ in differentiation and immune response in melanoma cell lineages (Carrington *et al*., 1988). A more recent study published that 1α,25$(OH)_2D_3$ mediates its anti-tumour effects in vitamin D resistant cells through the immune system, with enhancement of Fas expression on the surface of melanoma cells and increase of caspase sensitivity to Fas-mediated apoptotic pathway by natural killer (NK) cells (Lee *et al*., 2011). Interestingly, a synergistic effect of 1α,25$(OH)_2D_3$ and trichostatin A (a demethylating agent) was reported in vitamin D resistant cells IGR, suggesting that an epigenetic mechanism might be involved in vitamin D resistance (Essa *et al*., 2010). In the same study it was reported that micro-RNA mir-125b expression was inversely correlated with *VDR* expression in vitamin D sensitive cells (Essa *et al*., 2010). It has been shown that VDR expression is regulated by mir-125b at a post-transcriptional level (Mohri *et al*., 2009).

Supportive data that vitamin D has anti-tumour effects through different pathways such as proliferation, apoptosis and immune response has been shown in other cancers, reviewed by (Deeb *et al*., 2007). The vitamin D effect in cancer cells generally and specifically in melanoma cells however appears diverse, based on the above studies, and the information derived generally from a candidate approach to understanding the biological events rather than an agnostic one. In Chapter 5 I will describe my approach to validate previous findings, and possibly identify novel mechanisms of vitamin D anti-tumour effects by generating whole-genome gene expression data from two vitamin D responsive melanoma cell lines cultured with and in the absence of vitamin D.

1.13 The use of FFPE tumour tissue for genomic and transcriptomic analysis

As previously discussed primary melanomas are small and routinely fixed in formalin. Formalin-fixed tumours are then embedded in paraffin wax blocks to allow sections to be taken for staining and histological examination to assess features of prognostic significance, such as Breslow thickness and mitotic rate. Formalin is a fixative which causes chemical modifications of proteins and other tissue constituents to preserve the tissue in a close approximation of the *in vivo* state (Frankel, 2012, Srinivasan *et al*., 2002). The negative effects of formalin which affect the quality and quantity of nucleic acids have well been described: formation of cross-links between nucleic acids and proteins; addition of methylol groups in adenine and thymine in particular; random breaks in the nucleotide sequence (Frankel, 2012, Srinivasan *et al*., 2002). For these reasons the use of FFPE blocks for genomics has been limited in the past. Some of the effects of formalin are reversible. Cross-links are reversible through thorough proteinase K digestion and therefore nucleic acid yields can be improved although they will be lower than those from frozen tissue (Frankel, 2012). The addition of methylol groups in adenine in particular can prevent effective reverse transcription of RNA caused by alteration of the poly-A tail but the use of random instead of oligo(dT) primers in cDNA synthesis may overcome this problem (Frankel, 2012, Srinivasan *et al*., 2002). Nucleic acids fragmentation remains an issue, though. The development of new techniques is enabling the study of fragmented nucleic acids. This thesis reports use of NGS technology for DNA sequencing to identify CNAs and DASL assay to assess gene expression in FFPE melanoma samples. DASL is specifically designed for use with

degraded RNA and has been shown to produce reproducible results (Bibikova *et al*., 2004, Ravo *et al*., 2008, Conway *et al*., 2009). NGS technology utilises DNA sequencing of short fragments and recent studies have reported the successful generation of NGS data using low yields of DNA extracted from FFPE tumour samples (Wood *et al*., 2010, Belvedere *et al*., 2012). Using whole-genome NGS, Wood *et al* and Schweiger *et al* have reported reproducible CNAs between FFPE and frozen tumours (Schweiger *et al*., 2009, Wood *et al*., 2010). These results are promising but further investigation is needed to evaluate the feasibility of NGS technology to be used for the identification of prognostic and predictive biomarkers in melanoma using stocks of FFPE samples from patients with long-term follow-up clinical data. In Chapters 3 and 4, I report such a study.

1.14 Outline of thesis chapters

To address the major aims of the thesis (section 1.1), I will briefly outline the next chapters of this thesis:

Chapter 2 describes patients, samples, cell lines and methods used throughout the thesis. It describes techniques of tumour sampling and methodological work to assess the use of new technologies using FFPE samples. The results from the authentication of the cell lines used are also presented.

Chapter 3 presents the evaluation and optimisation of different protocols for sample preparation for whole-genome NGS analysis. Chapter 4 describes the feasibility to use whole-genome NGS using FFPE primary melanomas. It includes an assessment of quality control measures to predict success of sample preparation and assay performance. An attempt to identify copy number alterations associated with outcome and clinico-pathological characteristics known to predict survival but also *BRAF/NRAS* mutation status is also presented. Finally, an assessment of NGS and gene expression data integration is described.

Chapter 5 reports the effect of vitamin D on the growth of cultured melanoma cells. Immunocytochemistry experiments to assess the expression of vitamin D receptor and melanocyte markers in the cell lines used are described. This chapter also presents whole-genome gene expression data using microarrays technology. Differentially expressed genes between vitamin D treated cells and controls are identified and gene ontology and pathway analysis using the gene expression data generated is assessed.

Chapter 6 reports osteopontin plasma concentrations in patients with melanoma to identify their prognostic significance. The validation of the enzyme-linked immunosorbent assay (ELISA) used to measure plasma levels of osteopontin is described.

Finally, Chapter 7 summarises the main findings of my thesis, discusses limitations and describes future work.

Chapter 2
Methodology

2.1 Aims

Chapter 2 aims to describe:

- The studies which contributed patient samples and clinical data;
- The reasons of tissue microarray needle usage to sample the tumours
- The cell lines studied;
- The results of cell line authentication and assessment of their mutation status;
- The quality control assessment of WG-DASL HT data derived from FFPE samples to ensure that only high quality data are used in work described in Chapter 5;
- The successful preparation of samples for RNA-sequencing although the sequencing data are not presented in this thesis;
- The methods used throughout the thesis (generally statistical analysis relevant only to one chapter will be given within that chapter).

2.2 Studies contributing samples and clinical data

2.2.1 The Leeds Melanoma Cohort

The Leeds Melanoma Cohort is a population-based cohort of patients with melanoma predominantly from a geographically defined area: the Yorkshire and Northern region of the UK. In all, 2180 patients have been recruited to this study since 2001 and the Leeds group are in the process of retrieving and sampling primary tumours from all participants who gave consent to use of their tissues. The study was designed to identify genetic and environmental determinants of survival from melanoma.

Patients have been recruited within 6 months of diagnosis wherever possible (Newton-Bishop *et al.*, 2009). Detailed questionnaires have been filled in to collect

lifestyle indicators and lifetime environmental exposure data. Plasma, serum, DNA and lymphocytes have been stored from blood samples collected at the time of recruitment.

Research nurses within our group also collect blood samples and follow-up data from annual patient questionnaires and surveys completed by their general practitioners every 2 years. Participants in the study have also given consent to sample formalin-fixed paraffin-embedded (FFPE) primary tumour blocks stored in the referring hospitals.

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). A significant amendment was sought from the Northern and Yorkshire Research Ethics Committee (01/3/057) for using the plasma samples for biomarker discovery.

In Chapter 6 plasma samples from patients who relapsed after recruitment and samples from patients who haven't were selected to test the prognostic significance of osteopontin plasma levels. In Chapter 4 FFPE primary blocks from patients with tumours thicker than 0.75mm who have died from melanoma and patients who have not relapsed for at least 5 years were selected to identify if DNA copy number alterations are associated with survival.

2.2.2 Melanoma metastasis samples

Melanoma metastasis FFPE blocks from 11 patients were used for nucleic acid extraction to assess a number of assays throughout my thesis. These samples were used in the methodological work described in Chapter 3. Primary melanoma tumours are usually small and therefore we avoided using them in methodological assessments. In contrast metastasis samples are usually larger allowing more tissue to be used. Also, new targeted therapies have emerged for melanoma making primary FFPE blocks precious material which need to be kept for future clinical testing.

These samples were obtained from the Leeds Teaching Hospitals Trust pathology department (Dr Sara Edward) from patients who had given written consent to the use of their tissue samples at the time of surgery. Also, a number of samples from patients recruited in the Predicting Response to Chemotherapy in Malignant Melanoma study were used for methodological assessments. Ethical approval for this study has been granted by the Yorkshire and Humber Central Research Ethics

Committee (10/H1313/72) and the National Information Governance Board for Health and Social Care (formerly PIAG) (ECC 8-02 (FT2)/2010).

2.2.3 EDTA plasma and serum samples for validation studies

In the early phases of the Leeds Melanoma Cohort study, a standard approach to preparing blood samples was developed. Samples arrive in the lab at varying intervals, as some are taken in St James's and can be delivered within an hour and others are mailed to the laboratory. All samples are therefore processed for storage as near to 48 hours after the blood is taken with a proportion processed up to 4 days after the blood is taken. The assays developed to study these samples must therefore be shown to be stable when samples are left "on the bench" for the maximum period of 4 days, and thus validation of all test assays must be carried out both to look at the assay itself and how stable the protein/constituent is on the bench.

Blood samples were collected for this purpose from 5 patients with melanoma and 4 healthy volunteers in two 4ml EDTA-containing anticoagulant collection tubes from each person. Also, stored EDTA plasma and matched serum samples from 4 patients with renal cell carcinoma were kindly provided by Professor Rosamonde Banks. In all cases plasma was separated by centrifugation of blood samples at 1,500*g* for 15 minutes, prior to storage in aliquots at -80^{o}C.

Patients and healthy volunteers have given informed consent and sampling was done under approval of the Local Research Ethics Committee (LREC) (07/Q1206/28).

All samples were used in the validation assays of the osteopontin immunoassay described in Chapter 6.

2.3 Management of FFPE tissue

A standard operating procedure has been developed within our group on how to manage and sample tissue from FFPE blocks. Our Human Tissue Act Manager, Sandra Tovey was responsible for tracing the FFPE blocks. The blocks are mailed to our lab, sectioned, stained, reviewed for sampling, sampled wherever possible

and finally returned. Sectioning, staining and sampling were performed by skilled histopathologists: Dr Filomena Esteves and Dr Jonathan Laye.

2.3.1 Tissue sectioning

The blocks were mounted onto a microtome and 5 micron sections were cut allowing the sections to float in a water bath set at 45^{o}C until a clean section was achieved. One section from each block then mounted on a Superfrost glass slide (Solmedia, UK) and placed in an oven at 36^{o}C overnight to dry. The tissue was fixed by putting the slide on a heating block at 60^{o}C for 20 minutes with the tissue facing upwards.

2.3.2 Haematoxylin and Eosin (H&E) staining of slides

All sections were stained using Mayer's Haematoxylin and 1% Eosin to allow identification of an area for sampling after review.

Firstly, slides were de-waxed by immersing slides in xylene in a glass slide bath for 5 minutes. This step was repeated twice by immersing slides in different water baths containing fresh xylene. The slides were drained on tissue paper and a dehydration step followed by immersing slides in absolute ethanol for 2 minutes, repeating twice. The slides were transferred to baths containing 90% ethanol for 2 minutes and 75% ethanol for 2 further minutes. Then they were placed under running water for 30 seconds to 1 minute. For staining, the slides were immersed in Mayer's haematoxyin for 2-3 minutes, moved to a fresh container and then placed under running water for 1 minute. Next they were immersed in Scott's Tap water for 1 minute, rinsed in tap water for 1 minute, immersed in eosin stain for 3 minutes, rinsed under running water for 30 seconds to 1 minute and left to drain on tissue paper. The slides were then rehydrated after immersion in absolute ethanol for 1 minute, repeating two more times. They were air-dried for 1 minute, immersed in xylene for 1 minute, and then transferred to fresh xylene for a further minute. A drop of Depex mounting medium (VWR International, UK) was placed onto a glass cover slip and the slide was lowered onto the cover slip tissue side down, taking care not to introduce air-bubbles between the slide and cover slip. The slides were left in a fume hood overnight to dry before proceeding to review.

2.3.3 Slide review and tumour sampling

H&E slides were reviewed by Professor Julia Newton Bishop and Dr Jonathan Laye using a double headed microscope to identify an area of the tumour suitable for sampling (additional details are given in the next section).

A fine tip permanent marker was used to mark the area on the slide. Sampling was performed using Tissue Micro-Array (TMA) Building Apparatus. The needle was cleaned between blocks by taking cores from a blank paraffin block to avoid tumour contamination between blocks. The marked H&E slide was placed over the tumour block. The TMA needle was aligned above the marked spot, the slide removed carefully and the needle guided down to the block. A 0.8 x 2mm core was taken from the block, placed in a 1.5ml microcentrifuge tube and stored at 4^{o}C prior to processing.

2.3.4 Justification for tumour core sampling using a TMA needle

In this work, tumours have been sampled as described above using TMA needles inserted horizontally ideally through the deepest part of the tumour with high tumour cell content and minimal stromal invasion: thus, the intent was to consistently sample the deepest part of the primary tumour most likely to contain least stroma or inflammatory tissue. The considerable variation in the morphology of primary melanomas means that the sample taken might have been close to, and potentially include, contaminating stromal epidermis or dermis.

Where the tumour was too small to allow sampling without removing the majority of the tumour sample (which would potentially prejudice the patient's subsequent access to tissue for clinical biomarker testing) no sample was taken at all.

Other research groups habitually take whole tissue sections from FFPE blocks and some additionally perform laser capture microdissection (LCM). The decision to use a TMA needle was made by our group after considering that:

- TMA sampling leaves the block intact so that the area of sampling can be subsequently reviewed. This also means that the shape of the tumour and its relationship with the stroma in not affected in case clinical re-review of the block is needed;
- Sectioning of the block for LCM wastes tissue as many sections are discarded;
- LCM is a time-consuming method for large-scale studies;

- LCM is more likely to avoid stromal contamination. However, Dr Caroline Conway, a previous PhD student, showed that the tumour content of cores is estimated to be at least 70% (Conway *et al*., 2009). Briefly, horizontally sectioned cores were embedded in wax blocks and blocks were sectioned to obtain 5 micron sections throughout the tissue core. H&E sections prepared from these sections were reviewed using light microscopy and the percentage of tumour content was determined.
- Much of the work within the group aims to identify biomarkers. The group argues that TMA sampling would be appropriate for such work. We have consistently sampled the least inflamed part of the tumour and have not sampled tumour where inflammation was dominant. Using this approach there is a recognisable bias as it is argued that all samples should be sampled for biomarker work. However sampling was avoided when tumour cell content could be <70%.

2.3.5 Melanin score

Some melanoma tumours are highly pigmented while some others are not. Melanin is usually co-purified with nucleic acids from pigmented melanoma tumours. Melanin binds to the DNA polymerase and inhibits downstream reactions such as the polymerase chain reaction (PCR) (Eckhart *et al*., 2000). A melanin score (0-3) was therefore assigned for each tumour core so that technological variation which could be attributed to melanin could be investigated. No melanin content was noted as 0, and high melanin content was noted as 3. This procedure was performed by myself, Dr Rosalyn Jewell, Dr Jonathan Laye, or Minttu Polso before nucleic acid extraction.

2.3.6 Nucleic acid extraction from FFPE tissues

2.3.6.1 Tissue de-paraffinisation

1ml xylene was added in the tubes containing the FFPE tissue cores and incubated at 37^{o}C for 30 minutes with pulse vortexing at regular intervals. The xylene was carefully removed using a pipette and 1ml 100% ethanol was added. The sample was mixed by vortexing, quickly microcentrifuged at 13000rpm and the ethanol was carefully removed using a pipette. A second wash step with 70% ethanol was performed to remove residual xylene. Finally, the tubes were placed at 37^{o}C with an open lid to let the tissue dry completely.

2.3.6.2 Qiagen QiAamp® DNA FFPE tissue kit

A proportion of DNA samples extracted with this kit were used for work described in Chapters 3 and 4. In the course of my work a newer kit was assessed which proved to have higher yields of DNA, and also allowed dual extraction of DNA and RNA, and this kit was subsequently abandoned. Samples were processed according to the supplied protocol.

Briefly, the deparaffinised tissue core was resuspended in 180µl buffer ATL and 20µl proteinase K, vortexed and incubated at 56^{o}C with agitation for up to 48 hours to ensure complete tissue lysis. The lysed sample was then incubated at 90^{o}C for 1 hour. 2µl RNase A (100mg/ml) was then added and the mixed sample was incubated at room temperature for 2 minutes. 200µl buffer AL was mixed with the sample followed by further mixing after addition of 200µl absolute ethanol. The sample was transferred to a QIAamp MinElute column placed in a 2ml collection tube. The column was centrifuged at 6,000*g* for 1 minute. The column was put in a new collection tube, 500µl buffer AW1 added and the column was washed by centrifugation at 6,000*g* for 1 minute. A second wash step was performed after addition of 500µl buffer AW2. The column was transferred to a new collection tube and the column centrifuged at 20,000*g* for 3 minutes to dry the silica-gel membrane. The column was transferred to a new 1.5ml collection tube, 30µl buffer ATE applied to the column, and the column incubated at room temperature for 5 minutes. The DNA was eluted after centrifugation at 16,000*g* for 1 minute and stored at -20^{o}C.

2.3.6.3 Qiagen AllPrep® DNA/RNA FFPE kit

RNA and DNA extraction was performed using the AllPrep® DNA/RNA FFPE kit (Qiagen, Crawley UK) according to the manufacturer's instructions with a few modifications. Using this protocol both RNA and DNA could be extracted using a single tissue core. DNA & RNA extracted using this kit was used for work in Chapters 3 and 4.

2.3.6.3.1 Tissue digestion

The deparaffinised core was resuspended in 150µl buffer PKD. 10µl proteinase K (20 mg/ml) was added and the sample was mixed by vortexing. The sample was incubated at 56^{o}C overnight and then put on ice for 3 minutes to ensure efficient precipitation. The sample was centrifuged for 15 minutes at 20,000*g* and the supernatant transferred to a new tube. The pellet was retained for DNA purification.

2.3.6.3.2 RNA extraction

The supernatant was incubated at 80°C for 15 minutes. 320µl buffer RLT was mixed with the sample, followed by the addition of either 720µl or 1120µl ethanol (1120µl ethanol was added for total RNA extraction that includes microRNAs). 700µl of the mixed sample was transferred to an RNeasy MinElute column, and the column was centrifuged for 15 seconds at 8,000*g*. The flow-through was discarded and this step was repeated until the entire sample had passed through the column. 350µl buffer FRN was then added to the column, the column centrifuged for 15 seconds at 8,000*g* and the flow-through discarded. The DNase incubation mix of 10µl DNase I and 70µl buffer RDD was added to the column and left at room temperature for 15 minutes. Following the addition of 500µL buffer FRN the column was centrifuged for 15 seconds at 8,000*g* and the flow-through was kept for use in the next step. The column was placed in a new collection tube and the flow-through from the previous step was re-applied to the column. The centrifugation step was repeated for 15 seconds at 8,000*g* and the flow-through was now discarded. 500µl buffer RPE was added to the column, the column centrifuged for 15 seconds at 8,000*g* and the flow-through was discarded. The wash step with 500µl buffer RPE was repeated and then the column was transferred to a new collection tube. 26µl RNase-free water was finally applied to the column and incubated at room temperature for 1 minute. The RNA was eluted after centrifugation for 1 minute at 20,000*g* and stored at -80°C.

2.3.6.3.3 DNA extraction

The pellet retained after the tissue digestion described in section 2.3.6.3.1 was resuspended in 180µl buffer ATL and 40µl proteinase K. The mix was incubated at 56°C overnight and then at 90°C for 2 hours. 4µl RNase (100mg/ml) was added and the sample was incubated for 2 minutes at room temperature. 200µl buffer AL was added, the sample mixed thoroughly, and then 200µL ethanol was further mixed with the sample. The entire sample was transferred to a QIAamp MinElute column and the column centrifuged for 1 minute at 8,000*g*. The column was placed in a new collection tube, 700µl buffer AW1 added, the column centrifuged for 15 seconds at 8,000*g*, and the flow-through discarded. The same step was performed using 700µl buffer AW2 followed by 700µl ethanol. Finally, 30µl buffer ATE was applied to the column and left to stand at room temperature for 5 minutes. The DNA was eluted after centrifugation at 20,000*g* for 1 minute and stored at -20°C.

2.3.7 cDNA synthesis of RNA extracted from FFPE tissue

2.3.7.1 The Invitrogen™ High Capacity cDNA Reverse Transcription kit

Production of cDNA from both FFPE RNA and frozen RNA was carried out using the High Capacity cDNA Reverse Transcription kit (Invitrogen™ by Life Technologies, USA). This protocol uses random hexamer primers which is important for successful cDNA synthesis from fragmented RNA. Oligo-dT (Oligodeoxythymidylic acid residues (dT)) primers are usually avoided as they bind to the poly-A tail, which is missing in fragmented RNA derived from FFPE blocks.

In brief, a master mix was prepared by combining 2µl 10x reverse transcriptase (RT) buffer, 0.8µl 25x dNTP (deoxyribonucleotide triphosphate) mix, 2µl 10x RT random primers, 1µl Multiscribe Reverse Transcriptase, 1µl RNase Inhibitor and 3.2µl nuclease-free water. 10µl of master mix were combined with 200ng total RNA (up to 10µl) in a 96-well plate placed on ice, quickly centrifuged and incubated on the Applied Biosystems 9700 Thermal Cycler at 25^{o}C for 10 minutes, 37^{o}C for 120 minutes, and 85^{o}C for 5 minutes. The prepared cDNA was stored at -20^{o}C.

2.4 Mutation screening

BRAF and *NRAS* mutation screening was performed in all tumour samples described in Chapter 4 and in all cell lines used throughout this thesis. The assays were performed by Dr Philip Chambers, Genomics Facility, Leeds, using pyrosequencing.

2.4.1 Pyrosequencing

Pyrosequencing is a sequencing method where DNA polymerase initiates the synthesis of a single-stranded fragment while deoxynucleotides are added sequentially (Ronaghi *et al.*, 1998). When a deoxynucleotide is incorporated pyrophospate is released which is converted to adenosine-5'-triphosphate (ATP) by ATP sulfurylase. Luciferase oxidises luciferin in the presence of ATP and light is generated. The amount of light emitted is proportional to the amount of nucleotides incorporated. The unincorporated nucleotides are degraded by apyrase before the addition of the next nucleotide (Ronaghi *et al.*, 1998).

The assays were performed based on NCBI RefSeq accession number NM_004333 for *BRAF* and NM_002524 for *NRAS (Pruitt et al., 2012)*. Primers for PCR amplification and pyrosequencing assays were designed using proprietary Pyrosequencing assay design software version 2.0.1.15 (Qiagen, Crawley UK). Three PCR reactions were performed; one reaction for *BRAF* codon 600, one for *NRAS* codon 61 and one for both *NRAS* codons 12 and 13 (Table 2-1). Each PCR reaction was prepared by mixing 20ng genomic DNA or 2µl cDNA, 12.5µl HotStar Taq Master Mix (Qiagen, Crawley UK), 2mM magnesium chloride, 200nM each of forward and reverse primers and nuclease-free water to make a final volume of 25µl. Thermal cycling conditions were 94^{o}C for 12 minutes, 40 cycles of 94^{o}C for 10 seconds, 55^{o}C for 20 seconds and 72^{o}C for 20 seconds. The biotinylated PCR product was captured with streptavidin-coated magnetic beads and sequenced on a PyroMark ID system (Qiagen, Crawley UK) according to manufacturer's protocols. Data were analysed by visual inspection of pyrograms and by analysis of peak height data.

Table 2-1: PCR and pyrosequencing primers for *BRAF* and *NRAS* assays

Region of interest	PCR primers	Pyrosequencing primer	Amplicon length (bp)
***BRAF* codon 600**	F:5'-TGAAGACCTCA CAGTAAAAATAGG-3' R: 5'-biotin-TCCAGACAACTGTTCAAACTGAT-3'	5'-TGATTTTGGTCT AGCTACA-3'	91
***NRAS* codon 61**	F: 5'-biotin-GAAACCTGTTTGTTGGACATACTG-3' R:5'-TCGCCTGTCCTCATGTATTG-3'	5'-CTCTCATGGCA CT GTACT-3'	83
NRAS codons 12 & 13	F:5'-CTTGCTGGTGTGAAATGACTGAG-3' R: 5'-biotin-TGGATTGTCAGTGCGCTTTTC-3'	5'-CTGGTGGTGGT TGGA-3'	79

Pyrosequencing assays developed previously have shown higher sensitivity than Sanger sequencing assay for *BRAF* exon 15 (Spittle *et al*., 2007) and comparable sensitivity to Sanger sequencing assays for *NRAS* codons 12, 13 and 61 (Sivertsson *et al*., 2002). Also, matched frozen and FFPE melanoma DNA samples produced comparable results in a *BRAF* pyrosequencing assay, showing that FFPE DNA is suitable for use in pyrosequencing (Spittle *et al*., 2007). The percentage of the mutated allele compared to the wild-type allele was calculated based on peak height data using the formula: $peak^{mutant}/(peak^{mutant}/peak^{wild\text{-}type})*100$ (Yancovitz *et al*., 2012). Dr Philip Chambers has advised that the sensitivity of the assay is 5%

where a peak can be distinguished from background data (based on his previous experience with the assay). Therefore, the DNA sample was considered mutated when a mutant allele was present in >5% of the cells assessed or >10% if the data was of poor quality. Example pyrograms are illustrated in Figure 2-1.

2.4.2 Sanger sequencing

Traditional Sanger sequencing was performed on a selection of samples to confirm *BRAF* and *NRAS* mutations identified by pyrosequencing. Non-biotinylated pyrosequencing primers as described previously in Table 2-1 were used as sequencing primers and both forward and reverse strands were sequenced.

To remove any unincorporated dNTPs and primers, 2µl of ExoSAP-IT® (USB, Affymetrix, High Wycombe, UK) was added to 5µl of the PCR product, mixed and incubated at 37^{o}C for 15 minutes, followed by 80^{o}C for 15 minutes and finally cooled to 15^{o}C. The sequencing reaction comprised of 0.5µl of 3.2pmol/µl primer, 1µl of purified PCR product, 4.5µl of molecular biology grade water, 3.5µl of ABI 5x Sequencing Buffer (Applied Biosystems, UK) and 0.5µl of 2.5x Ready Reaction Mix v1.1 (from Applied Biosystems BigDye® Terminator Cycle Sequencing kit v1.1). Thermal cycling conditions were 25 cycles of 96^{o}C for 10 seconds, 60^{o}C for 5 seconds and 60^{o}C for 4 minutes, and finally the mixture was cooled to 15^{o}C. Sequencing products were concentrated and purified using ethanol precipitation. Precipitation was achieved by adding 1µl of 3M sodium acetate and 25µl of 95% ethanol to the product and the mixture incubated at room temperature for 15-30 minutes. The samples were centrifuged at 2250*g* for 30 minutes so that the product forms a pellet. The supernatant was removed by inverting the 96-well plate onto absorbent paper and the plate was centrifuged again at 180*g* for 1 minute. 70µl of 70% ethanol was added to wash away residual sodium acetate and the mixture centrifuged at 1650*g* at 4^{o}C for 15 minutes. Supernatant was discarded by inverting the plate onto absorbent paper and centrifugation at 180*g* for 1 minute followed. The plate was heated to 95^{o}C for 1 minute to dry the pellet. The sequencing product was suspended in 20µl of Hi-Di™ formamide (Applied Biosystems), heated to 95^{o}C for 1 minute and then immediately cooled on ice water prior to electrophoresis. Products were separated using the Applied Biosystems PRISM 3130xl Genetic Analyzer with a 36cm length 16-capillary array and POP-7™ polymer. Sequence analysis was performed using Sequencing Analysis 5.2 (Applied Biosystems) and Mutation Surveyor 3.2 (Soft Genetics, State College, PA).

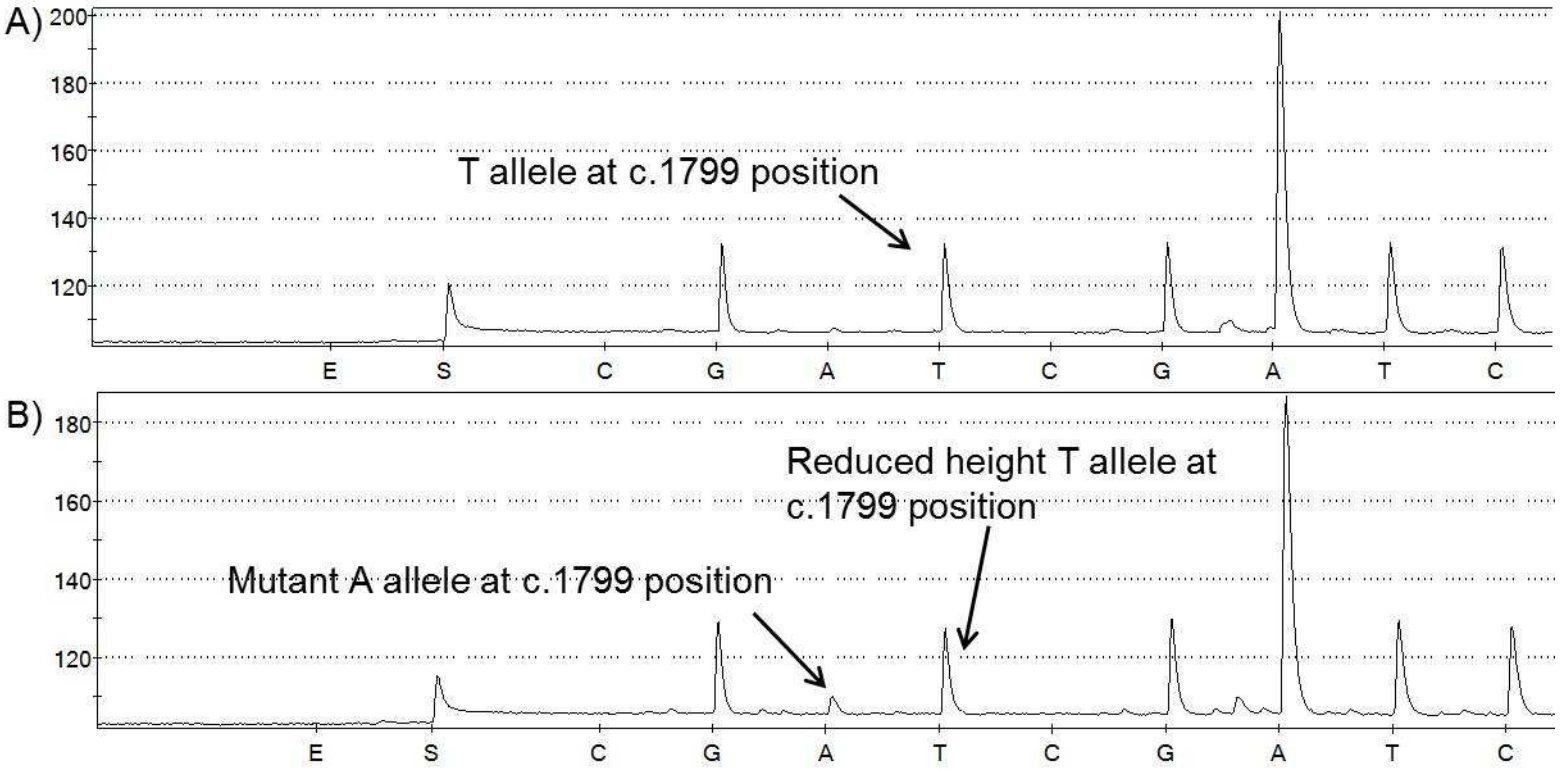


Figure 2-1: Pyrograms for BRAF codon 600.

A) An example of a sample with a wild-type allele, and B) an example of a sample with a V600E (c.1799T>A).

2.5 Cell lines

The SkMel5 (ATCC), MelJuso (ATCC) and SkMel28 (courtesy of Professor Alan Melcher, Leeds CRUK cell line bank) melanoma cell lines were used to for RNA-sequencing as described below in section 2.9.2. The MeWo (courtesy of Professor Alan Melcher, Leeds CRUK cell line bank) cell line was used for a number of methodological assessments but not elsewhere as, following cell line authentication, it appeared to be contaminated with the MM96 cell line (results shown in Appendix A.1.1).

Cell culture experiments described in Chapter 5 were performed in collaboration with Professor Jorg Reichrath, Saarland University, Homburg, Germany. The melanoma cell lines MeWo, SkMel28, MelJuso and SkMel5, which were used in this project, were alternatively sourced and kindly provided by Professor Reichrath. These lines were also subjected to cell line authentication in Leeds (results shown in Appendix A.1.2).

2.5.1 Cell line authentication

Cell line authentication was performed using the PowerPlex® 16 System (Promega, USA). This system detects 16 loci, including Penta E, D18S51, D21S11, TH01, D3S1358, FGA, TPOX, D8S1179, vWA, Penta D, CSF1PO, D16S539, D7S820, D13S317 and D5S818, which are short tandem repeat (STR) loci and the sex marker Amelogenin. STR loci are repetitive sequences of 3-7 base pairs and are distributed throughout the human genome. All STR loci are amplified in a single reaction for each sample using PCR. The alleles of each locus are differentiated by the number of copies of the repeat sequence using fluorescence detection following electrophoretic separation. This work was carried out by Dr Claire Taylor.

2.5.1.1 PCR amplification

The PCR Amplification mix was prepared by combining 2.5µl Gold ST★R 10x buffer, 2.5µl PowerPlex® 16 10x Primer Pair Mix, 0.8µl AmpliTaq Gold® DNA Polymerase (4u), up to 19.2µl of 0.5ng DNA and nuclease-free water up to a final volume of 25µl. A positive and a negative control reaction were prepared using the controls supplied with the kit. The mix was vortexed and pipetted into a 96-well plate. The plate was then loaded onto a GeneAmp® PCR System 9700 thermal cycler, being set up at 95^{o}C for 11 minutes, 96^{o}C for 1 minute and then: ramp 100% to 94^{o}C for 30 seconds, ramp 29% to 60^{o}C for 30 seconds, ramp 23% to 70^{o}C for

45 seconds for 10 cycles, then: ramp 100% to 90°C for 30 seconds, ramp 29% to 60°C for 30 seconds, ramp 23% to 70°C for 45 seconds for 22 cycles and then at 60°C for 30 minutes. The programme was run in the 9600 ramp mode. The reactions were stored at -20°C prior to proceeding to the next step.

2.5.1.2 Detection of amplified products and data analysis

A loading cocktail was prepared by mixing 0.5µl Internal Lane Standard 600 and 9.5µl Hi-Di™ (highly deionized) formamide for each sample. The mix was vortexed and pipetted into each well of a 96-well plate. 1µl amplified sample was then added to each well and sample denaturation performed at 95^{0}C for 3 minutes. The cocktail/sample mix was placed on ice for 3 minutes and then loaded onto the Applied Biosystems 3130xl Genetic Analyser instrument using a 36cm capillary array and POP-7™ polymer. Data analysis was performed using the GeneMarker version 1.95 software (SoftGenetics, USA). A reference profile for each cell line was sought from either the DSMZ (German Collection of Microorganisms and Cell Culture), Cell Line Integrated Molecular Authentication (CLIMA) or Cancer Genome Project (CGP) STR databases and matched against the profile retrieved after analysis.

An example can be seen in Figure 2-2, where the locus D3S1358 is shown in two MeWo cell line samples, one pure and one mixed with MM96 cell line. For this locus MeWo sample has allele 17 and MM96 sample has allele 16. The pure sample clearly carries allele 17 and has lost allele 16 whereas the mixed sample has a small peak of allele 17 and large peak of allele 16 which is the contribution of MM96 contamination. STR profiles of all cell lines can be found in Appendix A.1.

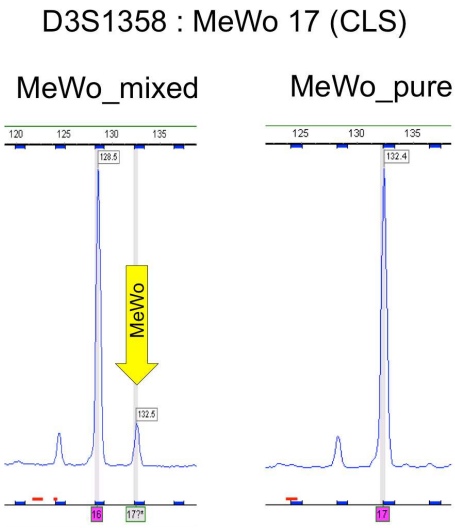


Figure 2-2: D3S1358 locus is shown in two MeWo samples, one pure and one mixed with a different cell line.

2.5.2 Cell line authentication results

The SkMel28, SkMel5 and MelJuso cell lines available in Leeds were correctly authenticated. Unfortunately, cell line authentication in the German cell lines was performed at a later time when the experiments using these cells had already been done. I went to work on the German cell lines in Homburg and DNA samples were subsequently sent to Leeds from the German collaborators to do the authentication test. The MeWo melanoma cell line was correctly authenticated, however SkMel5 appeared to be SkMel28, SkMel28 appeared to be MeWo and the MelJuso cell line did not match any reference sample. A second DNA sample was requested from a passage of frozen cells close to the ones I grew up while I was there. The second authentication showed the same results except that SkMel28 cells were this time correctly authenticated. Looking at the SkMel28 cells under the microscope whilst the cells were being cultured in Germany they had the expected appearance of SkMel28 cells and were very different to the MeWo cells. So, we decided to investigate further whether the cells in my hands were the correct ones. We performed *BRAF* and *NRAS* mutation screening using cDNA derived from the stored RNA from these cells. cDNA from the Leeds cell lines was tested as a positive control. Table 2-2 shows the mutation status identified. The mutation status of the German cell line MeWo, and the Leeds cell lines SkMel28 and MelJuso agrees with the literature (all 3 correctly authenticated). The German SkMel28 which was correctly authenticated the second time appeared to have a p.V600E

BRAF mutation, which agrees with both the Leeds SkMel28 sample and the literature. If it was a MeWo sample as the first authentication results suggested we would expect to observe wild type alleles for both genes tested. So, we demonstrated support for the view that the SkMel28 cells in my hands were the correct cells and not contaminated. The mutation status of the German and Leeds MelJuso sample did not match, which further confirmed that the German MelJuso cells are not MelJuso.

Table 2-2: Mutation screening using cDNA generated from the Leeds versus the German cell lines

Cell line	German cDNA		Leeds cDNA	
	BRAF	***NRAS***	***BRAF***	***NRAS***
MeWo	Wt	Wt	Wt	Wt
SkMel28	p.V600E	Wt	p.V600E	Wt
MelJuso	p.V600E	Wt	Wt	p.Q61L

In summary, we showed evidence that the queried German SkMel28 sample was correct and as a result RNA from this cell line was used in further experiments. The overall conclusions of the cell line authentication for the cell lines from both Leeds and Germany are summarised in Table 2-3. RNA from SkMel5 was not used in further experiments, as it appeared to be a SkMel28 sample in both authentication tests. The German line labelled as MelJuso (henceforth referred in quotes, 'MelJuso') was stained for melanocyte markers to identify whether this cell line is of melanocytic origin, as its profile did not match any reference profile at both authentication tests (see Chapter 5).

Table 2-3: Identity of cell lines as determined by cell line authentication against a reference profile

Cell Line	Leeds	Germany
MeWo	Contaminated with MM96	Authenticated
SkMel28	Authenticated	Authenticated
MelJuso	Authenticated	Unknown profile
SkMel5	Authenticated	SkMel28

Thus, the need to authenticate experimental cell lines attributed to previously extensively characterised lines in the literature has been clearly demonstrated.

2.5.3 Cell culture

SkMel28, SkMel5, MeWo and MelJuso lines were grown in RPMI-1640 medium (Sigma-Aldrich, UK) supplemented with 10% foetal calf serum (FCS) (Sigma-Aldrich, UK) and L-glutamine (2mM) (Sigma-Aldrich, UK) added fresh to the medium for experiments on the effect of adding vitamin D to cultures (described in Chapter 5). Cells were detached from the surface of the flask when they were 80-90% confluent. The medium was removed, cells washed with phosphate buffered saline (PBS) to remove excess serum and 1ml trypsin added. The cells were incubated with trypsin at 37^{o}C until the cells detached from the surface. 5-10ml medium was added; the cells were collected in 50ml conical tubes and pelleted after centrifugation at 400*g* for 4 minutes. The cells were resuspended in 10ml medium and counted using trypan blue exclusion and a haemocytometer prior to 1α,25$(OH)_2D_3$ treatment as described in section 2.5.5. Otherwise, pellets were frozen at -80^{o}C for later DNA/RNA extraction or fixed in formalin as described in section 2.5.4.

2.5.4 Embedding of cells in paraffin wax blocks

Cultured SkMel28, MelJuso and SkMel5 cells in FFPE blocks were prepared and used for nucleic acids extraction and subsequent RNA-sequencing as described in section 2.9.2. Four 150 cm^2 sized flasks of confluent cells for each cell line were pelleted as described in section 2.5.3. Each cell pellet was resuspended in PBS; the cells were split into two 50ml conical tubes and pelleted again. One pellet was frozen at -80^{o}C for later RNA extraction and one pellet was resuspended in 10ml 10% neutral buffered formalin and fixed overnight at room temperature. The next day the cells were centrifuged at 1500rpm for 5 minutes and washed twice with 10ml PBS. At this stage the cells were transferred into a 1.5ml tube and centrifuged at 8000rpm for 5 minutes. After centrifugation the supernatant was discarded, leaving only a small meniscus of PBS. The cell pellets were put at 45^{o}C on a heating block for 2 minutes. A 2% agar solution in deionised water was prepared and allowed to cool to about 55^{o}C. 500µl agar solution was added to the cells and resuspended. The cells/agar mixture were kept at 45^{o}C for 15 minutes and then rapidly pelleted at 8000rpm for 2 minutes before the agar was set. After centrifugation the agar was left to become solid and the tube containing the agar-

embedded cells was filled up with 70% ethanol and subsequently stored at 4°C until processing. The agar-embedded pellets were processed and paraffin-embedded by Dr. Mike Shires (Histopathology team, Leeds University) by routine overnight method.

2.5.5 1α,25(OH)$_2$D$_3$ treatment

MeWo, SkMel28, MelJuso and SkMel5 melanoma cells were treated with 1α,25(OH)$_2$D$_3$ (Sigma-Aldrich, UK) to study the anti-proliferative effects of 1α,25(OH)$_2$D$_3$ that has been reported previously in various cell lineages (Deeb *et al.*, 2007). Vitamin D is a fat soluble hormone and the approach usually described in the literature has been to dissolve in ethanol. It is also sensitive to light. I looked at a variety of approaches in Leeds but had technical difficulties and subsequently decided to travel to Homburg, Germany to carry out the experiments in the laboratory of Professor Reichrath.

The methodology described here is therefore that used and published by the German research group. 5000 cells were seeded in 24-well plates and left to grow for 24 hours as described in section 2.5.3. 1α,25(OH)$_2$D$_3$ was first dissolved in ethanol in the glass bottle provided and mixed well. Then the 1α,25(OH)$_2$D$_3$ solution was further diluted in dark plastic tubes (due to the light sensitive nature of the vitamin D). The final dilution was performed in treatment medium to reach the concentration required. The medium was replenished after 24, 48, 72, and 96 hours and the cells were left to grow for a total of 144 hours. The treatment medium was RPMI-1640 supplemented with 1% bovine serum albumin (BSA), 1% penicillin/streptomycin and 1α,25(OH)$_2$D$_3$ at a final concentration of 10^{-10}M or 10^{-8}M or 10^{-7}M. Control cells were treated with medium only, medium supplemented with 1% BSA and medium supplemented with 1% BSA and 0.01% ethanol. Three independent experiments were performed, and 4 replicates per treatment/control group were included in each experiment.

2.5.6 Proliferation assay using crystal violet dye

Proliferation was measured using crystal violet dye as described previously (Sertznig *et al.*, 2009a). Professor Jorg Reichrath suggested the use of this assay as being the most reproducible compared to superior assays such as WST-1 (work carried out in his lab previously but not reported in the literature). Crystal violet is a deep purple dye which is absorbed by DNA stoichiometrically and therefore absorbance measurement directly correlates with the number of the cells. Crystal

violet addition to the cells and absorbance measurement was done after 24, 48, 72 and 144 hours of 1α,25$(OH)_2D_3$ treatment. A working solution was prepared (1% w/v in 20% ethanol). The medium was removed from the 24-well plates and the cells were washed twice with 500µl cold PBS. The cells were then fixed by addition of 500µl 70% ethanol and incubation at 4^{o}C for 1 hour. The ethanol was removed, 500µl crystal violet working solution added in each well and the plate incubated at room temperature on a shaker for 30 minutes. The plate was then washed using distilled water until the running water was clear. The plate was placed upside down to dry completely. 400µl 70% ethanol was added to each well and the plate incubated at room temperature on a shaker for 30 minutes. The crystal violet dye was solubilized in the ethanol solution and the absorbance at 550nm was measured using a microplate reader (Titertek Multiskan Plus MK II, Labsystems, Finland).

2.5.7 1α,25$(OH)_2D_3$ treatment for subsequent RNA extraction

500,000 melanoma cells were seeded in petri dishes and treated with 1α,25$(OH)_2D_3$ as described in section 2.5.5 for 6, 24 and 48 hours. After these time points, the cells were pelleted and kept frozen for later RNA extraction. Three independent experiments were performed.

2.5.8 Immunocytochemistry (ICC)

Immunocytochemical staining was performed by Alexandra Stark at the Department of Dermatology at Saarland University in Homburg, Germany. The German cell lines SkMel28, MeWo and 'MelJuso' were stained for the vitamin D receptor (VDR) (Table 2-4).

The cell line of unknown origin, 'MelJuso', was also stained for the following melanocytic markers: Melan-A, S100, HMB45, Pan Melanoma cocktail (HMB45, Melan-A, Tyrosinase) to identify if these cells are of melanocytic origin (Table 2-4). The results from this work are presented in Chapter 5.

Immunocytochemistry (ICC), the staining of cultured cells, or immunohistochemistry (IHC), the staining of tissue sections, offers the direct visualization of specific antigens within the examined cells using specific antibodies (Ramos-Vara, 2005). An unlabelled primary antibody specifically binds to the cells followed by a biotinylated IgG secondary antibody which binds to the constant fragment (Fc) of the primary antibody. Then, streptavidin conjugated enzymes bind to biotin on the secondary antibody and generate an insoluble substrate which can be visualised by light microscopy (Ramos-Vara, 2005).

Table 2-4: Primary antibodies used for ICC

Antibodies used are reported with manufacturer details and dilutions used for ICC staining. Anti-VDR was a new antibody tested in the laboratory and a titration experiment was performed showing that 1:500 is the optimal dilution (details in Chapter 5).

Antibody	Staining	Manufacturer	Dilution
Monoclonal mouse anti-**VDR**(D-6)	Cytoplasmic, nuclear	Santa-Cruz Biotechnology	1:500 1:1000 1:2000
Monoclonal mouse anti-**Melan-A** Clone A103	Cytoplasmic	Dako, Germany	1:100
Polyclonal rabbit anti-**S100**	Cytoplasmic	Dako, Germany	1:1000
Monoclonal mouse anti-Melanosome Clone **HMB45**	Cytoplasmic	Dako, Germany	1:100
Monoclonal mouse anti-**Pan Melanoma Cocktail**	Cytoplasmic	Diagnostic Biosystems, Germany	1:50

S-100 is a very sensitive marker for melanoma but its specificity is somewhat limited as it is expressed in other cell types (such as myoepithelial cells and Langerhans cells) (Ohsie *et al*., 2008). Therefore, its use in combination with more specific melanocyte lineage markers such as HMB45, Melan-A and tyrosinase is recommended (Ohsie *et al*., 2008). The latter three markers are less sensitive than S100 but their specificity is 97-100% (Ohsie *et al*., 2008). The use of IHC for the diagnosis of metastasizing melanoma cells in the excised lymph nodes using some or all of the mentioned markers is now routine practice in the clinic (Marsden *et al*., 2010, Balch *et al*., 2009).

ICC was performed using the Dako REAL™ Detection System, Alkaline Phosphatase/RED,Rabbit/Mouse kit on the Autostainer (Dako, Germany). Briefly, $1x10^4$ cells were seeded on a glass slide and left to grow overnight. The next day, the cells were fixed using 4% paraformaldehyde in PBS for 15 minutes and then washed with PBS. Slides to be stained with HMB45 were pre-treated with proteinase K while slides to be stained with the rest of the antibodies were subjected to heat-induced epitope retrieval using a microwave. Diluted primary antibodies were incubated with the cells for 30 minutes, then the biotinylated secondary antibody was added for 15 minutes followed by streptavidin alkaline phosphatase for 15 minutes and the substrate working solution was added for 7

minutes. The slides were then counterstained with haematoxylin, dehydrated and finally mounted with Entellan® (Merck Millipore) mounting medium. Pictured of the stained cells are presented in Chapter 5. The staining was evaluated semi-quantitatively by Dr Jonathan Laye and myself by estimating the staining intensity where 0 represents no staining and 3 strong staining.

2.5.9 Nucleic acids extraction and purification from cell lines

2.5.9.1 Ambion mirVana™ microRNA isolation kit

This kit was used for both total RNA and microRNA isolation from cultured melanoma cell lines used for work in Chapter 5. The supplied organic extraction protocol, as described in brief below, was followed for both total RNA and microRNA isolation.

Cells were trypsinised to detach them and centrifuged at 300*g* for 4 minutes to form a pellet. The pellet was then washed by resuspension in 1ml PBS and pelleted again. The PBS was discarded and the pellet stored at -20^{o}C for up to a month for later RNA extraction. The pellet was removed from the freezer, put on ice and resuspended in 600µl lysis/binding solution. The sample was then vortexed to ensure complete lysis of the cells. Then 60µl microRNA homogenate additive was added to the lysate, mixed well and left on ice for 10 minutes. 600µl Acid:Phenol:Chloroform (Ambion by Life Technologies, USA) was added and mixed well by vortexing. The sample was centrifuged at 10,000*g* for 5 minutes at room temperature to separate the aqueous and organic phases. The aqueous phase (upper) was carefully transferred to a fresh tube without disturbing the lower organic phase which contained the proteins and the volume was measured. The aqueous phase was mixed with ⅓ volume of absolute ethanol; 700µl sample was transferred to a filter cartridge applied to a collection tube and the cartridge was centrifuged at 10,000*g* for 15 seconds. The flow-through was transferred to a new tube as it contained the small RNAs. This step was repeated until all the lysate-ethanol mixture was passed through the filter and the final volume of the flow-through was measured. The total RNA depleted of microRNAs was bound to the filter and it was recovered as described below in section 2.5.9.1.2.

2.5.9.1.1 miRNA isolation

A ⅔ volume of absolute ethanol was added to the flow-through retained in the previous step and mixed well. The mixed sample was transferred to a new filter

cartridge applied to a collection tube and the cartridge was centrifuged at 10,000*g* for 15 seconds. The flow-through was discarded and the step was repeated until the entire sample had passed through the filter. A wash step was then performed by addition of 700µl Wash 1 buffer and centrifugation at 10,000*g* for 15 seconds. Two more wash steps were performed by adding 500µl Wash 2 buffer to the cartridge. After discarding the flow-through from the last wash step the cartridge was centrifuged at 10,000*g* for 1 minute to remove residual fluid. The cartridge was transferred to a new collection tube, 100µl nuclease-free water was applied to the cartridge and it was left stand at room temperature for 1 minute. After centrifugation at 10,000*g* for 1 minute the eluate contained the microRNAs. This sample was stored at -80^{o}C.

2.5.9.1.2 Total RNA isolation

700µl microRNA Wash buffer 1 was applied to the cartridge from the last step described in section 2.5.9.1. The cartridge was centrifuged at 10,000*g* for 15 seconds and the flow-through was discarded. Two more wash steps were performed after addition of 500µl Wash buffer 2/3. After the last wash step the cartridge was centrifuged at 10,000*g* for 1 minute to remove residual fluid and then it was transferred to a new collection tube. Then, 100µl nuclease-free water was added and the cartridge was left at room temperature for 1 minute. Total RNA was eluted after centrifugation at 10,000*g* for 1 minute. The RNA was stored at -80^{o}C

2.5.9.2 Qiagen RNeasy® mini kit

Total RNA was extracted using this kit according to the supplied protocol and was further used for RNA-sequencing (section 2.9.2).

Briefly, cells were harvested as a pellet and stored at -80^{o}C prior to processing. 600µl RLT Plus buffer was added to each pellet and mixed well by vortexing for 30 seconds. Then 600µl 70% ethanol was mixed well with the sample by vortexing and up to 700µl of it was transferred to an RNeasy spin column placed in a 2ml collection tube. The column was centrifuged for 15 seconds at 8,000*g* and the flow-through was discarded. This step was repeated until the entire sample had passed through the column. The column was washed by addition of 350µl RW1 buffer to the column, and centrifugation, after which the flow-through was discarded. 80µl DNase I incubation mix (70µl RDD buffer and 10µl DNase I; Rnase-free DNase set, Qiagen, Crawley UK) was added to the column and incubated for 15 minutes at room temperature. Another wash step was performed by addition of 350µl RW1

buffer and then two more wash steps by addition of 500µl RPE buffer to the column. After the last wash step, the column was transferred to a new 2 ml collection tube, and centrifuged at full speed for 1 minute to dry the membrane. The column was then placed in a new 1.5ml microcentrifuge tube, 100µl RNase water added and the column left to stand at room temperature for 1 minute. The column was centrifuged for 1 minute at 8,000*g* and the eluted RNA was stored at -80^{o}C.

2.5.9.3 Qiagen AllPrep® DNA/RNA FFPE kit

RNA and DNA was extracted from cells embedded in FFPE blocks using the AllPrep® DNA/RNA FFPE kit and RNA was subsequently used for RNA-sequencing (section 2.9.2). Three cores were taken from each block and deparaffinised as described in section 2.3.6.1. The Qiagen AllPrep® kit was used as described in section 2.3.6.3 with the following modification: the RNA was eluted in 100µl RNase-free water and the DNA in 100µl elution buffer.

2.5.9.4 Ambion Turbo™ DNase-free kit

DNase treatment was essential to remove contaminating DNA from RNA samples used on the Affymetrix gene expression array. RNA extracted using the Ambion miRvana kit was DNase treated consequently using the Turbo™ DNase-free kit (Ambion by Life Technologies, UK).

A DNase treatment was performed on the eluted RNA by addition of 0.1 volume 10x Turbo™ DNase buffer and 0.02x volume Turbo™ DNase. The sample was mixed gently and incubated at 37^{o}C for 30 minutes. Then 0.1x volume of resuspended DNase Inactivation reagent was added and the sample incubated for 5 minutes at room temperature, mixing occasionally. The tube was then centrifuged at 10,000*g* for 15 minutes and the RNA transferred to a new tube without disturbing the pellet formed by the DNase Inactivation Reagent.

2.5.9.5 Qiagen RNeasy® MinElute® Cleanup kit

DNase treated RNA with the Turbo™ DNase kit was immediately cleaned up using the RNeasy® MinElute® CleanUp kit (Qiagen, Crawley UK) to ensure no contaminants were left in the RNA sample. The RNA sample was adjusted to a volume of 100µl with RNase-free water, 350µl RLT buffer added and the sample mixed well. Then 250µl absolute ethanol was added, mixed well and up to 700µl of sample was transferred to an RNeasy® MinElute® column placed in a 2ml

collection tube. The column was centrifuged at 8,000*g* for 15 seconds, the flow-through was discarded and the step was repeated until the entire sample had passed through the column. The column was then transferred to a fresh 2ml collection tube, and two wash steps were performed by addition of 500µl RPE buffer followed by 500µl 80% ethanol. After the last centrifugation, the column was placed in a new 2ml collection tube and centrifuged at full speed for 5 minutes to dry the membrane. The column was then placed in a new 1.5ml microcentrifuge tube. 15µl RNase-free water was added to the centre of the column, left at room temperature for 1 minute and centrifuged at full speed for 1 minute. The eluted RNA was stored at 80^{o}C prior to processing.

2.5.10 cDNA synthesis of RNA extracted from cell lines

Production of cDNA from RNA extracted from cell lines was carried out using the High Capacity cDNA Reverse Transcription kit (Invitrogen™ by Life Technologies, USA) as described in section 2.3.7.1.

2.6 Nucleic acid quantification and fragment size assessment

2.6.1 DNA/RNA quantification using the ND-8000

The concentrations of DNA and RNA extracted from FFPE or frozen samples was measured using the ND-8000 (Nanodrop, Thermo Scientific, USA), a spectrophotometer measuring absorbance in a spectrum of 220nm-750nm. 1.5µl nucleic acid was used and absorbance was measured at the appropriate wavelength. Single-stranded DNA, double-stranded DNA and RNA absorb at 260nm. Sample quality was also determined by calculation of the A_{260}/A_{280} and A_{260}/A_{230} ratios, indicative of sample contamination as organic compounds absorb UV light at 230nm and protein at 280nm. DNA is considered pure when the A_{260}/A_{280} ratio is 1.7-1.9 and RNA when the A_{260}/A_{280} ratio is 1.9-2.1.

2.6.2 DNA quantification using Picogreen

The concentration of DNA used for next-generation sequencing was measured using picogreen as an accurate measurement of double-stranded DNA was required. Picogreen is a dye which binds and intercalates with DNA meaning that

only double-stranded DNA is measured. This is in contrast with spectrophotometers which measure both single and double-stranded DNA. The Quant-iT™ broad range ds-DNA assay kit (Invitrogen by Life Technologies, USA) was used. 198µl picogreen mix (197µl reagent buffer and 1µl picogreen reagent) was mixed with 2µl of standard DNAs of known-concentration (kit-supplied) or 2µl DNA to be quantified and loaded onto a black flat-bottom 96-well plate. The plate was loaded onto the FLUOstar Galaxy spectrofluorometer (BMG Labtech, Germany) and intensity was measured using an excitation filter at 480nm and an emission filter at 520nm. The standard curve using the 8 DNAs of known concentration and the concentration of the unknown samples were calculated by the FLUOStar Galaxy software.

2.6.3 DNA/RNA fragment size assessment

DNA/RNA fragment size was determined during preparation of samples for next-generation sequencing described in section 2.9.1 and 2.9.2. RNA fragment size was checked to determine the quality of RNA samples before running them on the Affymetrix microarray described in section 2.7 and before proceeding to library preparation for RNA-sequencing described in section 2.9.2. Both Agilent 2100 Bioanalyser and 2200 Tapestation automated electrophoretic systems were used. The assay on 2100 Bioanalyser is in a chip format and uses micro-fluidics technology to enable electrophoretic nucleic acid fragment separation (Mueller *et al.*, 2000). The assay on the most recent Agilent 2200 Tapestation instrument is in a tape format and contains acrylamide gel and integrated electrodes to enable nucleic acid separation. The fragment sizes were evaluated by looking at the electropherograms.

2.6.3.1 Agilent 2100 Bioanalyser

2.6.3.1.1 DNA 1000/High Sensitivity kit for DNA analysis

The Agilent DNA 1000 kit was used for the detection of DNA fragments after shearing or the detection of DNA library size and the manufacturer's protocol was followed. The gel-dye matrix was prepared by adding 25µl DNA blue dye concentrate to a DNA gel matrix vial (both equilibrated at room temperature), mixed well by vortexing. This was transferred to the spin column provided and centrifuged at 2240*g* for 15 minutes. 9µl DNA gel-dye mix was added to the appropriate well of the chip. The chip was placed in the priming station, the syringe was plunged down and the chip left in place for 1 minute. Then the chip was taken out of the priming

station and 9µl DNA gel-dye mix, 5µl DNA marker, 1µl DNA ladder and 1µl sample were added to the appropriate wells. The chip was vortexed for 1 minute and finally loaded onto the instrument. Most FFPE libraries were of low concentration and for these the High Sensitivity version of the kit was used, especially when comparisons were done between samples prepared using different methods. The manufacturer's protocol was followed (similar to the DNA 1000 kit). The data from both chips were analysed using the 2100 Expert software.

2.6.3.2 Agilent 2200 Tapestation

2.6.3.2.1 Standard D1K/High Sensitivity HD1K Tape for DNA analysis

The Agilent 2200 Tapestation was also used for the detection of DNA fragments after shearing or the detection of DNA library size when it replaced the Agilent 2100 Bioanalyser in the laboratory. The Standard D1K kit or the High Sensitivity HD1K kit was used and the manufacturer's protocol was followed. 16 samples can be run on each tape, usually comprising a ladder and 15 DNA samples. The samples are prepared in 8-well strip tubes and then specialist loading tips transfer the samples from the tubes to the tape in the instrument. 3µl D1K ladder was placed into the first well of the 8-well strip, with 3µl sample buffer and 1µl DNA or 2µl sample buffer and 1µl DNA being prepared for the D1K kit or the HD1K kit respectively. The tubes were quickly vortexed and centrifuged, and loaded onto sample cradle of the instrument, together with the loading tips and appropriate tape. The data were analysed using the supplied 2200 Tapestation Controller software.

2.6.3.2.2 R6K kit for RNA analysis

Total RNA fragment size assessment was done using the R6K kit and the manufacturer's protocol was followed. 4µl R6K sample buffer was mixed with 1µl RNA. Sample denaturation was achieved by heating the samples at 72^{o}C for 3 minutes and then placing them on ice for 2 minutes. The samples, loading tips and R6K tape were loaded onto the instrument and the data analysis was performed using the 2200 Tapestation Controller software. A software ladder was assigned to each run and the sizes were assessed. The RNA Integrity Number (RIN) was calculated by the software and used as a quality assessment measure. RIN values start from 0 indicating totally degraded RNA to 10 indicating totally intact RNA. The algorithm that calculates the RIN takes into account several features of the electrophoretic curve as described previously (Schroeder *et al.*, 2006).

2.7 Affymetrix GeneChip® Human Genome U133 plus 2.0 Array

The Affymetrix GeneChip® Human Genome U133 plus 2.0 platform array was used for the gene expression work described in Chapter 5. 500ng total RNA was sent to the Affymetrix service provider at the Patterson Institute, University of Manchester, UK. The Human Genome U133 plus 2.0 array is a high-density probe-based array (Lockhart *et al*., 1996) currently comprised of 54,120 probe sets, translating into 47,401 gene transcripts. A probe is a 25-mer oligonucleotide which targets a specific gene transcript and millions of copies of a specific probe are immobilised on the solid surface of the array in a specific region called a probe cell. Each transcript is represented by multiple probe sets and each probe set is comprised of a perfect match (PM) and a mismatch probe (MM) (Lockhart *et al*., 1996, Lipshutz *et al*., 1999). The mismatch probe has the same sequence as the perfect match probe except for a non-complementary nucleotide in its 13^{th} position. Theoretically, the mismatch probes can be used to remove non-specific hybridisation (Lipshutz *et al*., 1999). Sample preparation and array scanning were performed according to Affymetrix guidelines (Affymetrix, 2009) at the Patterson Institute. Briefly 40ng high quality RNA (RIN>7) was converted to double-stranded cDNA. Only one RNA sample had a RIN<7 and was therefore not processed further. The double-stranded cDNA was purified using a column-based step, amplified and then repurified. The clean, double-stranded cDNA was converted to biotin-labelled cRNA using *in vitro* transcription, purified and fragmented to produce fragments of 35-200 bases in length. A hybridisation cocktail was prepared containing the clean biotin-labelled cRNA fragments and the mixture was transferred to the array where the cRNA fragments were hybridised to the complementary oligonucleotides present on the array for 16 hours. After hybridisation, unbound cRNA was removed and a number of washing and staining steps were performed resulting in the fluorescent agent streptavidin-phycoeryhtrin being bound to the biotin-labelled cRNA (Figure 2-3). The array was then placed in the GeneArray® laser scanner where the intensity of the fluorescent agent was recorded and an image of the entire chip was taken. Using a control probe a grid is aligned on the image and a raw intensity probe value for each probe cell is calculated. The raw intensity values are stored in CEL files which is the format that the data were received and used for subsequent analysis.

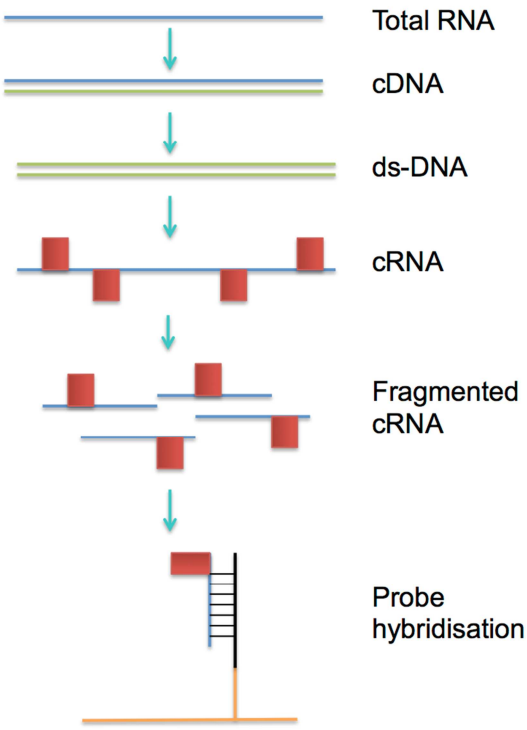


Figure 2-3: Affymetrix GeneChip array.

RNA is converted to double-stranded cDNA, which is then converted to biotin-labelled cRNA using *in vitro* transcription and fragmented to produce fragments of 35-200 bases in length. The cRNA fragments are hybridised to the complementary oligonucleotides present on the Affymetrix GeneChip® Human Genome U133 plus 2.0 platform array.

2.7.1 Data pre-processing and quality control

Data pre-processing is always performed before statistical analysis of array-based gene expression data to subtract technical variation from biological variation (Bolstad *et al.*, 2003). Technical variation usually arises from differences in sample preparation, processing or scanning of the arrays. A number of different methods have been compared for their performance so far (Irizarry *et al.*, 2003a, Giorgi *et al.*, 2010). Here, the Robust Multi-Chip Analysis (RMA) algorithm (Irizarry *et al.*, 2003b) was used to pre-process the data described in Chapter 5 using BioConductor software (Gentleman *et al.*, 2004) in R programming environment (R Development Core Team, 2012). RMA uses only the PM probe values and totally ignores the MM probe values while the originally proposed algorithm for Affymetrix array data, MAS5, uses both PM and MM probe values (Hubbell *et al.*, 2002). It has

been shown that subtracting the MM probe values from PM values mathematically does not always correct for non-specific hybridisation (Irizarry *et al*., 2003a, Irizarry *et al*., 2003b, Naef *et al*., 2002). Furthermore, MAS5 does not take into account across-array normalisation while RMA does (Irizarry *et al*., 2003b). Briefly, RMA pre-processing includes three steps: background adjustment to remove non-specific intensities, quantile normalisation across arrays to make them comparable, and summarisation using a log scale linear additive model to calculate a single expression measure for each probe set. Density plots of log-intensity distribution of each array before and after data pre-processing were used to identify arrays with abnormal distributions. The density distributions of raw probe-set log-intensities are not expected to be identical but still well correlated while the distributions of normalized probe-set log-intensities are expected to be more identical.

After data pre-processing quality control assessment was performed using the affyPLM package (Bolstad *et al*., 2005) available from BioConductor software in R which uses probe-level models. Chip pseudo images of the residuals of the models were produced to help us identify any artefacts on the arrays. Box plots of the Relative Log Expression (RLE) values for each probe set were also produced by comparing the expression value of each array against the median expression value across the arrays for the given probe set. The box plots should be centred to zero assuming that most of the expression values across arrays remain the same. Normalised Unscaled Standard Errors (NUSE) for each probe set on the array were also plotted. Standard error estimates for each probe set on each array were calculated and standardised across arrays so that the median standard error for that probe set is 1. Elevated standard errors relative to the other arrays are usually of lower quality. Finally, the average intensity of each probe across all probe-sets, ordered from the 5‘ to the 3’ end was calculated and plotted as a curve (RNA degradation plot) for each sample. Since degradation starts from the 5’ end of the RNA molecule, we would expect probe intensities to be lower at that end of a probe set compared to the 3’ end. An RNA sample that is too degraded will have a very high slope from 5' to 3'. The slope of the curves was thus used as a quantitative indicator of the RNA degradation. Two samples which did not perform as expected in most of the quality control plots were excluded from further analysis. The density distributions of log-intensities before and after pre-processing are included in Appendix A.2. Data pre-processing and quality control was performed by Dr Alastair Droop who is a bioinformatician in the Leeds Cancer Research UK Centre.

2.8 Illumina whole-genome cDNA-mediated annealing, selection, extension and ligation (WG-DASL HT) assay

The Illumina WG-DASL HT assay, which uses the HumanHT-12 v4 Expression BeadChip, was used for the gene expression work using FFPE samples described in Chapter 4. Two hundred nanograms total RNA (whenever possible) was sent to the Illumina service provider, ServiceXS (Leiden, Netherlands). The 12-sample HumanHT-12 v4 Expression BeadChip platform is composed of individual 40,000 probe-based bead arrays on a microscope slide surface. Three-micron beads assembled into the micro wells of the BeadChip substrate have probes attached to them. Each bead has thousands of copies of one type of probe and is present in multiple copies on the array. Each array has a 12-sample format enabling multiple samples assayed using a single array.

Sample preparation and array scanning were performed according to the Illumina guidelines (Illumina, 2012) by ServiceXS (Leiden, Netherlands). The WG-DASL HT assay can be used for partially degraded RNAs as it uses random priming in cDNA synthesis and a target sequence of only 50 nucleotides is required for probe groups to anneal (described below). RNA quality was measured using qRT-PCR of the housekeeping gene *RPL13a* as described in the Illumina manual (Illumina, 2012), by ServiceXS. Mean cycle threshold (Ct) values were determined for all samples. Previous DASL data within the group have demonstrated that this quality control step does not correlate with performance on the array and therefore all samples were used in the assay irrespective of this quality control step (Conway *et al.*, 2009, Jewell *et al.*, 2010).

The assay begins with conversion of total RNA to cDNA using biotinylated oligo(dT) and random nonamer primers. The biotinylated cDNA is then annealed to the DASL Assay Pool (DAP) probe groups. The DASL Assay Pool probe groups consist of 29,285 unique probes. The assay probe group consists of two 50-mer oligonucleotides; an upstream which contains a transcript-specific sequence and a universal PCR primer 1 at the 5’ end and a downstream which contains a transcript-specific sequence and a universal PCR primer 2 at the 3’ end. The oligos are hybridised to the targeted cDNA sites and non-hybridised oligos are washed away using streptavidin-conjugated paramagnetic particles. A polymerase is added and the upstream oligo extends and ligates to its corresponding downstream oligo to create a PCR template which is amplified with universal PCR primers 1 and 2. The PCR primer on the strand that is complementary to the array is fluorescently labelled. After amplification the labelled single-stranded product is hybridised to the

HumanHT-12 v4 Expression BeadChip (Figure 2-4). After hybridization, HumanHT-12 v4 Beadchips are scanned using the Illumina HiScan, iScan System or BeadArray Reader which measures fluorescence intensity at each bead location. The intensity of the signal corresponds to the quantity of the respective transcript in the original sample. The intensity values are stored in data files (*.bgx) and are analysed using Illumina's GenomeStudio® software.

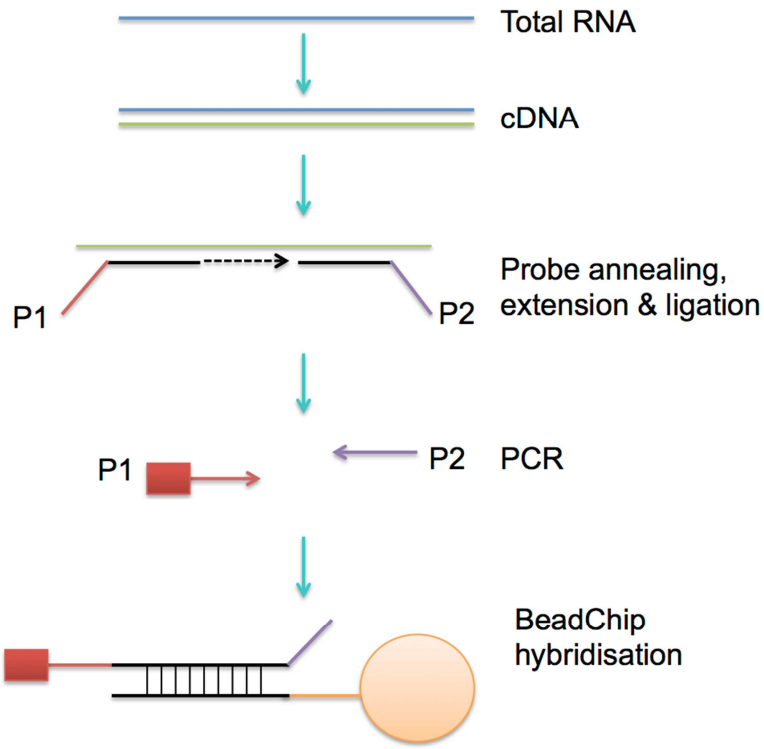


Figure 2-4: WG-DASL HT assay.

RNA is converted to cDNA using biotinylated primers. The biotinylated cDNA is then annealed to the DASL Assay Pool (DAP) probe groups. Correctly annealed oligos are extended, ligated and amplified using PCR. The PCR primer on the strand that is complementary to the array is fluorescently labelled. After amplification the labelled single-stranded product is hybridised to the HumanHT-12 v4 Expression BeadChip.

2.8.1 Control and replicate samples

To ensure the reproducibility of the assay, technical replicates (RNA samples from the same extraction) and biological replicates (RNA samples derived from different FFPE blocks of the same tumour) were assessed. To further monitor the performance of the assay MelJuso cell line RNA samples and Stratagene Universal Human Reference RNAs (Agilent Technologies, Edinburgh, UK) were also included on the array.

2.8.2 Data pre-processing and quality control

As discussed in section 2.7.1 data pre-processing is an important step before statistical analysis of array-based gene expression data (Bolstad *et al*., 2003). Data were pre-processed using Lumi package (Du *et al*., 2008) downloaded from BioConductor (Gentleman *et al*., 2004) in R (R Development Core Team, 2012).

The number of genes detected in each sample represents the genes for which target sequence signal is distinguishable from negative controls on the array at threshold p value<0.01. Previously published data have reported the number of genes detected as a measure of the array performance (Ravo *et al*., 2008). In this experiment the average number of genes detected was 13572 (range 6848-16194, Figure 2-5).

Data were background corrected to remove non-specific intensities and normalised using Robust Spline Normalisation (RSN), a method which is unique in the Lumi package (Du *et al*., 2008). Briefly, RSN includes features of quantile normalization (preservation of the rank order of genes) and loess method (continuous intensity data transformation) (Du *et al*., 2008).

Quality control (QC) plots including density plots of log-intensity distribution and boxplots of each sample before and after data pre-processing were used to identify samples with abnormal performance. These QC plots highlighted a few samples that did not perform well. Therefore, we explored the possibility that samples with low number of genes detected have not performed well on the array. The number of genes detected in each sample was found to be inversely correlated with both RNA concentration (Spearman's rho=-0.20, p=0.002) and Ct value of the housekeeping gene *RFLP13a* (Spearman's rho=-0.20, p=0.002) but not with melanin score or age of the block (data not shown). This finding shows evidence that a low number of genes detected might be due to a combination of low input material and low RNA quality. Consequently, data normalisation was rerun after exclusion of samples with

<9000 genes detected (2.5% of the samples). Log-intensity plots after normalisation showed three further outlier samples and we excluded them as well. Two of these samples had the number of genes detected in the lower bracket (respectively 9674 and 10021), which further supports the view that the number of genes detected is a good diagnostic tool. Based on the above observation the threshold applied for sample exclusion was <10021 genes detected. The number of genes detected for the third sample (14362) could not explain its abnormal behaviour after pre-processing. However, this sample was also excluded from the analysis as it was an outlier (possibly due to contamination or signal saturation). The cell line samples expressed more genes than tumour tissues (see Figure 2-5) and their log-intensity distributions were outlying before and after the pre-processing. For this reason we excluded them from the dataset prior to the final pre-processing.

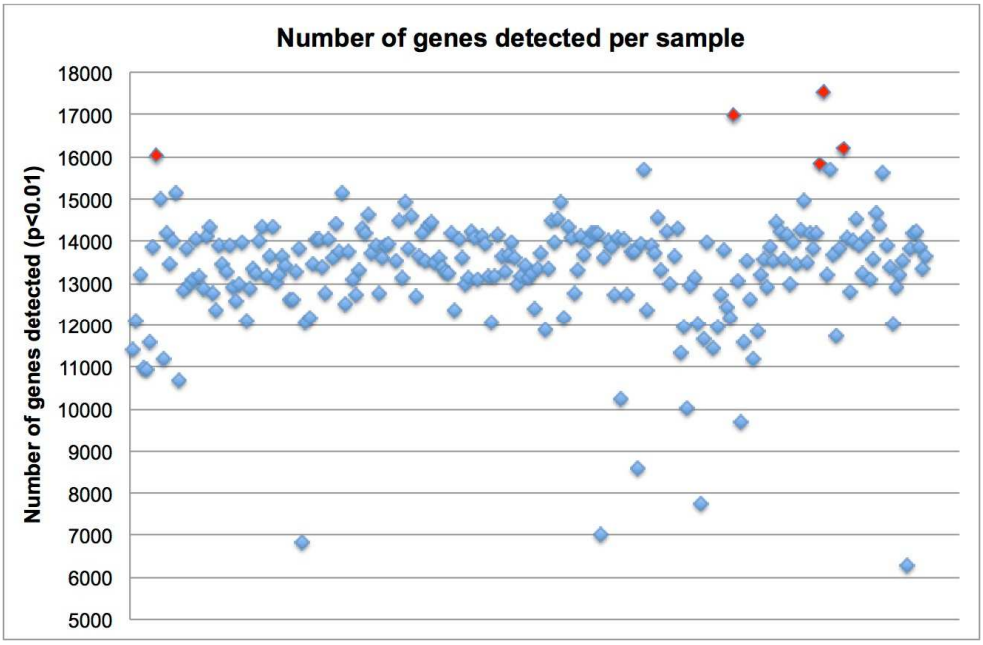


Figure 2-5: Number of genes detected (p<0.01) for each sample on the WG-DASL HT assay.

The red dots represent the number of genes detected for the cell line samples. All other samples were FFPE-derived.

In summary, data pre-processing was finally performed after exclusion of 5 samples with <10021 genes detected, 1 outlier sample and all cell line RNA samples. QC plots after sample exclusion before and after data pre-processing can be found in the Appendix A.2.3. The correlation between replicate samples after pre-processing is shown in Table 2-5. The correlation coefficients between normalised probe signals of replicate cell line RNAs are presented separately as they were before their exclusion to demonstrate the performance of the assay. Correlation coefficients across technical replicates are high, with slightly higher correlation across cell line RNA technical replicates, as expected (Table 2-5). The mean correlation coefficient for biological replicates was smaller which could be attributed to tumour heterogeneity (Table 2-5).

In summary, the data were of sufficiently good quality, and subsequently the replicate signals were averaged before log transformation and statistical analyses described in Chapter 5.

Table 2-5: Correlation coefficients between RSN transformed probe signal intensities of replicate samples.

Correlation coefficients for technical cell line replicates are presented as they were before cell line samples were finally excluded from the data.

Replicates (number of pairs)	Number of replicate pairs with failed samples	Mean of correlation coefficients (Range)
Technical FFPE replicates (22)	2	0.96 (0.81-0.99)
Biological FFPE replicates (5)	0	0.91 (0.84-0.93)
Technical cell line replicates (4)	0	0.98 (0.98-0.99)

2.9 Next-generation sequencing

A number of different next-generation sequencing platforms have been developed so far (reviewed by (Zhang *et al.*, 2011, Ross and Cronin, 2011, Mardis, 2008)). These are: a) the Roche GS-FLX/454 Genome Sequencer; b) the Illumina Genome Analyzer HiSeq2000; c) the ABI SOLiD system; d) the Danacher/Dover/Azco Polonator G.007; and e) the Helicos HeliScope. All platforms use slightly different technology to perform massively parallel sequencing differing in accuracy, error-rate and read-length (Zhang *et al.*, 2011). Here in Leeds the Illumina Genome Analyzer platforms have been used. Initially there was a Genome Analyzer IIx installed, however more recently a HiSeq2000 instrument replaced the Genome Analyzer IIx.

The Illumina system uses a sequencing-by-synthesis technology. The template requires amplification, through solid-phase amplification (Metzker, 2010), before sequencing, as this system cannot detect single fluorescent events. Double-stranded DNA, the input material, is fragmented mechanically or enzymatically and adapter sequences are ligated to the sample. A PCR step follows to amplify the ligated template which is then called a library and further attached to a glass slide, the flow cell. Oligos complementary to the adapters are attached to the surface of the flow cell. A flow cell consists of 8 lanes and thus 8 samples can be run simultaneously (7 test samples and 1 control). Each cluster is generated through bridge amplification by the addition of non-fluorescent nucleotides and a DNA polymerase (Metzker, 2010). In the end each fragment consists of millions of clones. The clusters provide free ends and a universal sequencing primer is hybridised and sequencing initiates. All four fluorescently-labelled nucleotides with 3’ blocked reversible terminators are added simultaneously but only the complementary one is incorporated at each position. When the first nucleotide is added to each sequence, the unincorporated nucleotides are washed away and imaging is performed. Then, the terminating group and the fluorescent dye are removed and the second incorporation takes place. This is called cyclic reversible termination (Metzker, 2010). The steps from DNA fragmentation to imaging are shown schematically in Figure 2-6.

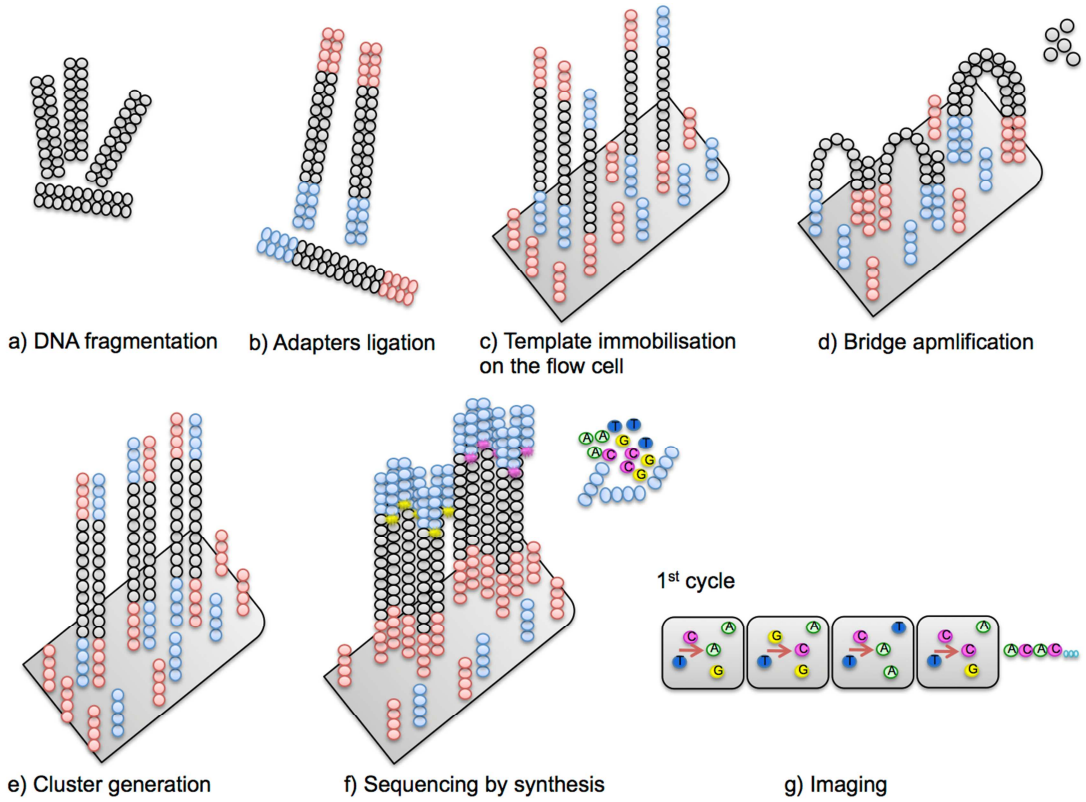


Figure 2-6: Illumina sequencing technology.

DNA is fragmented and adapters are ligated to each fragment, followed by PCR amplification. The amplified product is immobilised on the flow cell and each sequence is further amplified to form clusters comprised of millions of copies of that sequence. Sequencing is performed using sequencing-by-synthesis technology and imaging is performed. (More details are provided in the text) Adapted from (Metzker, 2010).

2.9.1 Whole-genome library preparation protocols using DNA for Illumina next-generation sequencing on the HiSeq2000 instruments

In the course of my thesis I attempted to generate next generation sequencing data from tiny formalin-fixed tumours, some of which were very old (and likely therefore degraded) and heavily contaminated with melanin. Library preparation was therefore potentially a serious issue both in terms of efficiency and time taken and therefore five different protocols were evaluated for the preparation of DNA libraries from FFPE derived DNA (Chapter 3). Protocols number 1, 2 and 5 as listed below

were performed by myself while protocols 3 and 4 by a service provider, Source Biosciences, UK.

All are based on the same principle except for the NEXTERA™ protocol which is described in detail in section 2.9.1.4. DNA is first fragmented into small fragments mechanically resulting in the formation of blunt ends. An end-repair reaction corrects the blunt ends and adenine addition follows to prepare DNA for adapter ligation. During ligation, adapters containing a 6-nucleotide barcode at the 5' end or indexed adapters are ligated to the DNA and a PCR step follows to amplify the ligated sample. Only one protocol (see section 2.9.1.5) used a different method to tag the samples where non-barcoded adapters were ligated to the DNA and indexed primers during PCR amplification were used to tag the sample. The indexed adapters/primers contain a 6-nucleotide barcode in the middle of the adapter/primer molecule, thus a different sequencing primer is used to read the barcode in an independent sequencing run. The barcodes are useful to identify samples when multiplexing more than one sample. The amplified sample is loaded onto the flow cell and cluster generation and sequencing are performed as described in section 2.9.

The concentration of the input DNA and the final library was measured using Picogreen (section 2.6.2). The size of the libraries was evaluated on either a standard or a high sensitivity chip on the Agilent Bioanalyser or the Tapestation System (section 2.6.3).

2.9.1.1 Library preparation protocol number 1 using 5'-end barcodes (manual protocol)

DNA in a final volume of 250µl (TE added) was sheared on the Covaris S2 sonicator (KBiosciences, UK) to produce an average fragment size of 200bp.

After shearing, DNA was cleaned up using the MinElute purification kit (Qiagen, Crawley UK). Briefly, 5x volume of the PB binding buffer was added to the sheared DNA, the solution was processed through a MinElute column, then a wash step of 750µl of PE wash buffer followed and the elution was done in 10µl of the elution buffer (EB) provided within the kit. 1µl of sheared sample was checked on the Bioanalyser or Tapestation system (Agilent Technologies, USA) using a standard or a high sensitivity chip.

The blunt DNA ends formed after shearing were repaired using the End-It DNA End Repair kit (Epicentre, USA). The reaction was prepared as shown in Table 2-6 and

the mix was incubated at room temperature for 45 minutes. The reaction was then purified using the QiaQuick purification kit (Qiagen, Crawley UK) using the same protocol when the MinElute purification kit was used but a QiaQuick column used instead of a MinElute. The elution was done in 34.45µl EB or 30µl EB if BSA was used.

Table 2-6: End-repair reaction

	Volume without BSA (µl)	Volume with BSA (µl)
DNA	9	9
End-it Buffer	5	5
dNTP mix	5	5
ATP	5	5
End-it Enzyme mix	1	1
BSA	-	5
dH_2O	25	20

The A-Addition reaction followed where adenine is added at the repaired DNA ends to prepare the molecule for the adapter ligation. The reaction was prepared as in Table 2-7 using 9 U/µL DNA polymerase I Large (Klenow) fragment, exonuclease minus (Promega, USA) and dATP (Promega, USA).

Table 2-7: A-Addition reaction

	Volume without BSA (µl)	Volume with BSA (µl)
DNA	34.45	30
DNA Pol (Klenow) Buffer	5	5
1mM dATP	10	10
Klenow	0.55	0.55
BSA	-	4.45

The ligation mix prepared as described in Table 2-8 and the mix was incubated at room temperature for 15 minutes followed by 20 minutes at 65^{o}C for to inactivate the ligase. The adapter sequences are listed in Appendix A.3.

Table 2-8: Ligation reaction

	Volume without BSA (µl)	Volume with BSA (µl)
DNA	10	9
Liga-fast reaction buffer	15	15
Adapter	0.5	0.5
T4 DNA Ligase	3	3
dH_2O	2	-
BSA	-	3

After ligation, size selection was performed using the AgenCourt Ampure XP Beads (Beckman Coulter, USA). The beads were added (0.9x of the reaction volume) and the sample mixed thoroughly by pipetting. The mixed sample was left at room temperature for 3-5 minutes and then placed on a magnetic separator rack (Invitrogen, UK) for 5-10 minutes and the beads were separated from the solution. The supernatant was aspirated and discarded. Two wash steps with 200µl 70% ethanol followed and after the last wash step the beads were air-dried for 15-20 minutes and the DNA was finally eluted in 30µl EB. All the steps were performed on the magnetic rack.

A PCR enrichment step was performed to enrich the ligated sample. The reaction was prepared as described in Table 2-9 and placed on a thermocycler at 98^{o}C for 30 seconds, 15 or 18 cycles of 98^{o}C for 10 seconds, 65^{o}C for 30 seconds and 72^{o}C for 30 seconds, and finally 72^{o}C for 5 minutes. The forward and reverse primers used were 5'-CAA GCA GAA GAC GGC ATA CGA GAT CGG TCT CGG CAT TCC TGC TGA ACC GCT CTT CCG ATC T-3’ and 5'-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T-3’ respectively.

Table 2-9: Enrichment reaction

	Volume without BSA (µl)	Volume with BSA (µl)
DNA	20	20
Phusion HF Master Mix	25	25
PE-PTO-F primer(25uM)	1	1
PE-PTO-R primer(25uM)	1	1
BSA	-	5
dH_2O	13	8

The reaction was finally purified using the AgenCourt Ampure XP Beads (1.8x of reaction volume) as described previously.

2.9.1.2 Library preparation protocol number 2 using 5'-end barcodes on the SPRIworks system (Beckman Coulter, USA)

Libraries were prepared using the SPRIworks automated system (Beckman Coulter, USA). The DNA was sheared using the Covaris S2 sonicator and it was placed straight on the SPRIworks instrument. The steps performed on the instrument were: DNA purification using beads; end-repair; purification with beads; A-addition; purification with beads; adapter ligation; and beads size selection. The 5'-end barcoded adapters were used in this system but were diluted first in 1:20. The enrichment and the final clean-up steps were manually performed as described in section 2.9.1.1.

2.9.1.3 Library preparation protocol number 3 on the Agilent Bravo robot using the TruSeq™ protocol (indexed adapters) (Illumina Inc., USA)

This is a semi-automated method where 96 libraries can be prepared in each run and requires manual intervention. The Truseq™ DNA sample preparation kit (Illumina Inc., USA) containing indexed adapters was used. Test FFPE DNA was sheared using a Covaris S2 sonicator with an average fragment size of 350bp. 50µl fragmented DNA was then mixed with 10µl End Repair Control and 40µl End Repair Mix and incubated for 30 minutes at 30^{o}C. The reaction was cleaned-up using 160µl AMPure XP Beads (Beckman, Coulter, USA) and eluted in 15µl Resuspension Buffer. 2.5µl A-Tailing Control and 12.5µl A-Tailing Mix were mixed with 15µl of clean end-repair reaction to perform the A-addition step. The reaction was incubated for 30 minutes at 37^{o}C. Adapter ligation was immediately performed by mixing 2.5µl Ligation Control, 2.5µl Ligation Mix and 2.5µl DNA Adapter Index with 15µl end-repaired fragments and incubating for 10 minutes at 30^{o}C. The ligation reaction was stopped by the addition of 5µl Stop Ligation Buffer. The ligation reaction was cleaned-up using 42.5µl AMPure XP Beads and eluted in 50µl Resuspension Buffer. A second clean-up step followed by adding 50µl AMpure XP Beads and the final elution was done in 20µl Resuspension Buffer. Size selection of the ligation products were performed after running them on a 2% agarose gel stained with 1x SyBR Gold. Gel bands were cut at 350bp. The bands were extracted from the gel using MinElute Gel Extraction kit (Qiagen, Crawley UK)

eluting in 25µl Qiagen EB buffer. 5µl PCR Cocktail Mix and 25µl PCR Master Mix were mixed with 20µl purified ligated fragments and put on a thermal cycler at 98°C for 30 seconds, 10 cycles of 98°C for 10 seconds, 60°C for 30 seconds and 72°C for 30 seconds, and 72°C for 5 minutes. The PCR reaction was cleaned-up using 50µl AMPure XP Beads and eluted in 30µl Resuspension Buffer. The Truseq™ libraries on the Agilent robot were prepared by Source BioScience, UK.

2.9.1.4 Library preparation protocol number 4 using the NEXTERA™ kit (indexed adapters) (Illumina Inc., USA)

This protocol uses an engineered transposome to randomly fragment the DNA and add adapter sequences to the ends in a single step called tagmentation. 25µl of tagment DNA buffer and 5µl of the tagment enzyme were added per 20µl of DNA sample at a 2.5ng/µl concentration (50ng in total). The reaction was incubated in a thermal cycler at 55°C for 5 minutes. Then the tagmented DNA was purified using the column-based Zymo purification kit. 180µl Zymo DNA binding buffer was mixed with the 50µl reaction, transferred to the Zymo-Spin I-96 plate and centrifuged at 1,300xg at 20°C for 2 minutes. The plate was washed twice after the addition of 300µl wash buffer and centrifugation at 1,300*g* for 2 minutes. The tagmented DNA was eluted in 25µl resuspension buffer. Then the PCR amplification step was performed by mixing 20µl cleaned tagmented DNA with 5µl of index 1 primers, 15µl of the NEXTERA™ PCR master mix and 5µl PCR primer cocktail. The reaction was put on a thermocycler at 72°C for 3 minutes, 98°C for 30 seconds and 5 cycles of 98°C for 10 seconds, 63°C for 30 seconds and 72°C for 3 minutes. Finally, the library was purified and the correct size fragments were selected using AMPure XP beads. The average size of libraries prepared with this protocol should be 500bp. The NEXTERA™ libraries were prepared by Source BioScience, UK.

2.9.1.5 Library preparation protocol number 5 using the NEBNext® Ultra DNA Library Prep kit for Illumina (indexed primers) (New England BioLabs, UK)

The NEBNext® kit was also considered for DNA library preparation from FFPE samples as it was optimised for as little as 5ng fragmented DNA. DNA fragmentation and purification steps were performed as described previously (see 2.9.1.1) except that the elution after purification was done in 14µl buffer EB. 55.5µl fragmented DNA (50ng maximum) was mixed with 3µl End Prep Enzyme Mix and 6.5µl End Repair Reaction Buffer (10x) and incubated on a thermal cycler at 20°C

for 30 minutes and 65°C for 30 minutes. The End Prep product was then mixed with 15µl Blunt/TA Ligase Master Mix, 2.5µl 1:10 diluted NEBNext Adaptor for Illumina and 1µl Ligation Enhancer and incubated at 20°C for 15 minutes in a thermal cycler. 3µl of USER™ Enzyme was immediately added to the ligation mixture and incubation at 37°C for 15 minutes followed. The ligated product was cleaned up using 86.5µl AMPure XP beads with three 80% ethanol washes and elution in 28µl of 0.1x TE. 23µl of clean ligated product was mixed with 25µl NEBNext High Fidelity 2x PCR Master Mix, 1µl Universal Primer and 1µl Index Primer (provided in NEBNext Multiplex Oligos for Illumina, Index Primers Set 1). The PCR amplification was performed using the following thermal cycler conditions: 98°C for 30 seconds, 12 or 15 cycles of 98°C for 10 seconds, 65°C for 30 seconds and 72°C for 30 seconds, and 72°C for 5 minutes. The PCR product was then purified using 50µl AMPure XP beads with 2 80% ethanol washes and eluted in 33µl 0.1x TE.

2.9.2 Whole-genome library preparation using total RNA for RNA-sequencing using the Illumina HiSeq2000

The aim of the work described in this section was to assess the feasibility to use RNA-sequencing technology using FFPE material. As a result RNA extracted from matched frozen and FFPE primary melanoma cell lines (SkMel28, MelJuso and SkMel5) was used to prepare libraries using the newly available TruSeq Stranded Total RNA Sample Preparation kit (Illumina Inc., USA). This protocol was chosen as it has been optimised to prepare libraries both from intact and fragmented RNA, i.e. RNA extracted from FFPE tissue. The quality of the RNA samples extracted from the cells was analysed on the Agilent Tapestation to determine fragmentation time as described below.

2.9.2.1 rRNA-removal

100ng total RNA measured by Nanodrop was used. Ribosomal RNA (rRNA) was depleted and the remaining RNA was purified, fragmented and primed for cDNA synthesis. The lid was pre-heated to 100°C during all thermal cycler programs and all incubations during this protocol were done at room temperature except otherwise stated below. The total RNA was first diluted with nuclease-free water up to a final volume of 10µl. 5µl rRNA Binding Buffer was added to the diluted RNA sample and mixed well. The addition of 5µl Ribo-Zero rRNA Removal Mix-Gold followed and the sample was mixed thoroughly. The sample was then placed on a thermal cycler at 68°C for 5 minutes to denature the RNA. Then, the sample was

left at room temperature for 1 minute. 35µl rRNA Removal Beads were mixed well with the sample for 1 minute. The tube was put on a magnetic stand for 1 minute and the resulting supernatant transferred to a new tube. The old tube was discarded and the new tube was then placed on the magnetic stand for 1 minute, repeating once more to ensure complete removal of all magnetic beads. After the third clean-up on the magnetic stand the supernatant was transferred to a new tube. 193µl RNAClean XP beads were added to a fresh tube and the rRNA-depleted RNA was transferred and mixed with the beads. The tube was incubated at room temperature for 15 minutes, then placed on the magnetic stand and incubated for a further 5 minutes. The supernatant was discarded. Two 200µl 70% ethanol washes were performed and rRNA-depleted RNA was finally eluted in 11µl Resuspension Buffer. 8.5µl clean rRNA-depleted RNA was transferred to a new tube.

2.9.2.2 RNA fragmentation

8.5µl rRNA-depleted RNA sample was fragmented after mixing with 8.5µl Elute, Prime, Fragment High Mix and put on a thermal cycler at 94^{o}C. The fragmentation time differs based on RIN scores and Agilent RNA traces by matching the sample profiles to example profiles illustrated in the protocol. Therefore the fragmentation time for intact frozen RNA samples was 8 minutes and for FFPE RNA was 7 minutes. After fragmentation the thermal cycler was set at 4^{o}C and on reaching temperature the sample was removed and briefly centrifuged.

2.9.2.3 First and second cDNA strand synthesis

SuperScript II was mixed with First Strand Synthesis Act D Mix at a ratio of 1:9 and 8µl of the mix was added to the fragmented RNA sample. The first cDNA strand synthesis was performed on a thermal cycler at 25^{o}C for 10 minutes, 42^{o}C for 15 minutes and 70^{o}C for 15 minutes. The second cDNA strand was then synthesized by adding 5µl Resuspension Buffer and 20µl Second Strand Marking Master Mix to the tube containing the RNA:cDNA sample. The reaction was performed on a thermal cycler at 16^{o}C for 1 hour having pre-heated the lid to 30^{o}C. The double-stranded cDNA reaction was purified by addition of 90µl AMPureXP beads. Two washes with 200µl 80% ethanol were performed and the double-stranded cDNA was eluted in 17.5µl Resuspension Buffer. 15µl clean double-stranded cDNA was transferred to a fresh tube.

2.9.2.4 3' end adenylation, adapters ligation and enrichment

The rest of the procedure includes 3'end adenylation, adapter ligation and PCR enrichment for the completion of the library preparation. 2.5µl Resuspension buffer and 12.5µl A-Tailing Mix were added to the 15µl double stranded (ds)-cDNA and mixed well. The adenylation was performed on a thermal cycler at 37^{o}C for 30 minutes and 70^{o}C for 5 minutes. The adenylated ds-cDNA was then mixed with 2.5µl Resuspension buffer, 2.5µl Ligation Mix and 2.5µl Indexed Adapter. The adapter ligation was performed on a thermal cycler at 30^{o}C for 10 minutes. Then the tube was removed from the thermal-cycler and 5µl Stop Ligation Buffer was mixed with the sample. The ligated sample was cleaned with 42µl AMPure XP beads as described in paragraph 2.9.2.3 and eluted in 52.5µl Resuspension buffer. The 50µl eluted cleaned ligated sample was cleaned again using 50µl AMPure XP bead and eluted in 22.5µl Resupension buffer. 20µl eluted clean ligated sample was transferred to a fresh tube. The enrichment of the DNA fragments was performed after addition of 5µl PCR Primer Cocktail and 25µl PCR Master Mix to the sample. The mixed sample was put on a thermal-cycler for 15 cycles of 98^{o}C for 10 seconds, 60^{o}C for 30 seconds and 72^{o}C for 30 seconds, finishing at 72^{o}C for 5 minutes. The resultant library was cleaned with 50µl AMPure XP beads and eluted in 32.5µl Resuspension buffer. 30µl of the clean eluted library was transferred to a fresh tube and was stored at -20^{o}C before proceeding to cluster generation and sequencing.

2.9.3 Results and future work

Here I present the successful library preparation using matched frozen and FFPE cell line RNA. The Agilent traces of the libraries and the concentrations are presented in Figure 2-7. These libraries were sequenced as 1 per lane but due to time limitations the data have not been analysed and therefore not presented in this thesis. The intent was to explore the feasibility of RNA-sequencing using FFPE material. To identify the reproducibility of the data generated from fragmented RNA, libraries were also prepared from matched frozen cells. Also, microarray data are available from the same frozen cell line RNA to compare the data generated from the two platforms. Quality control of the sequencing data, correlation between matched samples within the same platform and across platforms should be performed. If this analysis shows evidence that good quality data can be generated from FFPE cell line RNA then the RNA-sequencing platform could be tested using RNA from FFPE melanoma samples.

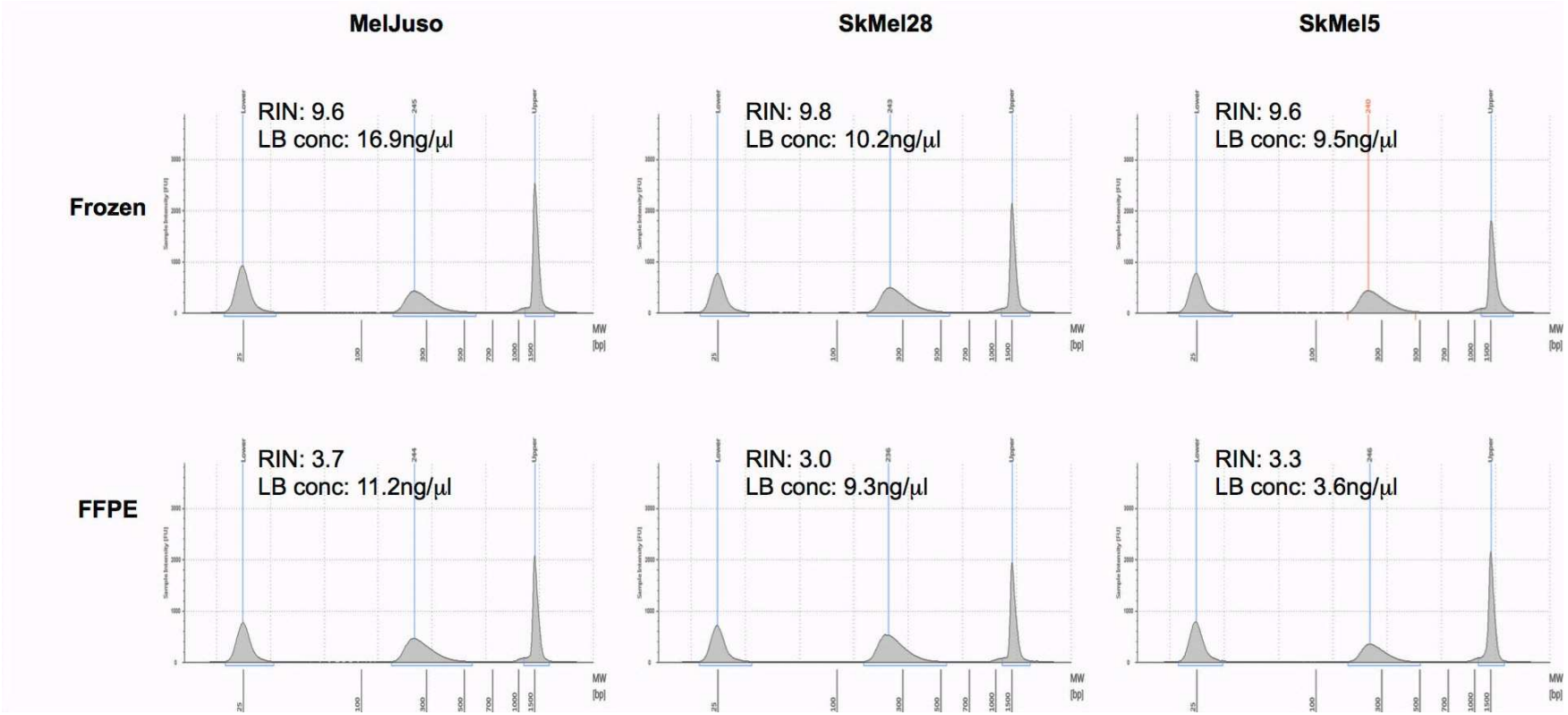


Figure 2-7: Libraries prepared from matched frozen and FFPE cell line RNA.

All libraries were prepared successfully even when the starting material was fragmented as indicated by the RIN number for each sample. The concentration of libraries generated using FFPE samples were lower compared to those from frozen samples, but were sufficient for sequencing.

2.9.4 Cluster generation and sequencing

Prior to sequencing, enriched libraries are clonally amplified through bridge amplification to form clusters of millions of copies of each DNA fragment on the flow cell as described in section 2.9. Cluster generation was performed on the cBOT instrument prior to sequencing on the HiSeq2000.

The DNA library was first diluted to 2nM using Tris-Cl 10mM, pH 8.5 with 0.1% Tween 20. Then, 10µl 20pM DNA library was mixed with 10µl 0.1N NaOH, centrifuged to 280*g* for 1 minute and incubated at room temperature for 5 minutes to denature the DNA into single strands. The 20µl denatured DNA was transferred to a tube containing 980µl pre-chilled Hybridization Buffer, mixed well, pulse vortexed and kept on ice. The fourth lane of the flow cell is the control lane for the 12pM PhiX library which was diluted and denatured in a similar manner. Denatured libraries were loaded onto an eight-tube strip. The eight-tube strip, the flow cell and the thawed reagent plate provided within the appropriate cluster kit version were placed on the cBot and v1.4 software was used. The Truseq PE Cluster Kit v3 with flow cell v3, Truseq SBS Kit v3-HS and HCS v1.4/RTA v1.12 were used for whole genome sequencing on the Illumina HiSeq2000 instrument. Paired-end runs were performed on the HiSeq2000 (2x100bp reads) according to Illumina standard protocols.

The NGS facility team, University of Leeds, performed both cluster generation and sequencing runs.

2.10 Bioinformatics analysis of whole genome next-generation sequencing paired-end data generated using the manual protocol

A pipeline generated by bioinformatician Dr Alastair Droop was used for aligning the sequencing data to a reference genome (Figure 2-8) and decisions made after discussion. Because the data sizes are so large the high performance computing facility at Leeds University (Arc1) was used.

Two files per lane in FASTQ format were collected from the Illumina Hiseq2000 sequencer; one file for read 1 and one for read 2. Each FASTQ file was uploaded to the FASTQC publicly available software (Babraham Bioinformatics, 2010). The FASTQC software generates plots used to assess data quality including quality

score per base, frequency of each base per cycle, GC content per base and overrepresented sequences. The Phred-based quality score represents the error probability for each base (Ewing and Green, 1998). A quality score of 30 is equivalent to the probability of an error base 1 in 1000 times (99.9% base accuracy). Typically, a quality score of greater than 28 is taken to indicate high quality data (Babraham Bioinformatics, 2010). A trimming procedure was followed to ensure that high quality data were used for alignment. Trimming bases with low quality scores has been proposed in the literature to reduce errors in the alignments (Minoche *et al*., 2011). Also, adapter sequences were removed if sequenced. Each sample was determined by the 6-base barcode (5 unique bases and a T at the 6th position). The sequencer needs a high diversity at each position and therefore 2 different 6-base tags were added to each sample and a total number of 10 barcodes were present in each lane.

The reads were then aligned to the human reference genome using BWA (Burrows-Wheeler transform) (Li and Durbin, 2009). The reference used was from Phase 2 of the 1000 genomes project. This reference is based on GRCh37 patch 4 (hg19) (Church *et al*., 2011), but with an updated mitochondrial sequence (NC_012920), herpes simplex sequences (NC_007605) and decoy sequences covering known repetitive regions. The repetitive decoys prevent reads mapping to repetitive elements from being included. The read count data were binned in 10kb regions before further processing.

PCR replicates were removed from the data using SAMtools (Li *et al*., 2009) based on the fact that reads that start and end at the same genomic coordinates must be PCR replicates.

A study in 2011 reported that multicopy regions due to incorrect assembly in the human reference genome lead to false positive CNV peaks in ChIP-sequencing data (Pickrell *et al*., 2011). These regions were listed as problematic by Pickrell *et al* and removed from further analysis (more details in Chapter 4).

GC content can cause uneven coverage of reads in next generation sequencing data (Chen *et al*., 2013). Therefore, base content per window was defined and an analysis of variance (ANOVA) test was run to identify the explained variance by GC content for each sample. For each sample log read counts were modelled as a linear function of chromosome and base content. The read counts were corrected for base content bias based on that model.

Finally the data were median normalised to make the samples comparable. Normalised read count data in 10kb regions were used for further analysis.

Decisions made on removal of PCR replicates, removal of problematic regions and GC correction are discussed in more detail in Chapter 4.

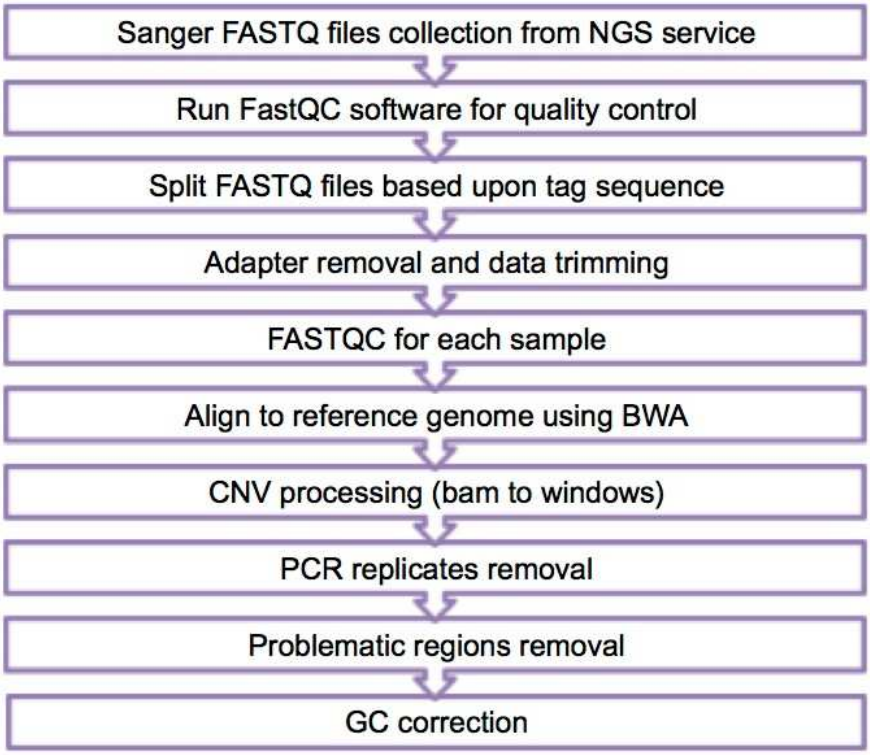


Figure 2-8: Next-generation sequencing data pre-processing workflow.

2.11 Human Osteopontin Immunoassay (Quantikine®, R&D systems)

The Human Osteopontin Immunoassay is a quantitative sandwich enzyme immunoassay and was used for work described in Chapter 6. Samples are added to a 96-well microplate coated with a monoclonal antibody specific for osteopontin. The plate is washed and anything that has not been bound is removed. A substrate solution is added and colour develops the intensity of which is proportional to the amount of osteopontin bound to the antibody. A dilution series of samples with known osteopontin concentrations are used to create a standard curve and from this the concentration of unknown samples is calculated.

The stored EDTA plasma samples from the Leeds Melanoma Cohort were used. In addition samples from 5 patients with melanoma, 4 patients with renal cell carcinoma and 4 healthy people were used to assess assay performance in the

stored plasma samples and make a comparison between plasma and serum osteopontin levels. The methodological work performed is described in more detail in Chapter 6.

A 25-fold dilution of EDTA plasma samples was prepared by combining 10µl sample and 240µl Calibrator Diluent RD5-24. A dilution series of the osteopontin standard sample (starting from 20ng/ml to 0.312ng/ml) was prepared. Low and high level control samples with known concentrations were used to ensure the assay performed as expected (Quantikine® osteopontin ELISA Kit controls, R&D systems). All samples, controls and standards were assayed in duplicate. 100µl of Assay Diluent RD1-6 was added to each well of the microplate and 50µl of standard, control or sample was added per well. The plate was covered with an adhesive film and incubated for 2 hours at room temperature. The content of each well was aspirated and washed by adding 400µl wash buffer, repeating the process three times for a total of four washes using a microplate Wellwash 4 Mk 2 autowasher (Thermo Scientific, USA). After the last wash, any remaining wash buffer was aspirated and the plate was inverted and blotted against clean paper towels. 200µl osteopontin conjugate was added to each well, the plate covered and incubated for 2 hours at room temperature. The aspiration/wash step was repeated as described above and 200µl Substrate Solution was added to each well. The plate was incubated for 30 minutes at room temperature protected from light. Next, 50µl Stop Solution was added to each well and thorough mixing was ensured. The colour in the wells changed from blue to yellow. The optical density of each well was determined within 30 minutes using a microplate reader set to 450nm. Wavelength correction was used when the optical density was measured at 570nm.

2.11.1 Data analysis of output from the osteopontin immunoassay

Standard curves were created by fitting the standard concentrations to a 4 parameter logistic model. Each standard curve was assessed for accuracy and precision of the back calculated standard concentrations. For a curve to be acceptable, precision and accuracy should be ≤20% at the limits of the assay and ≤10% throughout the remaining standards. ≥80% of standards must pass the criteria. Once standard curves have passed the above criteria, unknown concentrations of samples are calculated. As each sample is run twice, the mean concentration of the two replicates is calculated. Precision of the duplicates should be ≤10%. The average of the duplicates was used for statistical analysis which is described in Chapter 6.

2.12 Summary

In this chapter I presented methods used in this thesis. Statistical analysis relevant to a single chapter is presented in the relevant chapter. Here, I also reported the cell line authentication results derived from various cell lines used throughout the thesis, which highlight the importance of cell line authentication to avoid generation of misleading data. Finally, I presented the successful preparation of samples for RNA-sequencing using RNA from matched frozen and FFPE cultured melanoma cells with the intent to explore whether FFPE melanoma tumours can be assessed using this technology in the future.

Part II
Results and Discussion

Chapter 3
Evaluation of whole-genome library preparation techniques using DNA from formalin fixed paraffin embedded (FFPE) melanoma samples for next generation sequencing (NGS)

3.1 Aims

The aims of the work described in this chapter were:

- to evaluate the performance of different whole-genome library preparation techniques using DNA from FFPE melanoma samples;
- to assess if experimental DNA fragmentation is essential before library preparation considering that the FFPE samples might already be fragmented as a result of degradation in formalin;
- to assess the use of bovine serum albumin (BSA) during library preparation as a means of improving the generation of data from highly pigmented samples by comparing libraries prepared with and without BSA

3.2 Background

High-throughput DNA sequencing using NGS technology is a powerful tool to study the complex cancer genome. It can be used to detect copy number alterations, mutations and translocations (Meyerson *et al.*, 2010). Most NGS studies report usage of high quality DNA extracted from cell lines or fresh-frozen tumours. Frozen primary melanomas would inevitably be biased as only rare (large) selected tumours are cryopreserved. However, FFPE tumours which are routinely stored in pathology archives potentially could be a powerful source of genomic material for experimental studies. FFPE tissue available in pathology archives is potentially linked with long term follow-up data enabling the identification of prognostic and predictive biomarkers. For example predictive biomarkers might be identified from stored samples from mature clinical trials. The majority of genomic studies

performed in order to understand the genomics of melanoma have by necessity been carried out using metastatic tumours, but my aim was to identify genomic prognostic biomarkers for primary disease, therefore my intent was to develop methodologies to study FFPE melanoma primaries which are usually small in volume.

In Chapter 1 it was discussed how formalin fixation impedes the use of FFPE samples in various assays. Briefly, formalin affects the quality and quantity of nucleic acids by: the formation of cross-links between nucleic acids and proteins; the addition of methylol groups in adenine and thymine in particular; the introduction of random breaks in the nucleotide sequence (Frankel, 2012, Srinivasan *et al*., 2002). Recent studies have shown that DNA from FFPE tissue can be used successfully for next generation sequencing (Schweiger *et al*., 2009, Adams *et al*., 2012, Menon *et al*., 2012, Wood *et al*., 2010). Wood *et al* first demonstrated comparable results between NGS and aCGH technologies for copy number analysis (Wood *et al*., 2010). Two studies have illustrated that NGS data from FFPE tissue were comparable to data from matched fresh-frozen tissue (Wood *et al*., 2010, Schweiger *et al*., 2009). These studies report the use of 1-3µg DNA, which is not available from most FFPE melanoma samples as the majority of them are small. However, Wood *et al* also reported that lower amounts of FFPE-derived DNA from lung squamous cancer cell lines and oral tumours can be used to successfully generate NGS data to identify copy number alterations (Wood *et al*., 2010). This study showed similar patterns of copy number changes using sequencing data generated from serial dilutions of FFPE-derived DNA (Wood *et al*., 2010). Array CGH has been the leading technology for the identification for copy number alterations, but it is difficult to use with FFPE tissue (larger amounts are usually required) and an amplification step is needed when the amount of starting material is low (Little *et al*., 2006).

The data reported by Wood *et al* were encouraging but considering that primary melanomas are small, and DNA extracted from FFPE samples is limited in volume, a feasibility study was designed where DNA from 10 FFPE metastatic melanoma tumours was used for whole-genome NGS. I extracted the DNA from the samples but they were then processed, multiplexed and sequenced by the NGS facility team at Leeds University and the data were analysed by Dr Henry Wood using published techniques (Wood *et al*., 2010). All ten samples were sequenced twice in one lane on the Illumina GAIIx system using 75bp single-end reads. Dr. Henry Wood analysed the data using tumour:reference ratio where pooled data from the 1000 Genomes project from 10 healthy individuals were used as reference (Wood *et al*.,

2010). Successful generation of data from 90% of the samples as well as the observation of previously described CNAs in melanoma such as gains in chromosomes 6 and 7 and losses in chromosomes 9 and 10 in this feasibility study suggested that generation of reliable NGS data from FFPE melanomas (Figure 3-1) would be possible.

Based on the successful results of the feasibility study, a large-scale study using FFPE primary melanomas was designed to identify recurrent copy number alterations which could be linked to prognosis and driver mutation status. In this chapter I first explored the performance of the whole-genome library preparation technique used in the study by Wood *et al* using 21 DNA samples from primary melanoma tumours.

The process of library preparation has been described in Chapter 2. Briefly, DNA is fragmented, adapters are ligated and a PCR amplification step is performed to enrich the yield of the library. I explored a number of aspects of library preparation.

- DNA extracted from FFPE samples might already be sufficiently fragmented to allow library preparation and therefore I tested the hypothesis that further fragmentation during library preparation might be unnecessary and might be associated with loss of material during the consequent purification step. In the experiment therefore a comparison was performed between samples which were fragmented according to the protocol and others in which this fragmentation step was omitted.
- Melanin has been shown to co-purify with DNA and inhibit downstream applications such as PCR through binding to DNA polymerase (Eckhart *et al*., 2000). The inhibitory effect of melanin on DNA polymerase has however been shown to be reversible when BSA was added in the reaction (Eckhart *et al*., 2000). Leeds Genomics Facility had explored this in DNA from melanomas and showed *KIT* amplification to be successful when 5mg/ml BSA is added in the PCR reaction compared to reactions without BSA using samples enriched with synthetic melanin (data not shown). Here, I describe an investigation of the effects of BSA addition in all the enzymatic reactions during library preparation to prevent the inhibitory effects of melanin by comparing libraries prepared with and without BSA.
- Finally, in this chapter I also explore the comparative performance of a number of alternative library preparation techniques marketed at the time, including automated methods, as manual library preparation is a labour intensive and time consuming process for large-scale studies.

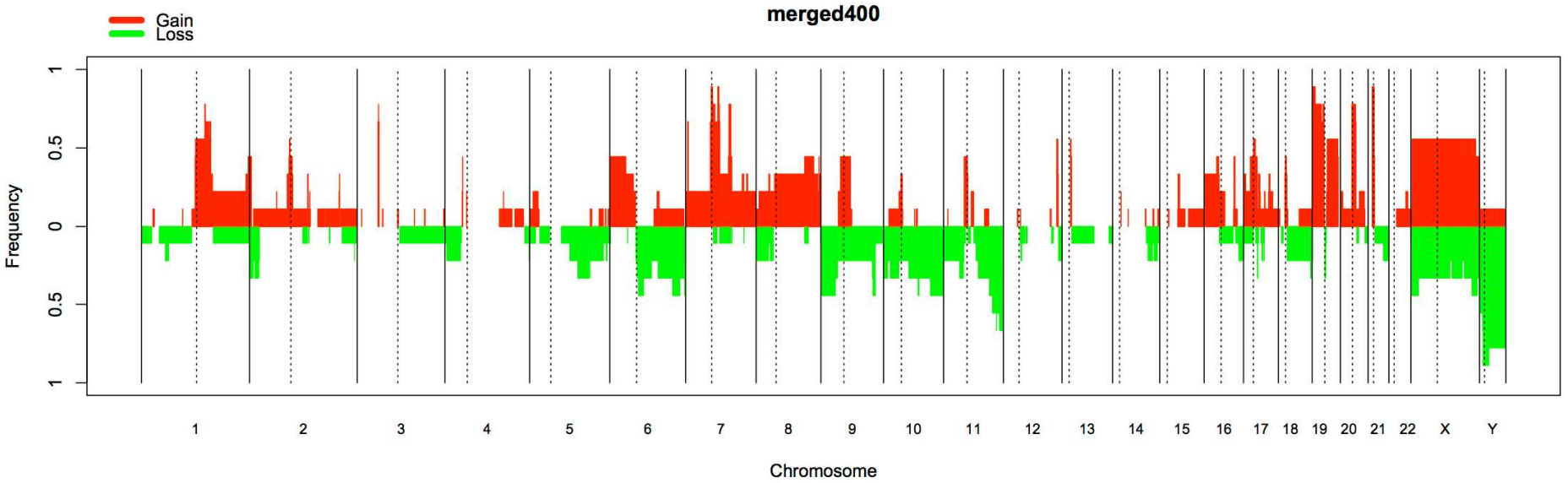


Figure 3-1: Frequency plot of copy number alterations in 9 FFPE metastatic melanoma samples (feasibility study).

The NGS data generated from the FFPE melanoma metastases confirmed copy number alterations such as gains in chromosomes 6 and 7 and losses in chromosomes 9 and 10, known to be altered in melanoma. The feasibility study was considered successful and enabled the design of a large-scale study.

3.3 Methodology

DNA from primary (Leeds Melanoma Cohort) and metastatic samples (details in section 2.2.2) was used to for the evaluation of the different protocols. DNA from 21/116 primary melanomas available in the lab was used to assess the performance of the manual library production protocol described by Wood *et al* (Wood *et al*., 2010). DNA was extracted from metastatic samples for the evaluation of the rest of the protocols using the Qiagen All Prep® FFPE RNA/DNA kit (section 2.3.6.3).

The protocols for each technique evaluated in this chapter were described in detail in section 2.9.1. In summary, the techniques used to test their performance using FFPE-derived DNA were:

a) Library preparation protocol using 5'-end barcodes (manual);

b) Library preparation using 5'-end barcodes on the Beckman Coulter SPRIworks system (SPRIworks);

c) Library preparation on the Agilent Bravo robot using the Illumina TruSeq™ protocol (indexing) (TruSeq);

d) Library preparation using the Illumina NEXTERA™ kit (indexing) (Nextera);

e) Library preparation using the New England BioLabs NEBNext® Ultra DNA Library Prep kit for Illumina (indexing) (NEB).

The protocols and the quality control checkpoints are summarised in Figure 3-2. To optimise the protocols, adapter concentrations and PCR cycles were adjusted according to the DNA input. The performance of the protocols was judged by looking at the size range of the libraries on the Agilent 2100 Bioanalyser or 2200 TapeStation and the concentration of the library, measured using the Quant-iT™ broad range ds-DNA assay kit (Invitrogen by Life Technologies, USA), as described in section 2.6. Fifty nanograms of each library are needed for sequencing. Henceforth, libraries with concentration greater than 2ng/μl in an elution volume of 30μl are considered successful. Examples of libraries judged to be good or contaminated by adapters are shown in Figure 3-3.

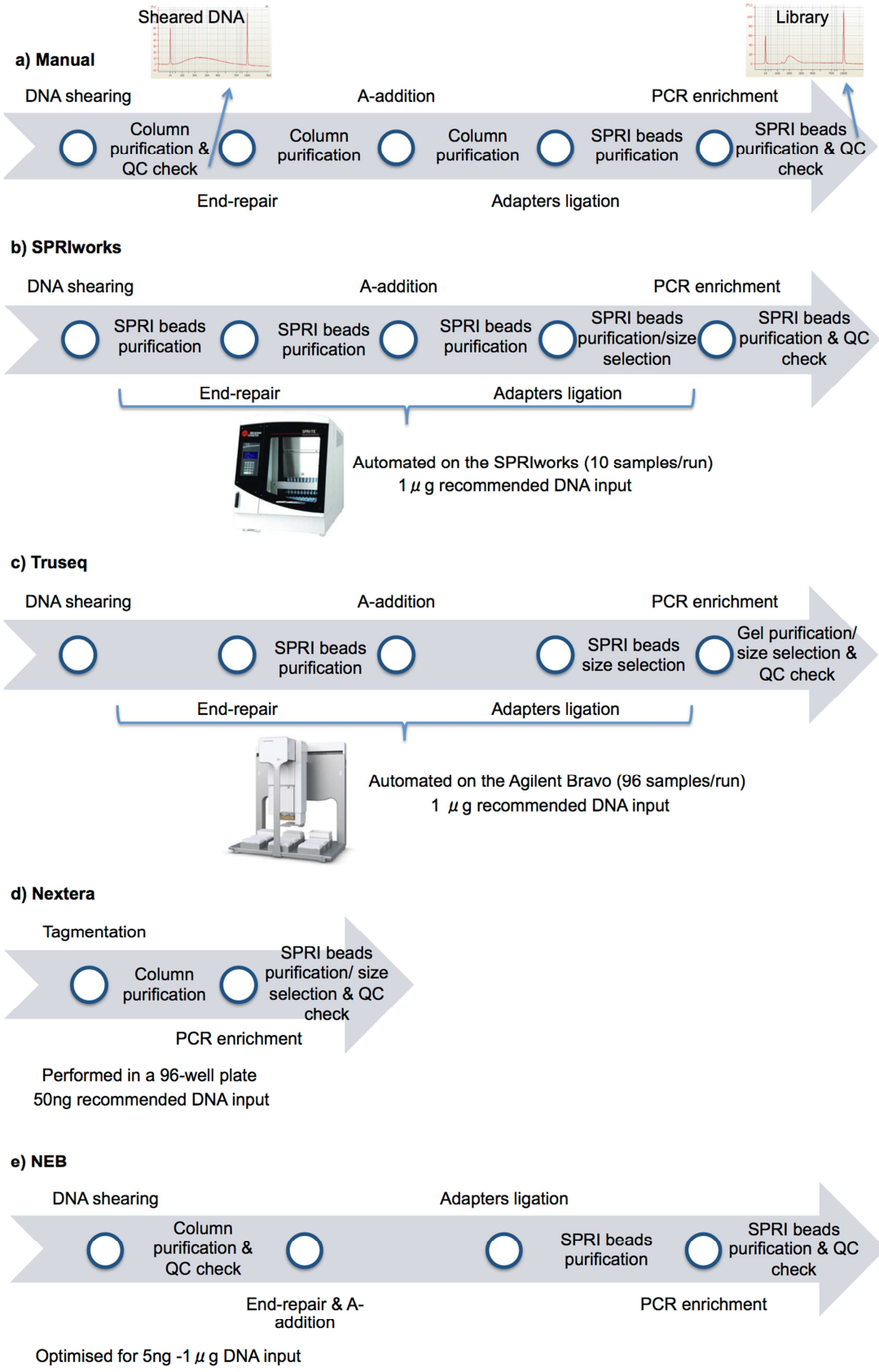


Figure 3-2: Whole-genome library preparation techniques and quality control (QC) checkpoints.

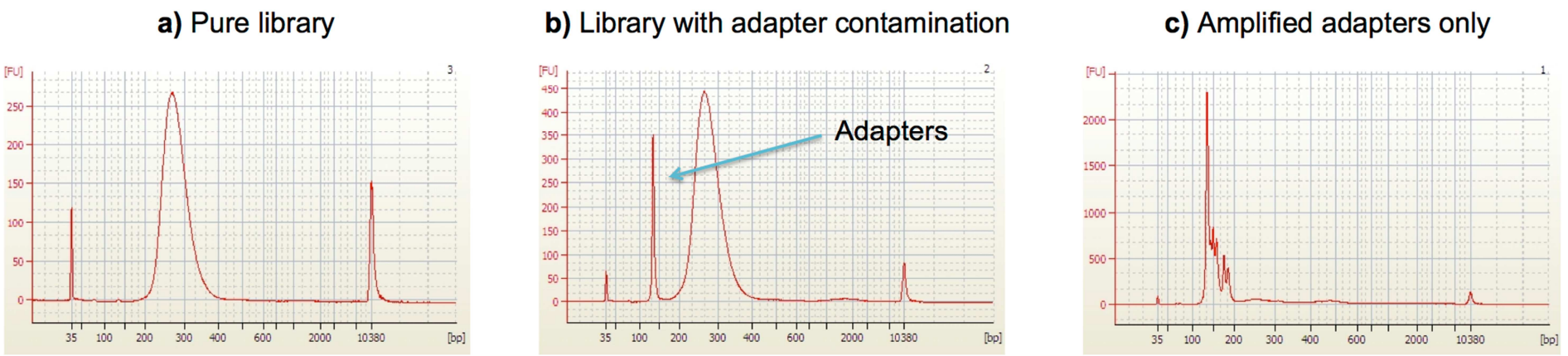


Figure 3-3: Agilent profiles of: a) a pure library free of adapters; b) a library contaminated with adapters; and c) adapters without library amplification.

If the ratio of adapters:DNA input during the ligation step is high, then adapter dimers are formed. Adapter dimers will bind to the flow cell, form clusters and be sequenced reducing the sequencing depth of the tested sample.

3.4 Results and discussion

3.4.1 Library preparation protocol using 5'-end barcodes for Illumina SPRI size selection (manual)

DNA from 116 primary FFPE melanoma samples was available in the lab from previous studies. The median DNA concentration measured using picogreen was 5.73ng/µl (range 0-125ng/µl). DNA is eluted in 30µl buffer and 60ng are used for *BRAF/NRAS* mutation screening routinely in the lab. Therefore, the average volume left from these samples was 20µl and the average DNA yield available for library preparation was 114ng.

Wood *et al* have illustrated that the manual protocol produced sufficient libraries with as little as 5ng FFPE-derived DNA (Wood *et al*., 2010) and therefore I tested this protocol using melanoma samples. First, 21/116 DNA samples with a wide range of concentrations (5.4-42.3ng/µl) were used to test if the manual protocol could produce libraries from small primary melanoma samples. All 21 samples were fragmented according to protocol and their profile was checked on the Agilent Bioanalyser. Seven of the 21 samples were considered unsuitable for library preparation as DNA was not visible on the Agilent profiles. Libraries were successfully prepared from the remaining 14 samples.

3.4.2 BSA addition using the manual protocol

The addition of BSA in PCR reactions can reverse the inhibitory effect of melanin on DNA polymerase. To test the effect of BSA addition during library preparation, 5 metastatic samples were identified, each with 2 highly pigmented cores (assigned melanin score of 3) available from each sample. The DNA extracted from the first core was used to prepare a library by adding 5mg/ml BSA to all enzymatic reactions of the manual protocol and the DNA from the second core was used to prepare a library using the manual protocol without BSA as described by Wood *et al*. The libraries produced with and without BSA are illustrated in Figure 3-4. Due to the presence of adapter dimers the library concentration was not used as a measure of library success. For samples 4 and 5 higher DNA input was used in the presence of BSA and therefore the better performance of the protocol could be either attributed to higher starting material or BSA or a combination of both. For samples 1 and 2 where similar or lower DNA input was used in the presence of BSA however, a library was generated only in the presence of BSA. For sample 3, where lower DNA

input was used in the presence of BSA higher library yield was obtained based on the Agilent trace in the presence of BSA. From this experiment, I concluded that BSA might improve the performance of the library preparation protocol. As a result, I decide to use BSA in all subsequent library preparations using the manual protocol even if samples were not visibly highly pigmented.

I then attempted to make libraries from the 7 out of the original set of 21 samples, where no DNA had been seen on Agilent profiles and I had therefore not attempted to make a library, but this time adding BSA during library preparation. I was able to generate successful libraries from 5 of these 7 samples. The samples that failed had very low DNA yield (18ng and 4ng) compared to the others (30-260ng). The fact that some of these samples prepared successful libraries even if DNA was not visible on the Agilent traces could be in part attributed to BSA addition (2/5 DNAs generated successful libraries had melanin score of 3). However, it is also likely that the standard Agilent kit failed to detect the low amount of fragmented DNA before library preparation as its sensitivity is low and as a result they were erroneously considered as unsuitable. It was therefore considered that the Agilent high sensitivity kits would be more reliable to check the fragmented DNA before library preparation in the next runs.

Figure 3-4 (following page): Libraries prepared in the presence or absence of BSA.

Library concentration is overestimated when there is adapter contamination and therefore library yields were determined based on the Agilent traces. For samples 1 and 2 libraries were only prepared successfully when BSA was used. For sample 3 a library with higher yield was observed in the presence of BSA. For samples 4 and 5 no conclusion for the effect of BSA could be made due to higher DNA input used in the protocol with BSA addition.

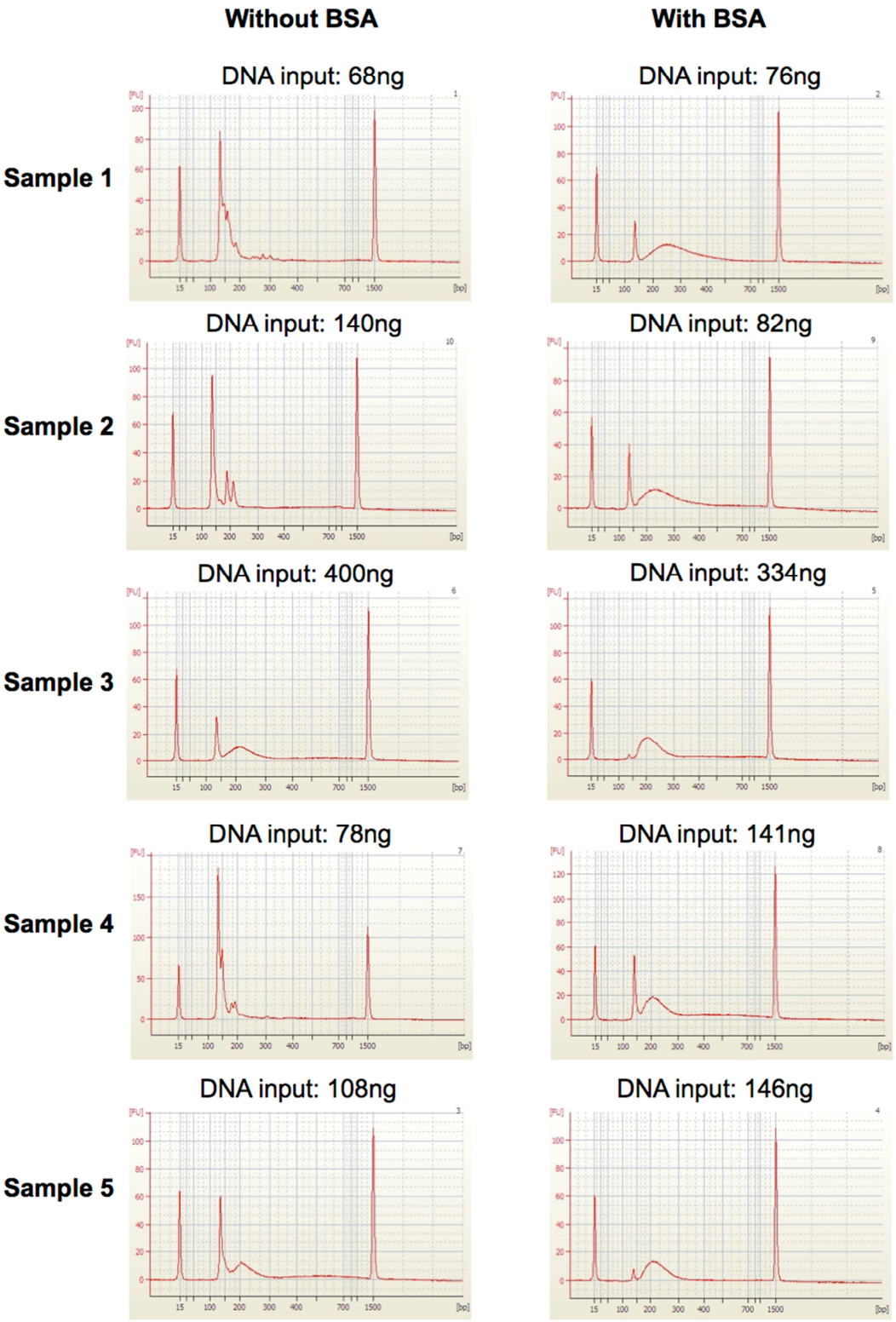


Figure 3-4: Libraries prepared in the presence or absence of BSA.

3.4.3 Adapter contamination

Some libraries were contaminated with adapters (Figure 3-3, Figure 3-4) and after discussion with other colleagues preparing libraries it was decided that the adapters concentration should be diluted proportionally according to the DNA input (Table 3-1).

Table 3-1: Adapter concentration based on the DNA input.

DNA input	Adapters
1µg	Undiluted
500ng	1:2
250ng	1:5
120ng	1:10
60ng	1:20

3.4.4 DNA library preparation using 5'-end barcodes on the Beckman Coulter SPRIworks system (SPRIworks)

Library preparation using the manual protocol is time consuming. Samples could be prepared in batches of 15 over 2 days. Therefore the semi-automated protocol on the Beckman Coulter SPRIworks system was explored as a faster option for a large-scale project. The adapters used in the SPRIworks protocol are the same used in the manual technique. The standard SPRIworks protocol suggests diluting the adapters 1:10, and recommends 1µg DNA input. In the trial runs <150ng DNA input was used to test if this protocol is useful for low DNA yield from primary melanoma samples.

3.4.4.1 First and second SPRIworks trial runs designed to identify the feasibility of the standard protocol using FFPE samples

Ten libraries were prepared using different concentrations of DNA extracted from 2 samples and different dilutions of adapters. Successful libraries were prepared when the standard 1:10 adapter dilution was used but libraries with low yield were prepared when further diluted adapters were used for sample 1 (25ng; Figure 3-5). Notably, better libraries were produced from lower DNA concentrations derived from sample 2 (Figure 3-5). The standard adapter dilution (1:10) was judged necessary and a second experiment was performed to validate this finding. In the second run, no libraries were generated from any of the samples tested (data not shown).

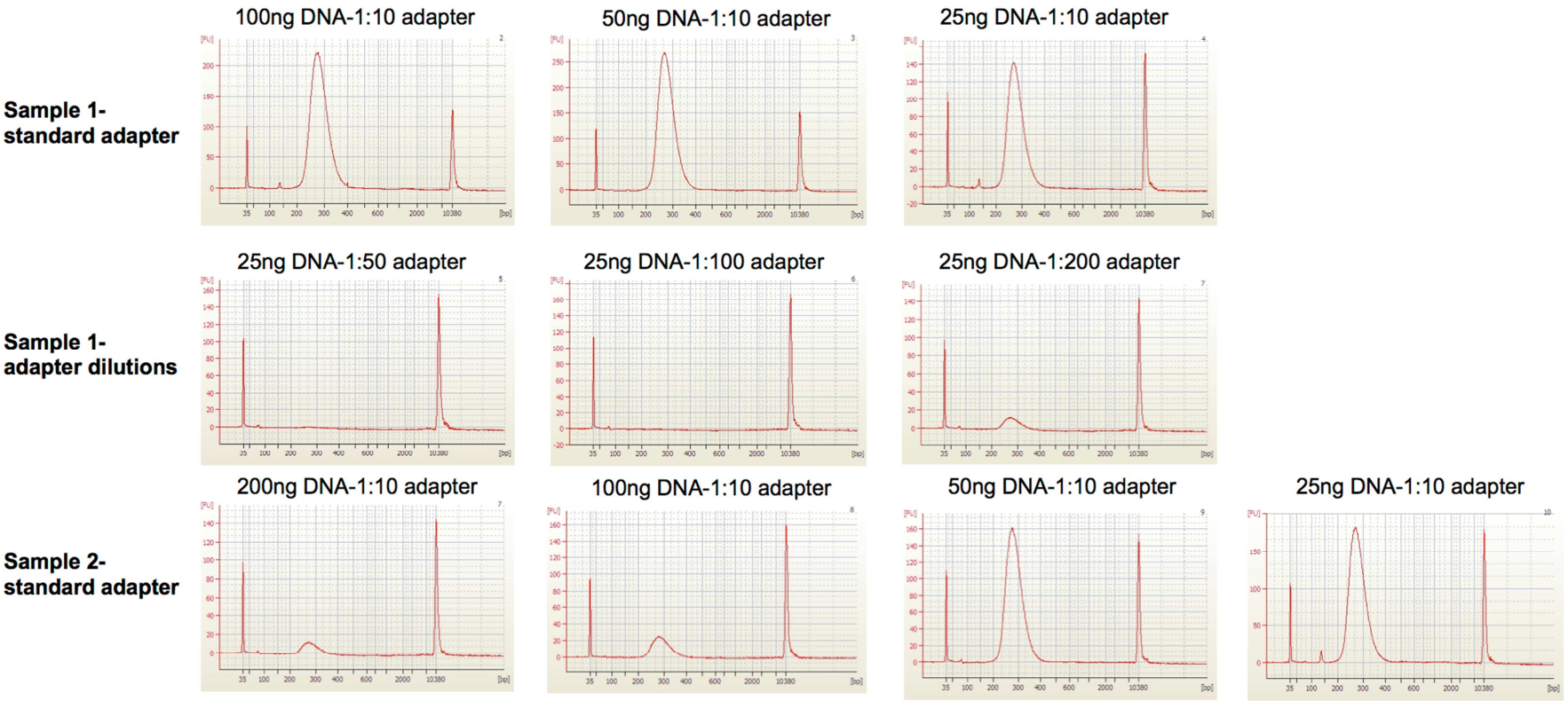


Figure 3-5: Libraries from the first SPRIworks trial run.

Libraries were successfully prepared when the standard 1:10 adapter was used. For sample 1 (25ng DNA input), library preparation was successful when the standard adapter concentration was used while it was unsuccessful when diluted adapters were used. For sample 2, lower library yields were observed when higher DNA input was used.

3.4.4.2 Third SPRIworks trial run designed in an attempt to produce libraries despite the inconsistent results seen in previous runs

Considering that the second run failed due to an experimental error a further trial run was designed to validate the initial finding. Two samples were tested using different dilutions of adapters and a DNA sample derived from a cell line was used as a positive control. Also, matched libraries were generated using the manual protocol to allow direct performance of the two protocols. In contrast with the initial run, here it was clear that the standard 1:10 dilution of adapters resulted in adapter dimers for less than 150ng DNA input. For sample 5 there was no library produced when a 1:10 dilution of adapters was used, a library with a huge adapter peak for a 1:50 dilution of adapters, a library with no adapter peak for a 1:100 dilution and unexpectedly a library with a small adapter peak for 1:200. The libraries from this trial run are shown in Figure 3-6. The manual protocol produced better libraries than the SPRIworks.

In this run libraries were produced using the SPRIworks technique at least for most of the samples but adapter dimers were present when the standard concentration of adapters was used. The technique failed to show a consistent pattern of lower adapter peaks when higher dilutions of adapters were used. We therefore contacted the company to seek advice and a fourth trial run was designed.

Figure 3-6 (following page): Third SPRIworks trial run.

Libraries were prepared from the majority of the samples using SPRIworks but the corresponding libraries using the manual protocol were always of higher yield with no adapter dimers except for one sample. The different colours of the small legends above the Agilent traces represent the pairs of matched libraries using both protocols. For sample 5 the SPRIworks technique failed to show a consistent pattern of lower adapter peaks when higher dilutions of adapters were used.

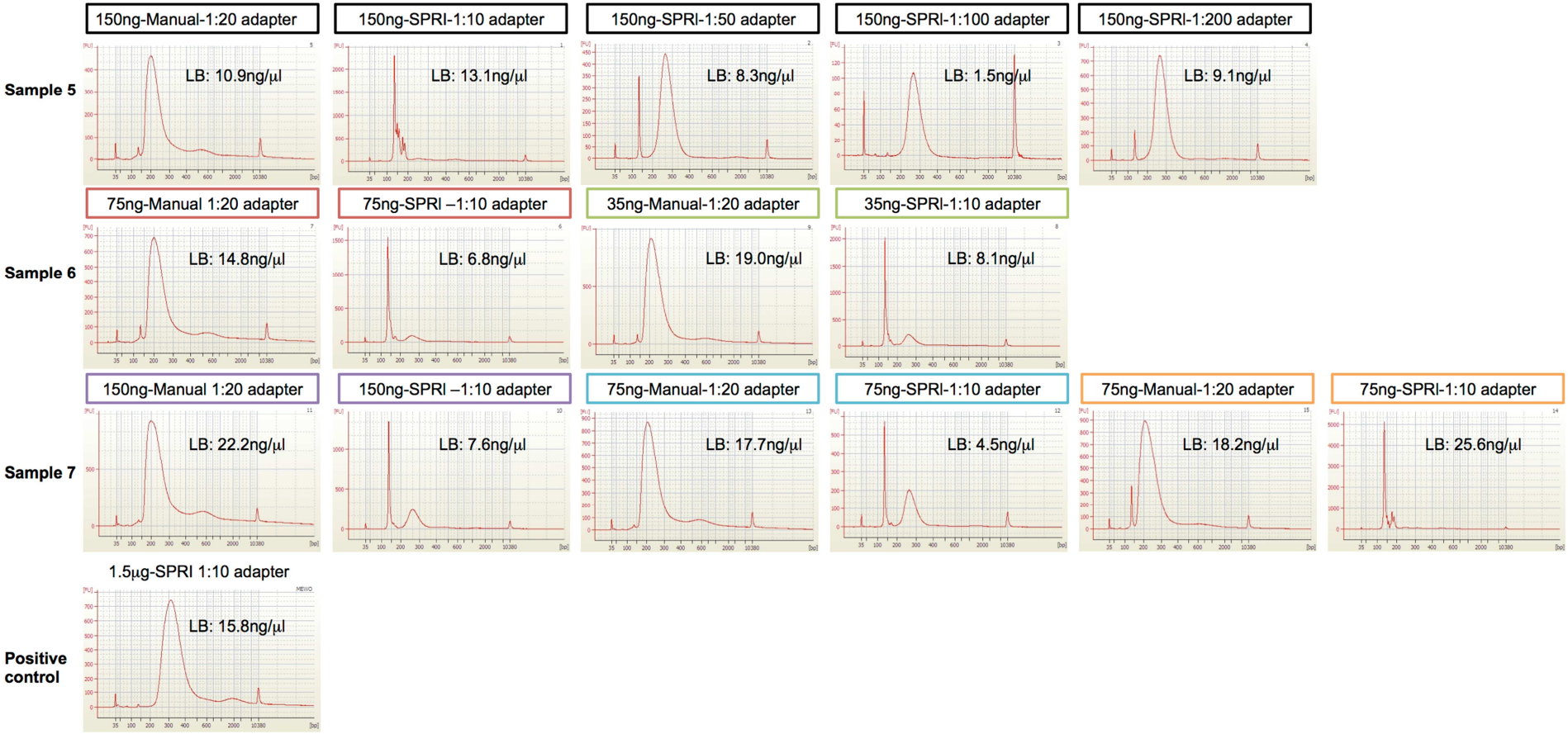


Figure 3-6: Third SPRIworks trial run.

3.4.4.3 Fourth SPRIworks trial run to assess both adapters concentration and fragmentation effects due to the inconsistent results on the previous assays

In the fourth trial run, both experimentally fragmented and non-fragmented samples were used. This run was designed with the help of a representative from Beckman Coulter, when we sought advice regarding the inconsistent results between runs. Generally, a very small library was prepared when non-fragmented samples were used compared to the fragmented samples. This suggests that fragmentation before library preparation is essential. For samples 8 and 9 a consistent lower peak was seen when higher dilution of adapter was used but the opposite was seen for fragmented sample 10. Libraries with higher yield were produced with the manual protocol. The Agilent profiles of the libraries in the fourth run are shown in Figure 3-7.

Figure 3-7 (following page): Fourth SPRIworks trial run.

Generally, a very small library was prepared when non-fragmented samples were used compared to fragmented samples. For samples 8 and 9 a consistent lower peak was seen when higher dilution of adapter was used but the opposite was seen for fragmented sample 10. Libraries with higher yield were produced with the manual protocol. The different colours of the small legends above the Agilent traces represent the pairs of matched SPRIworks libraries starting with fragmented and non-fragmented DNA using the same adapter dilution.

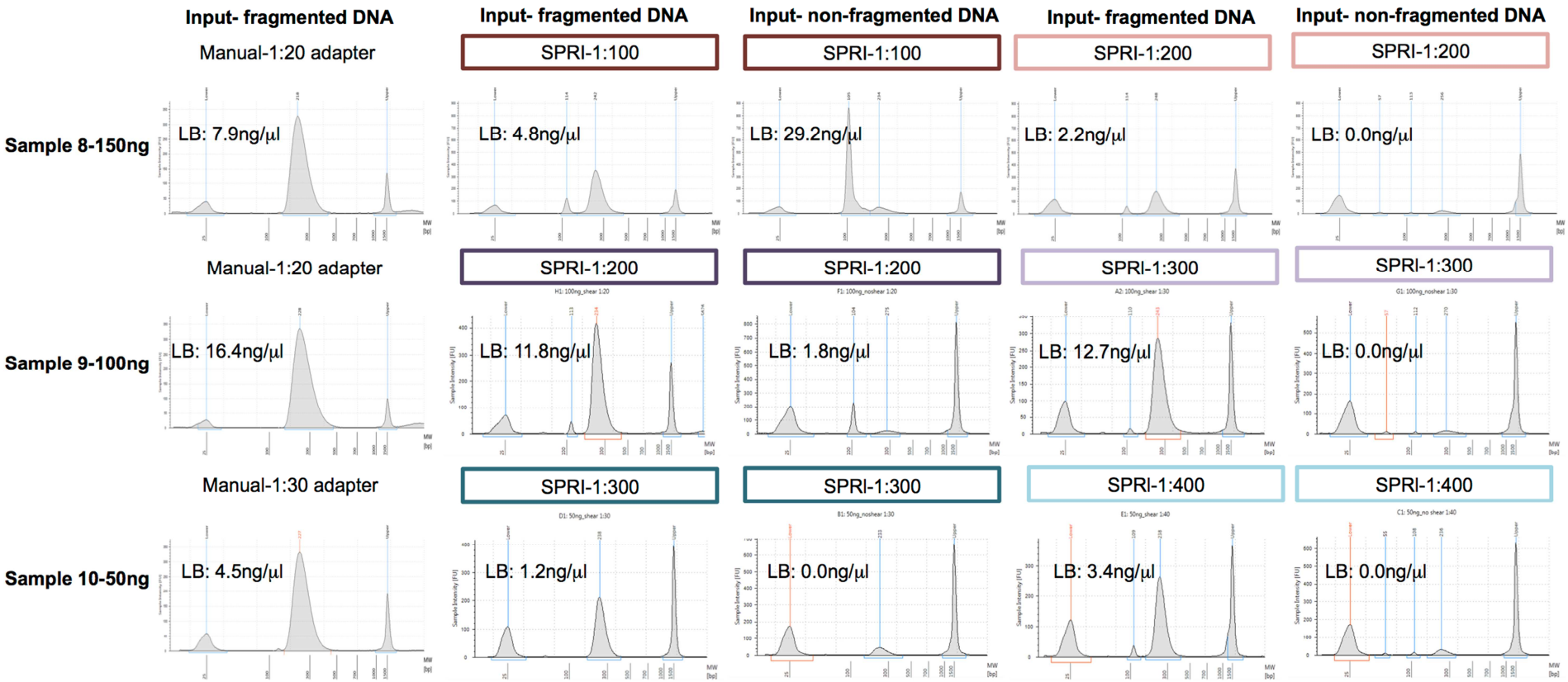


Figure 3-7: Fourth SPRIworks trial run.

3.4.4.4 Summary of the experiments with the SPRIworks technique

The SPRIworks protocol showed inconsistent results across the trial runs and the manual protocol always produced better libraries in terms of less adapter contamination and library yield. Some libraries prepared with the manual protocol still contained adapter peaks and therefore henceforth more diluted adapters were used for the low DNA input (Table 3-2).

Table 3-2: Adapter concentration for <100ng DNA input (updated table).

DNA input	Adapters
85-100ng	1:25
65-85ng	1:30
45-65ng	1:35
30-45ng	1:40
20-30ng	1:50
10-20ng	1:60
<10ng	1:70

3.4.5 DNA library preparation on the Agilent Bravo robot using the Illumina TruSeq™ protocol (indexing) (TruSeq)

The third technique which was tested was the Truseq protocol on the Agilent Bravo robot. Using the Agilent Bravo robot, up to 96 samples can be processed in one batch. The libraries using this technique were prepared by a service provider, Source Bioscience (Manchester, UK). Ten DNA samples were mailed to the company while some DNA was kept in the lab to prepare libraries using the manual protocol for comparison.

Using the SPRIworks system it was identified that smaller libraries were produced using non-fragmented compared to experimentally fragmented DNA (section 3.4.4.3). The DNA samples mailed to the company were run on the Agilent Bioanalyser to identify the degree of fragmentation present in each sample before experimental fragmentation. The Agilent traces showed that the DNA samples were not totally fragmented, although degraded (Figure 3-8). This means that experimental fragmentation is essential to produce a uniform pool of fragments before library preparation.

Libraries were prepared using a 1:100 dilution of adapters following the standard Truseq protocol on the Agilent Bravo robot. This technique uses a tight size selection step and therefore the average length of the libraries was higher compared to the average length of the libraries using the manual protocol (350bp vs 220bp). The Truseq protocol produced libraries without adapter dimers but of lower yield compared to the manual protocol (Figure 3-9). The number of PCR cycles used in the Truseq protocol was lower (10 compared to 18 cycles used in the manual protocol) and therefore could explain the very low library yields observed. A further trial run could be performed by preparing libraries of smaller length and increasing the number of PCR cycles. However, the technique was not further optimised due to limited time available.

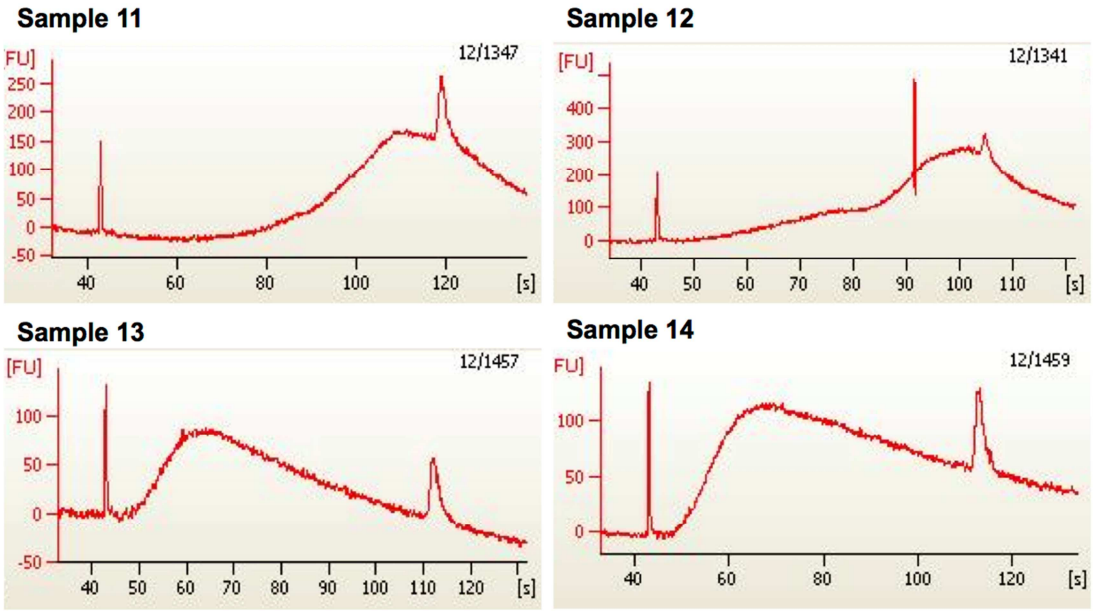


Figure 3-8: Agilent traces of FFPE-derived DNA samples.

Representative examples of DNA samples run on a high sensitivity chip. The fragment size of a high quality genomic DNA sample is usually 15-30kb. The 120s peak represents a length of 10kb. It is shown that samples 11 and 12 were less fragmented than samples 13 and 14. However, experimental fragmentation is needed to produce a uniform pool of fragments before library preparation.

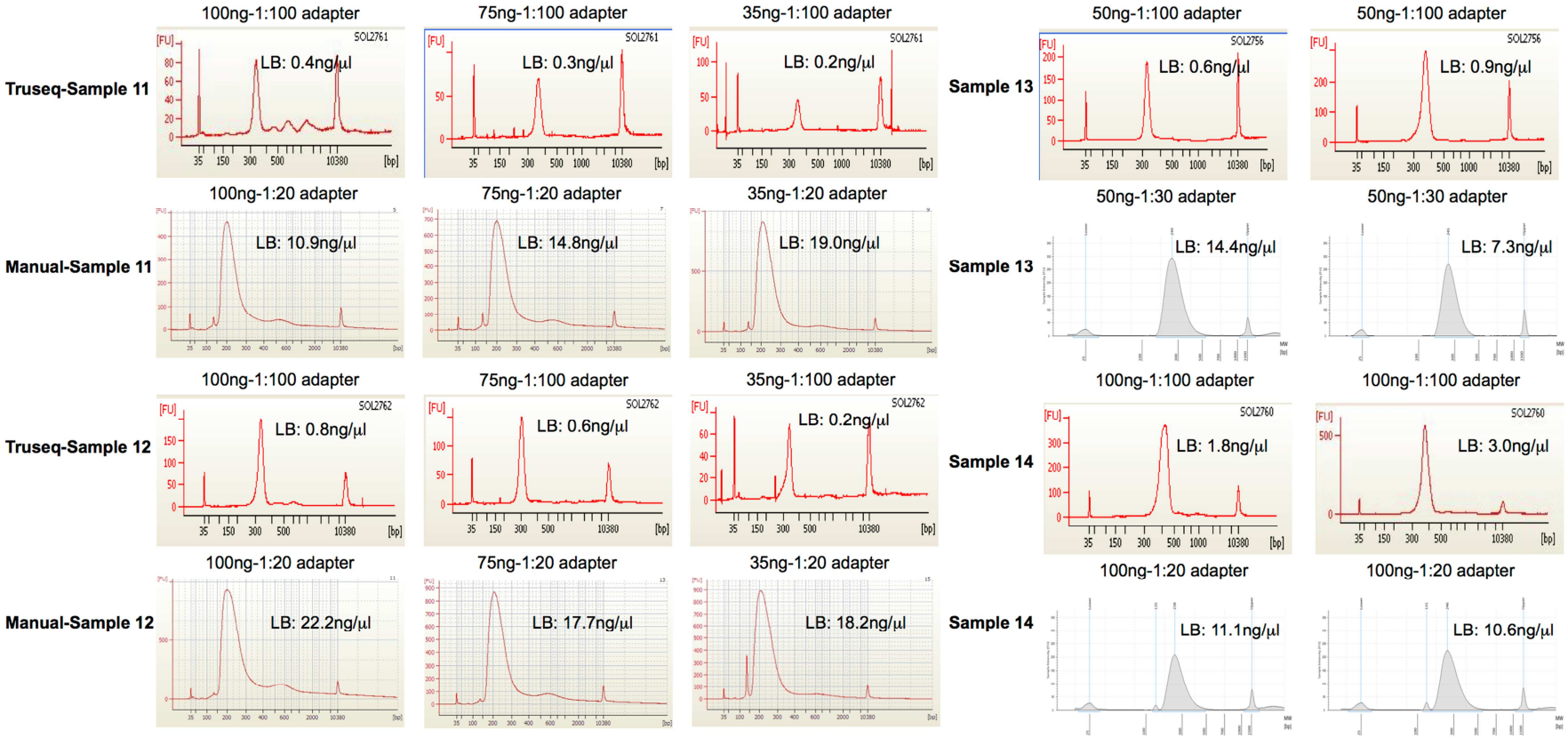


Figure 3-9: Truseq protocol on the Agilent Bravo compared to manual protocol.

The Truseq protocol produced libraries without adapter dimers but of lower yield compared to the manual protocol. The lower number of PCR cycles used in the Truseq protocol is the main reason of the lower library yields but this technique was not further optimised due to time limitations.

3.4.6 DNA library preparation using the Illumina NEXTERA™ kit (indexed) (Nextera)

Libraries using the Nextera protocol were also prepared by Source Bioscience. This protocol is performed in a 96-well plate and therefore 96 samples can be processed simultaneously. Nextera protocol requires only 50ng DNA input and appeared therefore to be a good option, considering the limited amount of DNA available from primary melanomas. Using this technique however 7/8 samples failed to generate libraries and only 1 sample generated a low yield library in both runs, (Figure 3-10). Nextera protocol uses transposase enzymes to fragment the DNA as it was described in section 2.9.1.4. The use of these enzymes can lead to over-fragmentation of the FFPE DNA and might therefore explain the poor results observed but I did not explore this further.

Figure 3-10 (following page): Nextera trial runs using 50ng DNA input.

Using this technique 7/8 samples failed to generate libraries and only 1 sample generated a low yield library in both runs (sample 14). A positive and a negative control sample were included in run B. The transposases (enzymes used to randomly fragment the DNA in Nextera protocol) might lead to over-fragmentation of the FFPE DNA which might therefore explain the poor results observed but I did not have the opportunity to explore this further.

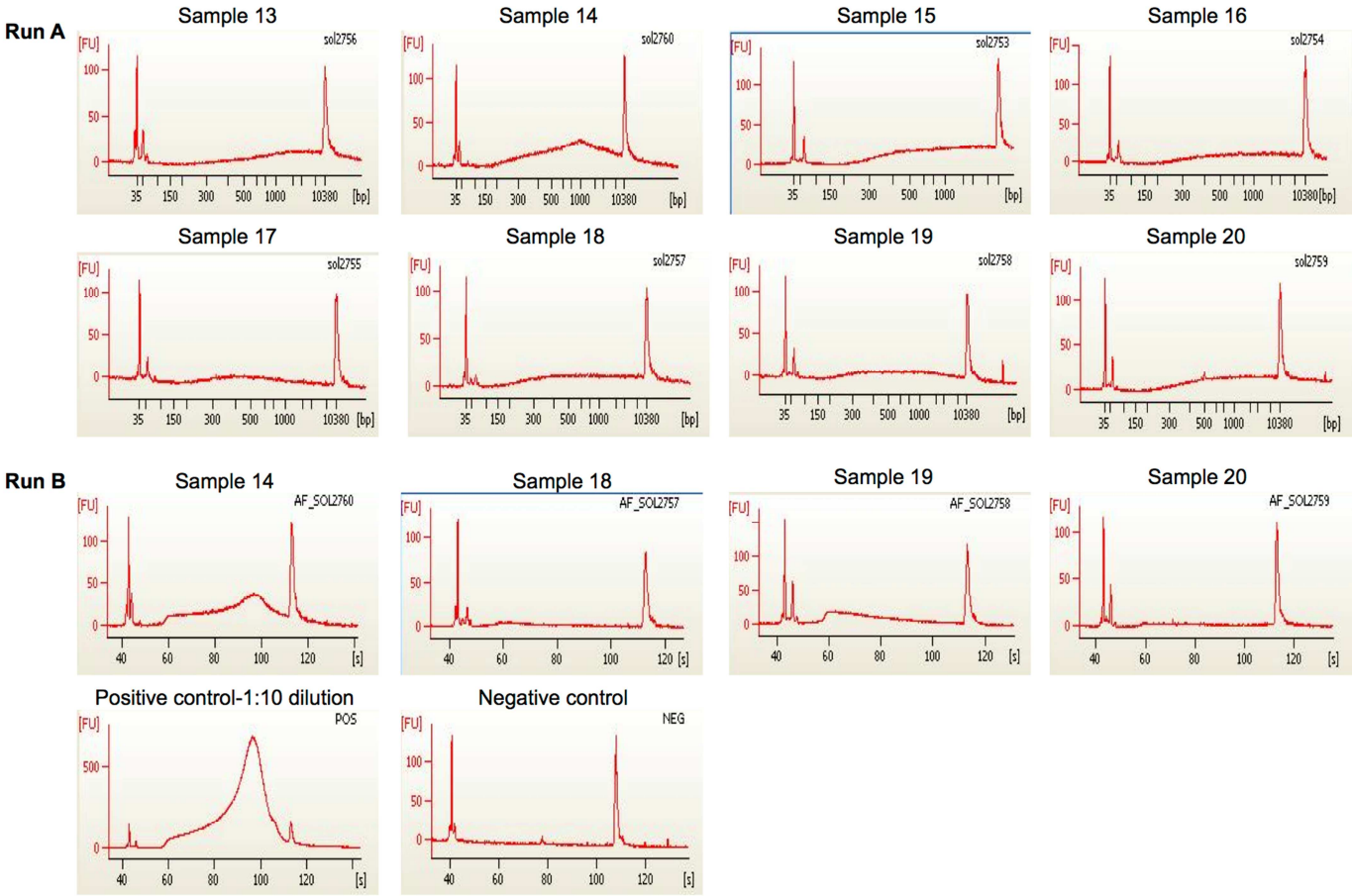


Figure 3-10: Nextera trial runs using 50ng DNA input.

3.4.7 DNA library preparation using the New England BioLabs NEBNext® Ultra DNA Library Prep kit for Illumina (indexing) (NEB)

The NEB protocol recently became available, and it is reported to be optimised for as little as 5ng fragmented DNA. It is quicker than the manual protocol as the End-repair and A-addition steps are performed in a single reaction. Three trial runs were performed using DNA from melanoma samples. Run A was performed by Sally Harrison from the NGS team in Leeds. Good library yields were produced from all 6 samples that were tested using 12 PCR cycles (Figure 3-11). I performed a second run (B) to validate the first result but low library yields were produced. Therefore, run C was performed using 6 more samples with 15 PCR cycles by Helen Snowden from the Genomics Facility in Leeds. In run C high library yields were obtained (Figure 3-11).

Figure 3-11 (following page): Libraries prepared using the NEB protocol.

Run A: Good library yields were produced from 2 samples (samples 21 and 22) using different DNA input using 12 PCR cycles; Run B: Low library yields were produced when different DNA input of sample 22 was used; Run C: Six more samples were tested with 15 PCR cycles and sufficient library yields were obtained.

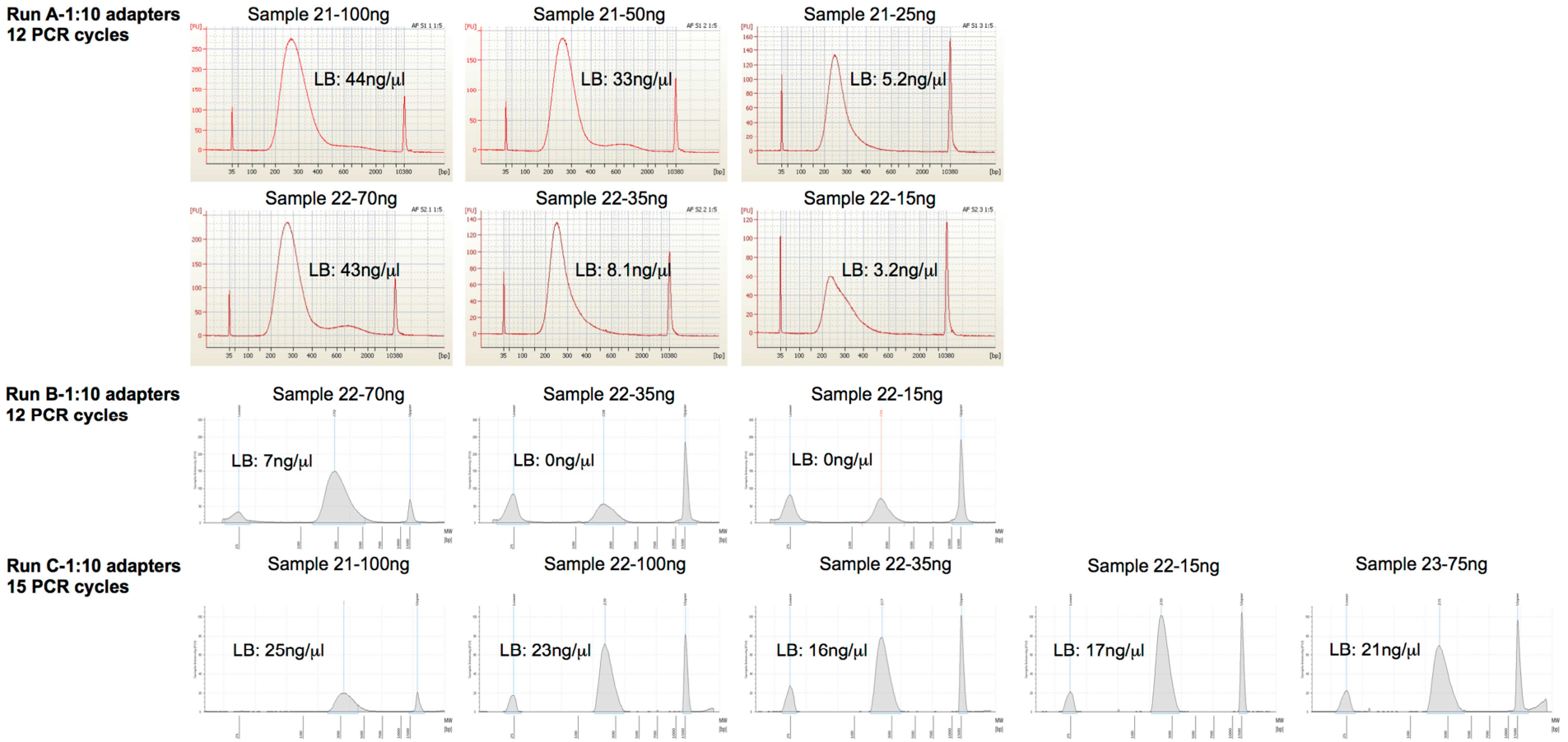


Figure 3-11: Libraries prepared using the NEB protocol.

3.5 Conclusion

Whilst NGS studies represent major developments in cancer research, analysis of FFPE tumours would enable studies using larger numbers of less selected samples, including primary melanomas, with mature follow-up data to allow identification of potential prognostic biomarkers.

Here, I have evaluated different protocols for library preparation for whole genome NGS including comparative performance of alternative library preparation techniques marketed at the time, including automated techniques, useful for large-scale studies.

The concentration of DNA extracted from FFPE primary melanoma samples is low (range 0-125ng/µl) even if TMA cores are used and most available library preparation protocols require large amounts of high quality DNA. Here, I have described the evaluation of the manual protocol for library preparation, which has previously been shown to produce successful libraries from FFPE samples of other tissues (Wood *et al*., 2010, Belvedere *et al*., 2012). I showed that the manual and subsequently the NEB protocol produced libraries successfully from melanoma samples. Optimisation of these protocols included adjustment of adapters concentration according to the DNA input to reduce adapter dimers in the libraries. Finally, it was shown that BSA might improve the performance of the library preparation protocol when libraries with and without BSA were compared and therefore BSA was used in all subsequent library preparations using both the manual and the NEB protocols even if samples were not visibly highly pigmented.

The SPRIworks protocol showed inconsistent results across runs and therefore it was considered unsuitable for samples with low DNA yield. However, it helped us conclude that experimental fragmentation might be essential as smaller libraries were generally generated when experimentally non-fragmented samples were used compared to matched experimentally fragmented samples. This was further confirmed when DNA traces before experimental fragmentation showed that FFPE-derived DNA was insufficiently fragmented, although degraded.

The Truseq protocol on the Agilent Bravo robot produced low yield libraries and further optimisation was needed to prove the good performance of this protocol using low DNA input. Unfortunately, further optimisation using this technique was not performed due to time limitations. Finally the Nextera protocol was proven unsuitable for FFPE samples. This protocol has not been tested using FFPE samples (Illumina technical support). A possible explanation could be that the

transposase enzymes used to fragment DNA in this protocol might result in over fragmentation resulting in failure to generate libraries, however this hypothesis has not been tested.

In conclusion, the manual and the NEB protocols produced higher library yields without adapter contamination compared to the other protocols tested. In Chapter 4 the manual and NEB protocols were used for library preparation from primary melanoma samples that were used for NGS to identify copy number alterations. In the next chapter I report a comparison between the two as both have been used in the preparation of the whole sample set.

Chapter 4
Identification of DNA copy number alterations in FFPE primary melanoma samples

4.1 Aims

The aims of the work described in this chapter were:

- To evaluate the feasibility of whole-genome next generation sequencing (NGS) technology using FFPE-derived DNA from primary melanoma samples. Specific aims were to investigate:
    - the characteristics of the sample associated with successful DNA library preparation;
    - the quality of sequencing data, using the FASTQC quality control tool;
    - the effect of PCR replicates in the data and their removal at analysis;
    - the relationship between previously identified regions in the reference genome reported to be “problematic” and sequencing data;
    - data quality in relation to GC-rich areas of the genome;
    - the characteristics of the sample associated with sequencing coverage.
- To explore the data and copy number alterations using an agnostic approach. Specific aims were to investigate:
    - the association between copy number alterations and survival or factors known to predict survival, such as Breslow thickness, using regression models;
    - the association between *BRAF*/*NRAS* mutation status and DNA copy number alterations using regression models;
    - if a correlation could be demonstrated between copy number and gene expression data generated from the same tumours.
- To explore the validity of the data by looking for copy number alterations previously reported in melanoma, for example evidence consistent with deletion of the 9p21 region of the genome where the tumour suppressor gene *CDKN2A* is located.

4.2 Background

In the previous chapter I have described the evaluation of different library preparation protocols in order to identify the best method to prepare FFPE melanoma samples for whole-genome next generation sequencing. I reported optimisation of the manual and the NEB protocol for small amounts of DNA. Reproducible results were obtained using the above protocols compared to other protocols tested.

Libraries with higher yield were obtained when BSA was added to all enzymatic reactions during library preparation in an attempt to reduce the adverse effect of melanin, a previously reported PCR inhibitor (Eckhart *et al.*, 2000).

Finally, libraries prepared using DNA which was not experimentally fragmented were smaller than libraries prepared using experimentally fragmented DNA derived from the same samples, implying that the DNA samples derived from FFPE melanoma samples are not sufficiently fragmented *per se*. The Agilent traces of 10 FFPE DNA samples confirmed that the DNA was not sufficiently fragmented, although degraded.

Sequencing data from a small feasibility study using 10 FFPE metastatic samples suggested that NGS could be used to study copy number alterations in melanoma. In this chapter I describe a consequent evaluation of the NGS platform for melanoma samples in a large-scale-study using FFPE primary melanomas from the Leeds Melanoma Cohort.

The Leeds Melanoma Cohort is the largest melanoma cohort in the world (consisting of data and samples collected from 2180 patients with a median follow up of 7.2 years). This cohort is powered to identify determinants of outcome. A study was designed to identify CNAs in melanoma and explore their prognostic implication. Libraries from 269 samples have been generated to date (as of September 2013) and sequencing data from the first 75 samples have now been analysed, and I describe the analysis of these samples here. I have worked on developing a pipeline for analysis using these initial 75 samples and will be able to work on the whole data set after submission of this thesis.

In this chapter, then,

- I report an investigation of sample characteristics as potential determinants of successful library preparation.

- I describe an investigation of the quality of the sequencing data and explore the effect of previously reported biases/problematic areas in the reference genome sequence that affect NGS data quality.
- I describe approaches to data analysis. The method used to identify copy number alterations was based on the read depth in pre-specified 10kb regions (otherwise known as windows). The concept is that lower numbers of reads indicate copy number loss while increased numbers of reads indicate copy number gain (Teo *et al*., 2012, Duan *et al*., 2013). This method is known as the depth of coverage (DOC) method (Teo *et al*., 2012, Duan *et al*., 2013). Several tools are available to analyse NGS data using this method and most of them are based on the presence of a reference sample where test:reference ratio is calculated to estimate the number of copies in a region (Teo *et al*., 2012, Duan *et al*., 2013). In this study no matched normal reference sample was sequenced and therefore the estimation of the exact copy number was not possible. Instead, the idea was to identify differences in copy number associated with samples with distinct characteristics. DOC methods rely on the assumption that the sequencing process is uniform (read depth is proportional to the number of copies) (Teo *et al*., 2012, Duan *et al*., 2013). However, previously reported factors, such as GC content, have been shown to be associated with sequencing coverage, suggesting that the above assumption is not always correct (Dohm *et al*., 2008, Aird *et al*., 2011, Chen *et al*., 2013). I report here attempts to reduce error associated with variation such as GC content using computational or statistical models. The analysis was performed by Dr Alastair Droop and Dr Helen Thygesen in an attempt to develop a pipeline for copy number analysis based on the needs of our dataset. Whilst aware that algorithms exist for analyses of this sort, we felt that these are in a state of evolution and we did not wish to assume that the published approaches were optimal.
- Here, I decribe an exploration of the effect of two biases related to GC-content and PCR amplification. Uneven coverage has been reported in GC-rich regions (Dohm *et al*., 2008, Chen *et al*., 2013). A potential explanation of GC-bias could be the different melting behaviour of double-stranded DNA: AT-rich DNA fragments denature at lower temperatures than GC-rich fragments. PCR amplification is thought to be the main cause of GC-bias where amplification of GC-rich regions is less efficient due to their ability to denature at higher temperatures (Dohm *et al*., 2008, Aird *et al*., 2011).

Potential denaturation of AT-rich fragments before adapter ligation has also been suggested to be a source of GC-bias (Dohm *et al.*, 2008).

- PCR amplification has also been reported to have an additional effect, irrespective of GC-bias. PCR replicates occur when multiple copies of a single fragment hybridise on different locations in the flow cell, and therefore multiple clusters of the same original molecule are formed while the goal is to generate 1 cluster per fragment. If too low DNA input is used a higher number of PCR cycles is required to generate sufficient library for sequencing. This results in low-complexity libraries: greater amounts of amplified DNA from fewer areas of the genome: a likely source of bias. Library preparation protocols therefore suggest the use of a limited number of PCR cycles in the enrichment step to reduce PCR replicated sequences in the sequencing data. The number of samples sequenced on a single lane has an effect on the visibility of this problem. For example a low-complexity library run with many other samples on a single lane will result in less PCR replicates sequenced while a low-complexity library run on its own would result in a higher level of PCR replicates sequenced. It seems likely that to avoid PCR replicated sequences there will be an optimal balance between DNA quantity and sequence depth. In this chapter I have attempted to describe efforts to explore this balance, which would be necessary if NGS was to be used for biomarker discovery using small DNA input.
- NGS is a technology under development and possibly more biases or on the other hand better solutions will be identified in the near future. In this chapter I describe the approach we followed to minimise GC-bias and PCR replication and then the statistical analysis performed in an attempt to identify the validity of the data. Both an agnostic approach and a candidate-gene approach were taken to identify previously reported copy number alterations in melanoma.
- Although individual datasets (genomic, transcriptomic, epigenomic, and proteomic) are able to provide useful information, data integration offers a potential advantage to answer complex biological questions (Hawkins *et al.*, 2010). Whole-genome DASL gene expression data from 47 of the 75 samples with NGS data were generated and here I describe an attempt to integrate copy number and gene expression data from matched samples to identify if a correlation could be established. Copy number change is only one cause of variation in gene expression but for some genes at least one would expect reasonable correlation.

4.3 Detailed methodology

4.3.1 Patient samples

Patients from the Leeds Melanoma Cohort Study were selected for the work described in this chapter, which was intended to identify NGS associations with death from melanoma. Samples were chosen for my work in a case control design, where cases were participants who had died, and controls were sex and Breslow thickness matched participants who have not died.

First, all patients who have died from melanoma were identified and were grouped into the following six categories: a) males with a tumour Breslow thickness 0.75-2mm; b) males with a tumour Breslow thickness 2-4mm; c) males with a tumour Breslow thickness >4mm; d) females with a tumour Breslow thickness 0.75-2mm; e) females with a tumour Breslow thickness 2-4mm; and f) females with a tumour Breslow thickness >4mm. Then, samples matched to those from participants in the above groups who have not died from melanoma and were recruited to the study at least 5 years ago were identified. Patients were excluded if: the primary block was not yet available from the local hospital (82/878); the primary block could not be cored (either the tumour was too small or there was an excessively mixed population of cells) (369/878) and DNA was not available (the limited amount of DNA had been used in other assays) (65/878). Out of the 878 samples, 362 (41%) were eligible for inclusion in the study including 15 samples with DNA concentration <1ng/μl which were judged too poor to generate libraries but we wished to test that assumption. The number of patients in each of the above groups after patient exclusion is shown in Figure 4-1.

Two metastatic samples from the Chemotherapy study with sufficient DNA were selected to generate five libraries using different DNA input. These libraries were sequenced but not used for statistical analysis using the clinical variables (only primaries from the Cohort study were analysed).

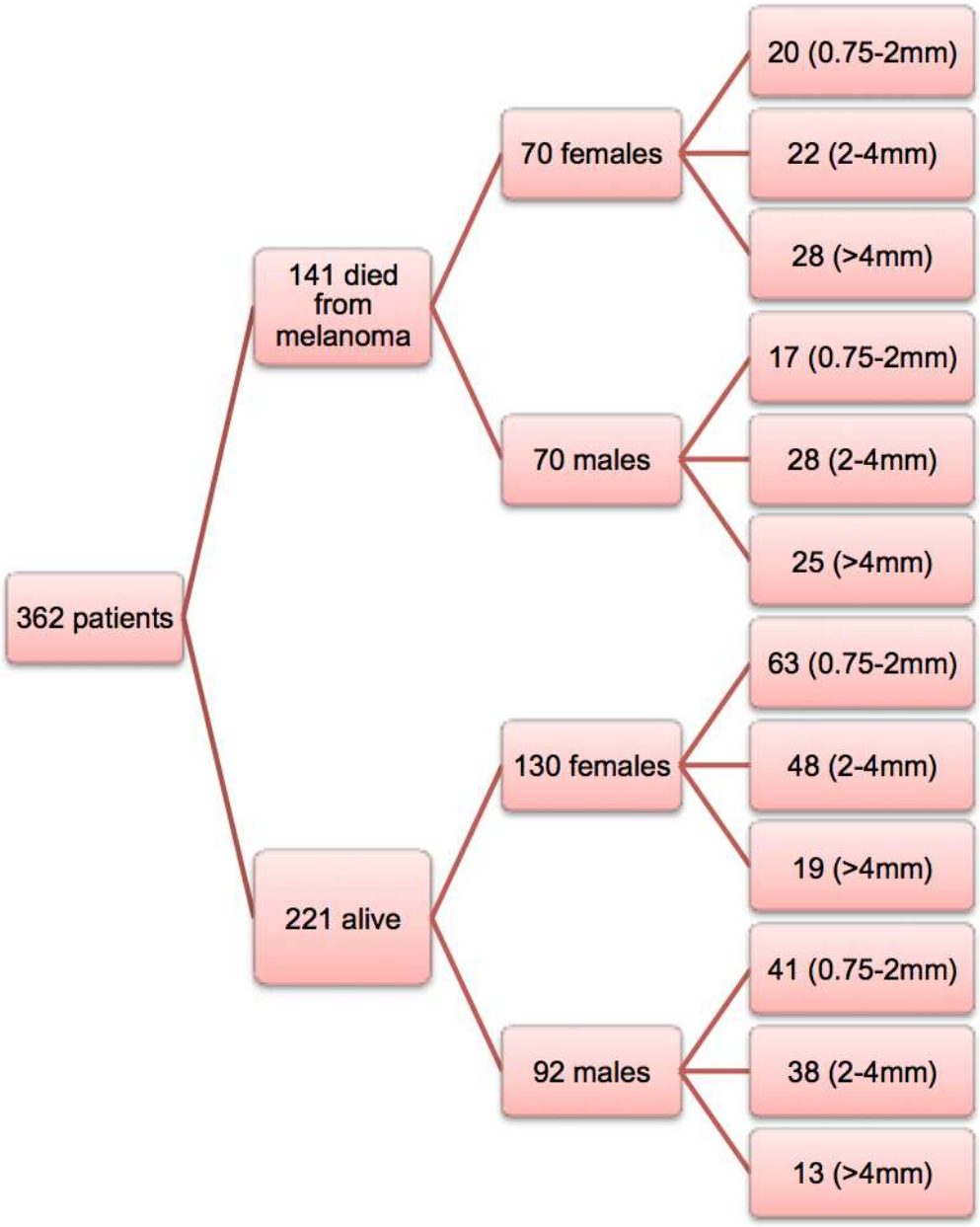


Figure 4-1: Patients included in the study.

The diagram shows the number of patients in each survival group (alive or dead), gender and Breslow thickness group with DNA sample available to use. In brackets the Breslow thickness (mm) group is indicated.

4.3.2 Sample processing

DNA already available in the lab (generated for other experiments) had been extracted using either the Qiagen QiAmp® FFPE DNA micro kit or the Qiagen Allprep® FFPE DNA/RNA kit by a previous PhD student called Rosalyn Jewell as described in sections 2.3.6.2 and 2.3.6.3 respectively. For the rest of the samples, DNA was extracted using the Qiagen Allprep® FFPE DNA/RNA kit either by me, histopathology technician, Minttu Polso, Dr Jonathan Laye or the service provider, GenProbe, Manchester, UK.

Whole-genome library preparation protocols were evaluated and this is described in Chapter 3. Libraries from 71 DNA samples were prepared using the manual protocol as described in section 2.9.1.1. The NEB protocol became available during my PhD and libraries from the rest of the samples were prepared using that protocol as described in section 2.9.1.5.

Samples were sequenced using the Illumina Hiseq2000 instrument by Sally Harrison or Christopher Watson from the NGS facility in the University of Leeds as described in section 2.9.4.

4.3.3 Sequencing data pre-processing

The data outputted from the sequencer were pre-processed by bioinformatician Dr Alastair Droop and statistician Dr Helene Thygesen in conjunction with myself according to the pipeline described in section 2.10. The last 3 steps of the pipeline (removal of problematic regions from downstream analysis, removal of PCR replicates and GC correction) are discussed in this chapter.

4.3.4 Quality control of library preparation

I performed the statistical analysis in this section using Stata (StataCorp., 2011). Library preparation is critical for the successful generation of sequencing data. Therefore, I also explored if DNA yield, degree of pigmentation of the tumour (melanin score), age of the block, and day of the week that the block was received in the NHS histopathology laboratory are determinants of successful library preparation. The day of the week was considered as a potential determinant of successful library preparation as a block processed at the end of the week is likely to be left in formalin for the weekend before processing and this could have led to lower DNA yield extraction from that block as a result of greater degradation of DNA. A successful library was considered present when the concentration of the

library was >2ng/µl and the amplified library was visible on the Agilent profile. Kruskal-Wallis and Spearman's correlation tests were used where appropriate to identify if any of the above tumour or block characteristics determined successful library preparation. In Chapter 3 I have described work in which I established that, library preparation using both the manual and NEB protocols was successful and in this chapter I report a comparison between the two using DNA from 16 samples. Library concentration was compared between the matched samples using the Wilcoxon matched pairs signed rank test.

4.3.5 Quality control of sequencing data

Libraries from 61 unique samples from the Cohort study and 14 replicate pairs were sequenced in the first batch and I will report here my analyses of this batch. Subsequent batches are being processed currently and I intend to analyse the whole data set after submission of this thesis. The FASTQC software (Babraham Bioinformatics, 2010) was used to assess quality of the sequencing data as described in section 2.10.

One of the steps during data pre-processing that we considered as a means to avoid bias was removal of PCR replicates. PCR replicates are defined as sequences with the same start and end coordinates (Li *et al*., 2009). Normally, only a single replicate read would be included in the final dataset. I first investigated whether the frequency of the PCR replicates removed for each sample was associated with DNA yield, melanin score, age of the block, and day of the week that the block was processed using Spearman's correlation and Kruskal-Wallis tests where appropriate, to identify if any of the tumour characteristics predict the frequency of PCR replicates observed. To assess if removal of PCR replicates is uniform across the genome, the frequency of the reads removed per chromosome, divided by the total number of reads for that chromosome was plotted for all samples. The decision we ultimately made to remove PCR replicates before final analysis is based on the work described here.

4.3.6 Technical replicate samples

Technical replicates were sequenced to assess the reproducibility of NGS technology using FFPE-derived DNA. Here, different technical replicates were assessed: a) replicates of the same libraries derived from 4 patient samples; b) 5 libraries which were generated using different DNA input from 2 patients (3 libraries using 250ng, 100ng and 25ng DNA from patient 1; 2 libraries using 100ng and 25ng

DNA from patient 2); and c) 10 libraries derived from 5 patients of which 1 library was prepared using 15 PCR cycles and 1 using 18 cycles. Spearman's correlation tests were performed between the read counts for each replicate pair.

4.3.7 Genome-wide statistical analysis of sequencing data

Linear, logistic and Cox regression models were constructed to identify if read counts from single 10kb windows were associated with the following variables: Breslow thickness (mm); mitotic rate (per mm^2); presence of tumour ulceration; AJCC stage (I/II/III); melanoma-specific survival; tumour site (Head/Neck/Trunk or Limbs/Acral); primary subtype (superficial spreading/lentigo maligna or nodular melanoma); and mutation status (*BRAF* versus *NRAS* versus none). Multiple testing correction was performed by applying false discovery rate (FDR) to the p-values (Benjamini and Hochberg, 1995). A threshold of FDR<0.1 was used. Median smoothing (using a bandwidth of 501 windows) of the coefficients was performed to reduce the noise. The threshold applied to the smoothed coefficients was selected so that the number of selected windows was equal to the number of windows with an FDR<0.1 prior to smoothing. The statistical analysis was performed by Dr Helene Thygesen using the statistical package R (R Development Core Team, 2012) although I was intimately involved in the process.

4.3.8 Integration of sequencing and gene expression data

Gene expression data using the Illumina whole-genome DASL array were generated from 47 of the 61 unique samples from the Cohort study with NGS data available. Gene expression and NGS data were both used on a log base 2 scale.

Gene expression DASL data were pre-processed at the probe level and, as described before, multiple probes on the array target a specific gene. In the rest of the chapter I have used the words probe and gene interchangeably.

First, the DASL probes were merged with the NGS data only for the 10kb regions significantly associated with tumour mitotic rate (based on the smoothed coefficients). A Spearman's correlation coefficient between the signal of a DASL probe and the read count of the corresponding NGS 10kb window where the probe was located was calculated. Multiple testing correction was performed and a threshold of FDR<0.1 was applied.

Secondly, to reduce noise, median smoothing (using a bandwidth of 501 windows) of the NGS read counts was performed and the DASL data were merged with the

NGS data which were significantly associated with tumour mitotic rate. Here, Spearman's correlation coefficients between the signal of a DASL probe and the smoothed read count of the NGS 10kb region where the probe was located were calculated.

Third, 15 genes reported previously to be associated with melanoma/melanoma prognosis in terms of gene expression or DNA copy number alterations/mutations were selected (*SPP1, RAD52, TOP2A, RAD54B, RAD51, CDKN2A, EGFR, PTEN, TERT, BRAF, NRAS, MITF, CKIT, CMYC, MC1R*). The DASL probes targeting each one of the above genes and the NGS 10kb region where they were located were identified. The correlation(s) between the read count of the window and the probe(s) targeting the gene were calculated using the Spearman's correlation test.

I recently identified that the DASL file including probe coordinates provided by Illumina was incorrect. Some coordinates were based on the human reference genome version hg18 while some others were based on version hg19. This means that a proportion of DASL probes have not mapped against the correct 10kb NGS window. Therefore, the data presented here must be treated with caution. The data will be re-analysed using the correct probe coordinates after the submission of this thesis.

4.3.9 Pathway analysis

The DASL probes significantly correlated with the smoothed read counts from NGS windows associated with tumour mitotic rate were used to identify significantly overrepresented pathways using MetaCore™, a commercially available pathway database and data analysis tool (Thomson Reuters, USA). MetaCore™ offers a supervised and informed meta-data collection that was reported to out-perform other publicly or commercially available pathway databases in terms of completeness and accuracy when compared with experimental standards (Shmelkov *et al.*, 2011).

4.4 Results

4.4.1 DNA concentrations, failure of library preparation and associations with tumour characteristics

Both DNA concentrations and rate of failure of library preparation (as defined in section 4.3.4) were assessed in relation to tumour characteristics.

DNA concentration was inversely correlated with melanin score (borderline statistical significance, Spearman's rho=-0.14, p=0.06; Table 4-1). There was no significant effect of age of the block or day of the week which the sample was received by the NHS laboratory on the likelihood of successful library preparation.

The median DNA input used for library preparation was 100ng/µl (range 0.00-1000.00ng/µl) as measured using picogreen. Library preparation was successful in 92% (247/269) of the samples processed up to the presentation of this thesis. Associations between library status (successful or failed) and tumour characteristics are shown in Table 4-2. The median DNA yield used for samples which failed library preparation (26.98ng/µl) was significantly lower compared to the median DNA yield used for samples which prepared libraries successfully (100ng/µl) (Mann-Whitney, z=-5.06, p<0.001; Table 4-2). DNA input was the only statistically significant predictor of library preparation failure although age of the block showed a borderline association (Mann-Whitney, z=1.90, p=0.06; Table 4-2,)

.

Table 4-1: Associations between DNA concentrations and tumour characteristics.

	Samples	Median DNA concentration, (range)	Associations between measure and DNA concentration
DNA concentration (ng/µl)	269	7.30 (0.00-89.40)	
Melanin score	172		**Spearman's rho=-0.14, p=0.06**
0	42	11.45 (1.00-124.69)	
1	38	12.75 (0.10-116.09)	
2	46	15.91 (0.00-77.49)	
3	46	5.54 (0.00-57.48)	
Age of block, years	267		Spearman's rho=-0.07, p=0.28
Day of the week block processed	267		Kruskal Wallis, χ^2=5.18, p=0.52
Monday	49	8.20 (0.60-89.40)	
Tuesday	66	7.57 (0.00-47.23)	
Wednesday	55	6.20 (0.72-124.69)	
Thursday	43	9.13 (0.00-116.09)	
Friday	52	7.40 (0.26-57.48)	
Saturday	1	9.19	
Sunday	1	39.98	

Table 4-2: Associations between library status and tumour characteristics.

	Failed libraries	Successful libraries	Association between measure and library success
Libraries, n (%)	22 (8.18)	247 (91.82)	
DNA input (ng), median (range)	26.98 (0.00-287.48)	100 (0.00-1000.00)	**Mann-Whitney, z=-5.06, p<0.001**
Protocol, n (%)			
Manual	6 (27.27)	74 (29.96)	χ^2=0.07, p=0.79
NEB	16 (72.73)	173 (70.04)	
Melanin score, n (%)	For 10 samples	For 162 samples	
0	2 (20.00)	40 (24.69)	
1	3 (30.00)	35 (21.60)	Fisher's exact=0.92
2	2 (20.00)	44 (27.16)	
3	3 (30.00)	43 (26.54)	
Age of block, years, median (range)	11.19 (4.89-14.27)	10.02 (2.63-20.38)	Mann-Whitney, z=1.90, p=0.06
Day of the week block processed, n (%)			
Monday	3 (13.64)	46 (18.78)	
Tuesday	8 (36.36)	58 (23.67)	
Wednesday	4 (18.18)	52 (21.22)	Fisher's exact=0.62
Thursday	4 (18.18)	38 (15.51)	
Friday	3 (13.64)	49 (20.00)	
Saturday	0 (0.00)	1 (0.41)	
Sunday	0 (0.00)	1 (0.41)	

4.4.2 Manual versus NEB protocol

In Chapter 3 I have reported experiments which suggested that both the manual and NEB protocols generated libraries which met my criteria for adequacy. The NEB protocol was identified to be more efficient as less PCR cycles were needed compared to the manual protocol. It also uses the new barcoding system (indexing) as described in section 2.9.1.5 the advantages of which are discussed later (in section 4.5.3). As it was expedient for me to change to the NEB protocol I report here an experiment to test consistency between libraries generated from the same tumour block using the two protocols. Sixteen samples that had generated a successful library using the manual protocol for which a second tumour core was available were used. The second core was used to extract DNA and libraries were also generated using the newer NEB protocol. Similar DNA yields were used for library preparation using both protocols except for two samples where less DNA was used for the NEB library preparation by chance as less DNA could be extracted from the second core. We failed to prepare a library using the NEB protocol from one sample. The manual protocol uses 18 PCR cycles while the NEB protocol uses 15 PCR cycles. The libraries generated using both protocols were free of adapters and therefore library concentrations from the matched samples were compared. Higher library concentrations were observed when the NEB was compared to the manual protocol (median library concentration 13.90ng/µl, 17.10ng/µl for the manual and the NEB protocol, respectively) although the difference was not statistically significant.

Table 4-3: Matched libraries using the manual and NEB protocols

	Manual protocol	NEB protocol	Test statistic
Successful libraries, n (%)	16 (100%)	15 (93.75)	Fisher's exact=1.00
DNA input, ng, median (range)	100.00 (44.5-100.00)	100.00 (23.19-100.00)	Mann-Whitney signed rank test, z=1.99, p=0.05
Library concentration, median (range)	13.90 (4.52-25.90)	17.10 (0.00-36.81)	Mann-Whitney signed rank test, z=-0.93, p=0.35

4.4.3 Sequencing data using the manual protocol only

As above, and as presented in Chapter 3, the NEB protocol performed better than the manual. Therefore, although 75 libraries (61 unique samples and 14 replicates) were prepared using the manual protocol of which sequencing data are presented in this thesis, subsequent samples were prepared using the NEB protocol.

4.4.4 Quality control of the sequencing data

FASTQC software (Babraham Bioinformatics, 2010) was used to assess the quality of the sequencing data, looking at the quality score per base, sequence content per base and overrepresented sequences. The quality score for each base is assigned by the Illumina base-calling software and represents the probability that a base has been called incorrectly. For example a quality score of 30 is equivalent to the probability of an error base 1 in 1000 times (99.9% base accuracy). Typically, a quality score of greater than 28 is taken to indicate high quality data, whilst a quality score of less than 20 indicates very low quality (the background colouration in the figures generated by FASTQC show this consensus). First, the quality control of the sequencing data from each lane was assessed. In Figure 4-2, a number of issues observed have been highlighted. The first 6 bases represent the barcodes used to identify each sample where a T is always present at the 6th position. The T at the end of the barcode (part of the adapter) is needed for the adapter to ligate to DNA fragments. Low quality scores were observed at the 6th base of pair 2 where the T is present and at the subsequent 4 bases. Interestingly, low quality scores were not assigned to the first 10 bases of read pair 1. Bases from both read pairs were also assigned lower quality scores towards the end of the run.

One hundred base pair fragments are sequenced, so if fragments are shorter than 100bp then the end of the sequencing read will contain Illumina adapter sequences. We therefore removed adapter sequences at the 3' end of the sequences as described in section 2.10. This was highlighted by the FASTQC software where Illumina adapter sequences were shown to be overrepresented in the data.

Based on the above observations, the pre-processing pipeline was adjusted to include a trimming procedure of: the bases at the 3' end of the fragments representing Illumina adapters; the bases after the 90th cycle; and the first 10 bases including the 6-base barcode. In Figure 4-2, we can see that sequencing failed during 60th-75th cycles in a single run due to an air conditioning failure, affecting 10 samples and therefore bases after the 50th position were trimmed from these samples.

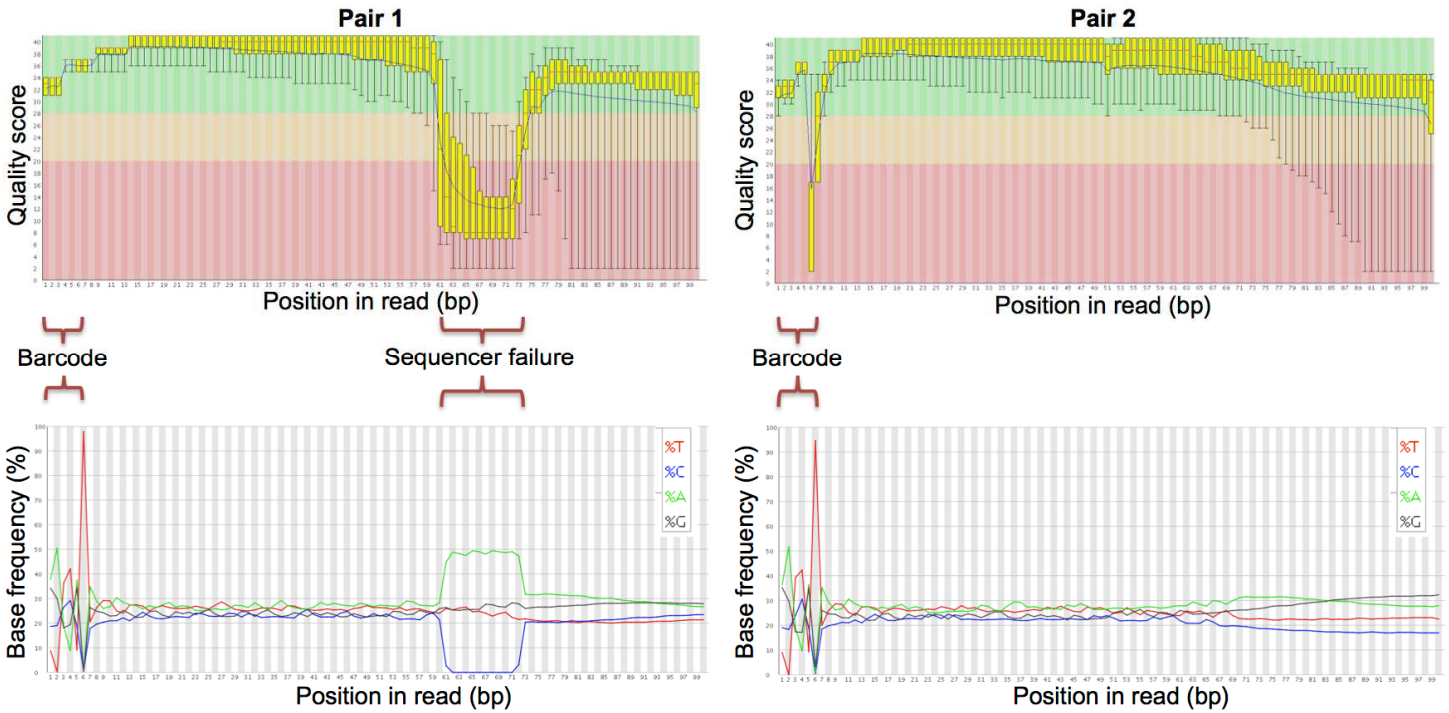


Figure 4-2: Quality scores and base frequency per position using the data from a single lane containing 5 samples.

FASTQC software was used to assess the quality of the sequencing data. Here, the lower quality scores at the 3' end of the sequences and the low quality score at the 6th base of pair 2 where the T is present (last base of the barcode used) are illustrated. The considerable dip in quality score during sequencing failure associated with loss of air-conditioning during 60th-75th cycles is shown (low quality scores in these cycles). The base frequency at a given cycle is expected to reflect the base frequency in the input sample. This holds when the quality score is high, but at low quality we can see a marked deviation in base frequency.

After adopting the above procedure, quality control was performed for each sample separately to reassure good quality data per sample. A high frequency of replicated sequences (>25% is assigned as high by the FASTQC software) was observed, an issue I therefore investigated.

4.4.5 PCR replicates

Sequencing data from eight out of 75 samples (11%) contained greater than 25% replicated sequences. Generally, more PCR cycles are used to produce enough library yield from low DNA input. PCR replicates arise when multiple copies of the same original molecule bind onto different positions on a flow cell. Therefore, it was considered that removal of PCR replicates should be performed from all samples using Samtools (Li *et al.*, 2009) as described in section 2.10. The median proportion of replicated sequences removed was 9.90% (range, 5.24-63.69%). The frequency of PCR replicates was inversely correlated with DNA input (Spearman's rho=-0.34, p=0.003) and positively correlated with the age of the block (Spearman's rho=0.33, p=0.004) as illustrated in Figure 4-3 and Figure 4-4 respectively and therefore very much was related to sample quality.

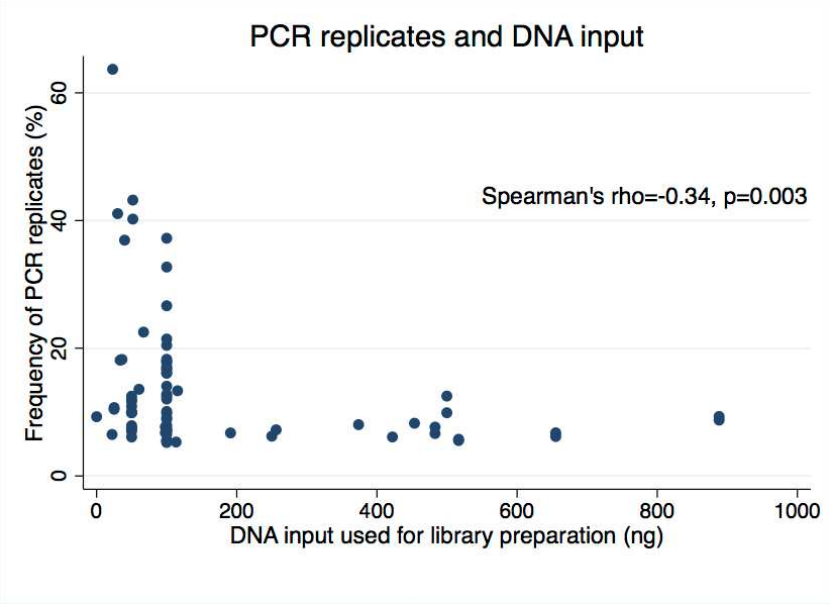


Figure 4-3: The correlation between frequency of PCR replicates and DNA input used for library preparation.

Replication was seen predominantly where DNA input was less than 200ng.

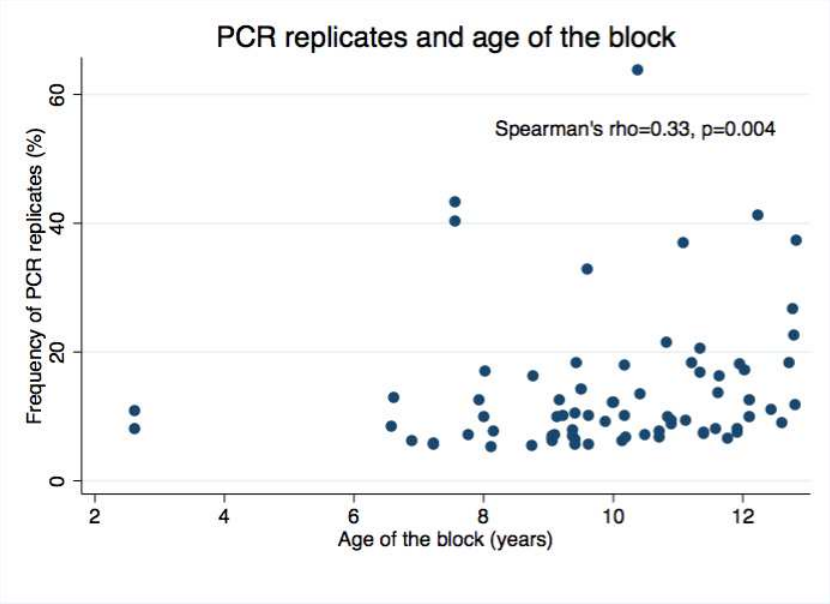


Figure 4-4: The correlation between PCR replicate frequency and age of the block.

Replicates were significantly associated with increased block age but 2 samples which were relatively recent (under 8 years) had high levels of PCR replicates.

Neither melanin score nor day of the week that the block was processed was associated with the frequency of PCR replicates (Spearman's rho=-0.08, p=0.48 for mitotic rate and x^2=7.21, p=0.13 for day of the week, data not shown). To investigate further what was removed using Samtools, the percentage of the number of PCR replicate sequence reads removed in each chromosome was plotted. The plot showed that similar numbers of replicate reads were removed across all chromosomes for each sample. There were significantly more PCR replicates seen however for chromosome 16 (Figure 4-5). Approximately 5% more reads were consistently removed from chromosome 16 compared to all other chromosomes in all samples (Figure 4-5). Work reported in section 4.4.6 suggests that this was because of "problematic" regions in this chromosome.

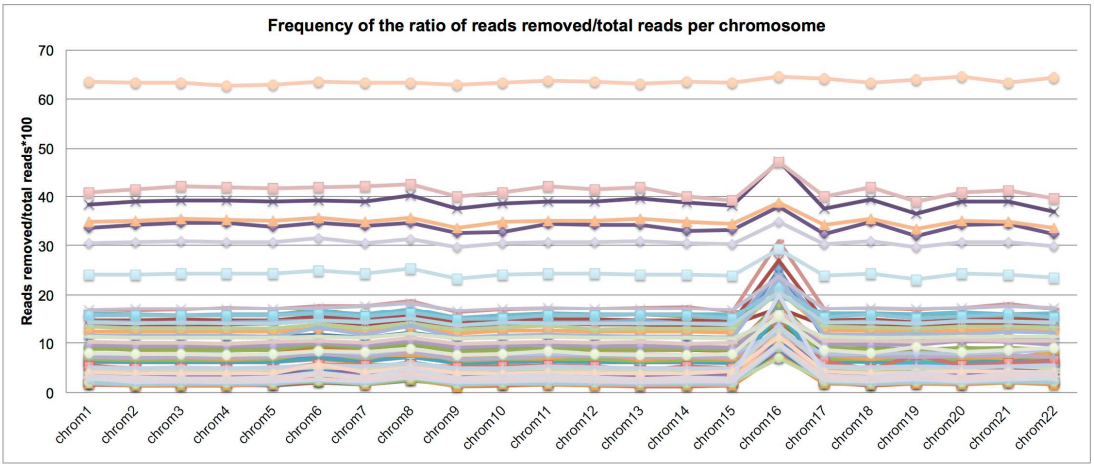


Figure 4-5: The frequency of the fraction of PCR replicate reads removed per chromosome for each sample.

It is illustrated that more reads aligned at chromosome 16 were consistently removed from all samples. Work reported in section 4.4.6 suggests that this was because of "problematic" regions in this chromosome.

4.4.6 Problematic regions

Raw data before and after removal of PCR replicates (dereplicated data) were used to assess the effect of regions reported by others as "problematic" (Pickrell *et al.*, 2011).

In 2011, Pickrell *et al* observed that errors in the human reference genome sequence cause copy number false-positive peaks in the ENCODE (Encyclopaedia of DNA Elements) NGS datasets (Pickrell *et al.*, 2011). The errors in the reference sequence are caused by collapsed repeats where a sequence in the reference is present in a single copy while in reality there are multiple copies of this sequence (Pickrell *et al.*, 2011). They observed a high read depth in these regions. The location of these regions is publicly available and different thresholds are used to define them as problematic (based on the distribution of sequencing coverage in the 1000 Genomes data, 0.1% right tail, 0.05% right tail, 0.01% right tail, 0.005% right tail, and 0.001% right tail) (Pickrell *et al.*, 2011). I therefore investigated how these problematic regions affect my data. The read counts of windows which belong to non-problematic and problematic regions based on the different thresholds are shown in Figure 4-6 for raw data and in Figure 4-7 for dereplicated data. Generally, increased numbers of reads were seen in the problematic regions.

There is no *de novo* rationale for choosing a particular threshold for data cleaning. However, we noted that the 0.1% threshold seem to reject too many regions and a decision to remove the problematic regions from subsequent analysis based on the 0.05% threshold was made.

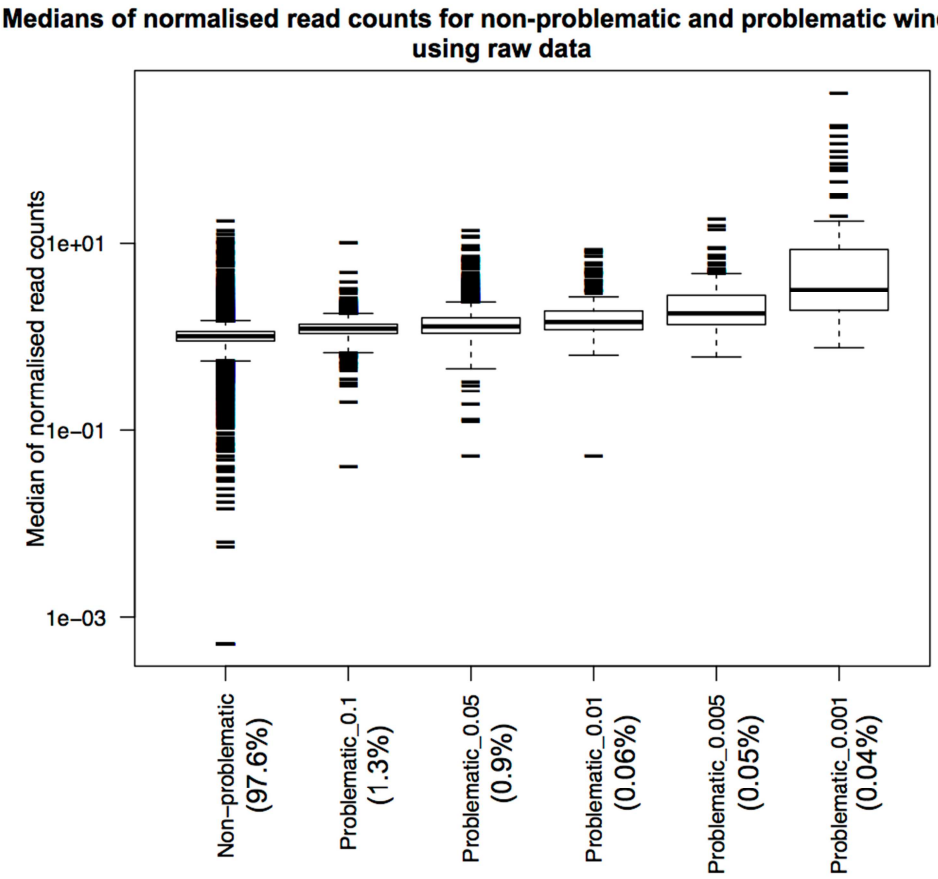


Figure 4-6: The medians of normalised read counts for non-problematic and problematic regions using raw data (before removal of PCR replicates).

Increased coverage was observed in the problematic regions i.e. that coverage was proportionally greater for the regions identified by Pickrell *et al* using more stringent criteria. The percentages in brackets represent the number of windows in our data that fall in each category. The most stringent threshold was the 0.001% right tail of the distribution of coverage in the data analysed by Pickrell *et al* (more details in the text).

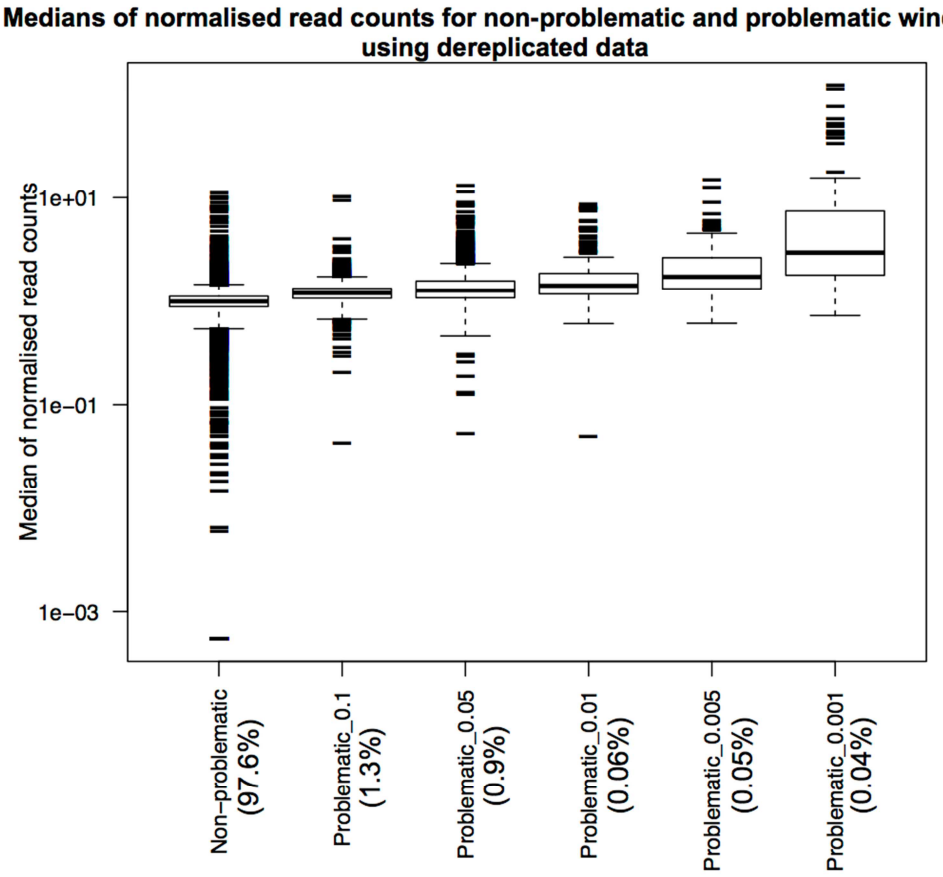


Figure 4-7: The medians of normalised read counts for non-problematic and problematic regions using dereplicated data.

Again in the dereplicated data, increased coverage was observed in the problematic regions: that coverage was proportionally greater for the regions identified by Pickrell *et al* using stringent criteria. The most stringent threshold was the 0.001% right tail of the distribution of coverage in the data analysed by Pickrell *et al* (more details in the text). The percentages in brackets represent the number of windows in our data that fall in each category. A decision to mask the problematic regions based on the 0.05% threshold was made.

The percentage of the number of PCR replicate reads removed in each chromosome was plotted again using the data after removal of problematic regions, Figure 4-8. For each sample the frequency of the reads removed was equal across all chromosomes suggesting that the consistent increased percentage of reads removed from chromosome 16 across samples (as shown in Figure 4-5) was due to one or more problematic regions located at that chromosome.

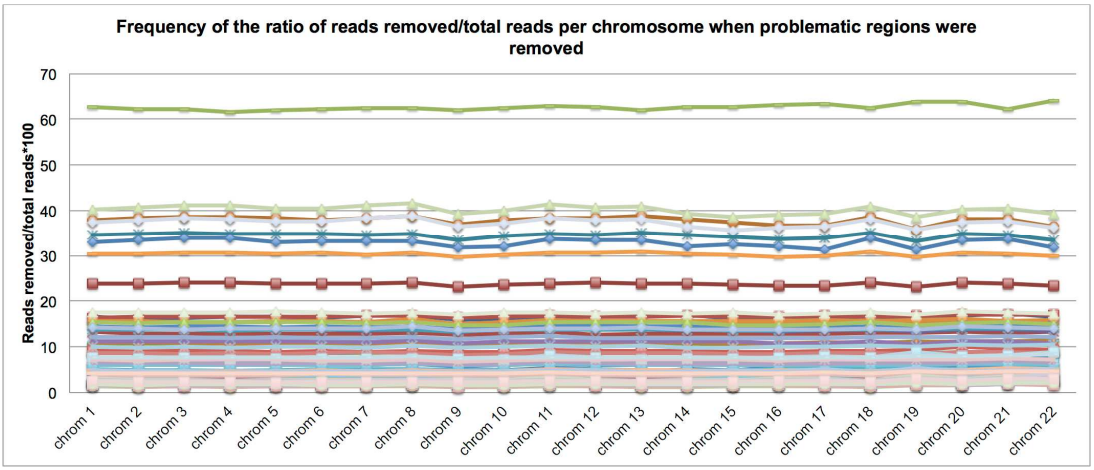


Figure 4-8: The frequency of the fraction of PCR replicate reads removed per chromosome after removal of problematic regions.

The increased percentage of reads aligned to chromosome 16 that were removed was due to false positive regions defined as problematic by Pickrell *et al*.

4.4.7 Coverage and tumour characteristics

The median number of reads after removal of PCR replicates and problematic regions for chromosomes 1-22 was 40,497,890 (range, 13,064,103-58,284,709). The median coverage was 1.01x (range, 0.33x-1.46x) if 80bp is considered the average read length (as the first and the last 10 bases were consistently removed from all samples).

Total read counts were associated with the DNA yield used for library preparation (Spearman's rho=0.29, p=0.01, Table 4-4). Neither melanin score, age of the block or day of the week that the block was processed were associated with total read counts (Table 4-4).

Table 4-4: Associations between total read counts and tumour characteristics

	Samples	Median read counts (range) ($x10^7$)	Association between measure and total read count
DNA input (ng)	75	4.05 (1.31-5.83)	**Spearman's rho=0.29, p=0.01**
Melanin score			
0	23 (30.67)	4.35 (1.31-5.83)	
1	15 (20.00)	4.10 (2.02-5.64)	Spearman's rho=-0.16, p=0.17
2	20 (26.67)	3.66 (2.14-5.68)	
3	17 (22.67)	3.82 (2.02-5.13)	
Age of block, years, median (range)	75		Spearman's rho=-0.0004, p=0.99
Day of the week block processed, n (%)	75		
Monday	14 (18.67)	3.26 (2.02-4.83)	
Tuesday	18 (24.00)	4.33 (2.02-5.59)	Kruskal-Wallis, χ^2=9.24, p=0.06
Wednesday	15 (20.00)	4.15 (1.31-5.20)	
Thursday	8 (10.67	3.62 (2.49-5.83)	
Friday	20 (26.67)	4.34 (2.14-5.64)	

4.4.8 Technical replicate samples

Spearman's correlation coefficients between read counts for each sample pair were calculated. We distinguished between pairs that were replicates of the same sample and pairs of unrelated samples (non-replicates). If the data are of high quality a high correlation between replicates compared to non-replicates is expected.

Table 4-5 shows that the correlation coefficients between non-replicates were much lower for non-problematic regions than for the problematic regions. This observation provides further evidence that the previously identified problematic regions by Pickrell *et al* are indeed questionable and data generated from them should probably be removed. There was no difference in correlations between raw and data from which PCR replicates had been removed (dereplicated data) (Table 4-5).

Table 4-5: Spearman's correlation coefficients between read counts per 10kb region for technical replicate and non-replicate pairs.

The median and the range of Spearman's rho for replicate pairs and non-replicate samples using raw and dereplicated data for the problematic and non-problematic regions is presented.

Raw data, Spearman's correlation coefficient median (range)			
Non-problematic		**Problematic**	
Replicates	**Non-replicates**	**Replicates**	**Non-replicates**
0.89 (0.80-0.92)	0.42 (-0.14-0.92)	0.93 (0.88-0.96)	0.64 (0.21-0.96)

Dereplicated data, Spearman's correlation coefficient median (range)			
Non-problematic		**Problematic**	
Replicates	**Non-replicates**	**Replicates**	**Non-replicates**
0.89 (0.80-0.94)	0.44 (-0.15-0.94)	0.93 (0.88-0.97)	0.65 (0.24-0.97)

Technical replicates were sequenced to assess the reproducibility of NGS technology using FFPE-derived DNA. Here, different technical replicates were assessed: a) the same libraries derived from 4 patient samples; b) 5 libraries which were generated using different DNA input from 2 patients (3 libraries using 250ng, 100ng and 25ng DNA from patient 1; 2 libraries using 100ng and 25ng DNA from patient 2); and c) 10 libraries derived from 5 patients of which 1 library was prepared using 15 PCR cycles and 1 using 18 cycles. The correlation coefficients between read counts in 10kb regions for each group of replicates were higher compared to non-replicate pairs and no difference was observed when different DNA input or different PCR cycles were used (Table 4-6). An example of chromosome 7 read counts is presented when the starting material was 250ng, 100ng and 25ng from the same patient (Figure 4-9).

Table 4-6: Spearman's correlation coefficients for the read counts per 10kb region of the 3 different groups of technical replicates.

Data free from problematic regions were used.

	Raw data Spearman's correlation coefficient, median (range)	**Dereplicated data** Spearman's correlation coefficient, median (range)
Same libraries	0.84 (0.80-0.94)	0.83 (0.81-0.91)
Libraries using 15 vs 18 PCR cycles	0.88 (0.81-0.92)	0.88 (0.81-0.92)
Libraries using different DNA input	0.90 (0.90-0.90)	0.90 (0.89-0.90)
Non-replicates	0.42 (-0.14-0.92)	0.44 (-0.15-0.94)

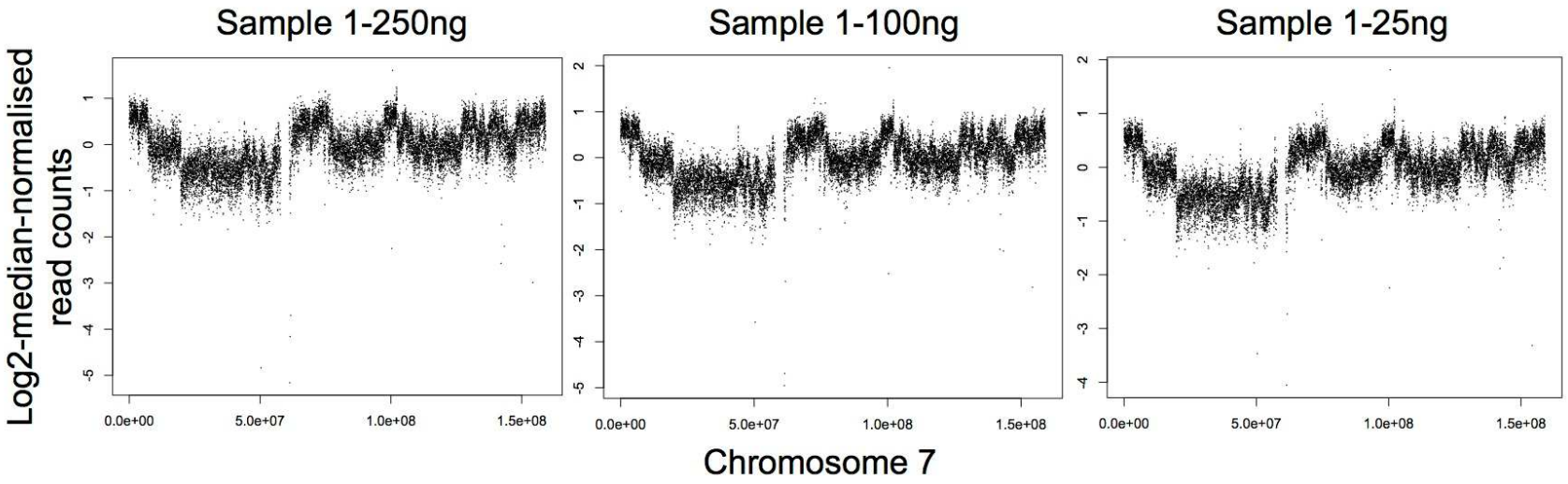


Figure 4-9: Examples of copy number analysis in chromosome 7 using different amounts of DNA from the same sample.

NGS data were generated using 250ng, 100ng and 25ng from the same sample. The log base 2 median normalised read counts for chromosome 7 are presented. Each dot represents a 10kb region. Similar patterns of copy number alterations are observed. Similar results were obtained for the other chromosomes (not shown).

4.4.9 GC correction

The data free from PCR replicates and problematic regions were used to identify if GC content affects the data. The explained variance by GC content identified using ANOVA, as described in section 2.10, varied across samples. A comparison between GC content and median-normalised read counts (GC bias) is illustrated in Figure 4-10 where three examples are shown. Both positive and negative linear associations but also non-linear associations between GC content and read counts were observed. Gusnanto *et al* have reported the use of a Loess model to correct for GC bias which does not allow for additional covariates to be incorporated into the model (Gusnanto *et al*., 2012). Although some of our data showed a non-linear GC bias, a linear model was used to correct for the observed GC bias, which also included chromosome as a covariate. Three examples of chromosome 7 read counts before and after GC correction are shown in Figure 4-11.

All the different associations observed in my data have been previously reported (Dohm *et al*., 2008, Chen *et al*., 2013) and therefore a correction was considered essential to remove the effect of GC content before downstream analysis.

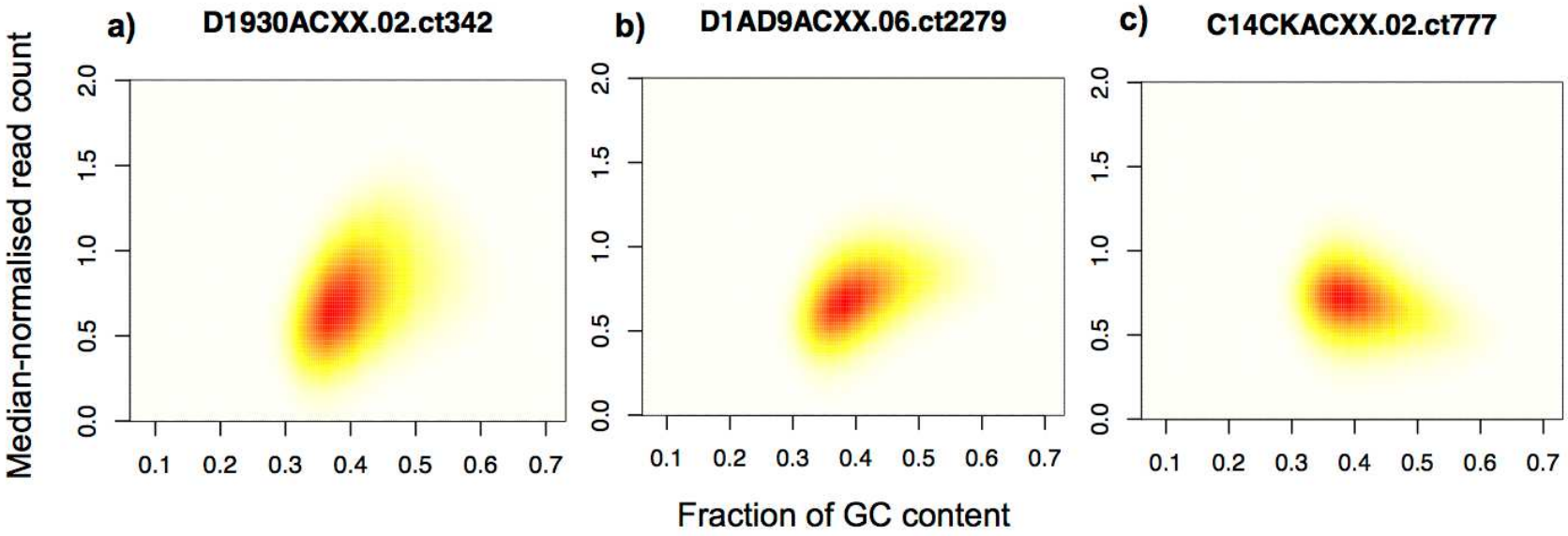


Figure 4-10: Comparison between GC content and median-normalised read counts.

Smoothed scatterplots between GC content and median-normalised read count per window, where red represents the highest density of windows (Eilers and Goeman, 2004). a) A positive linear association between GC content and read counts; b) a positive non-linear association between GC content and read counts; c) a negative linear association between GC content and read counts. All of these associations have been previously reported and a correction model should be applied to remove the effect of GC content before downstream analysis. A linear model was used to correct for the GC-bias which also included chromosome as a covariate.

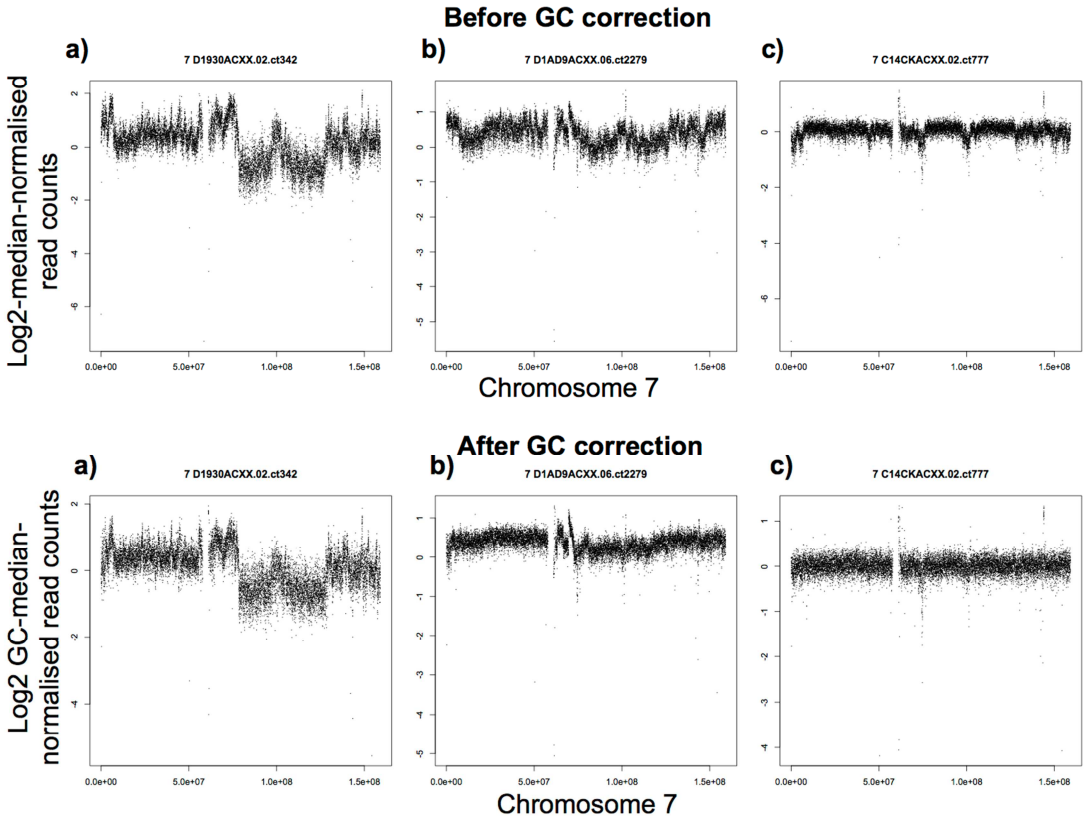


Figure 4-11: Chromosome 7 read counts before and after GC correction.

The data are shown for samples with: a) positive linear GC bias; b) positive non-linear GC bias; and c) negative linear GC bias. It can be seen that some variation is smoothed after correction for GC content. For the first sample (a) some variation in read counts is preserved which might be interpreted as that this is real variation, and for the other samples (b and c) a lot of variation is removed.

4.4.10 Genome wide analysis for associations between copy number and patients characteristics

The optimal approach to whole-genome NGS data analysis is still an active research area. Most methods utilise matched normal reference samples (which we did not have) and they have not been tried for FFPE material. We wanted to stay as close to the raw data as possible: that is that we did not want to adopt approaches to analysis which might be reported in the literature but flawed. In the absence of a standard metric for quality we tried to correct for the biases we observed but also exclude low quality data from subsequent analysis.

For the development of this protocol, where there were replicates data from a single replicate were used. To look for associations between copy number and patients' characteristics median and GC-corrected data free from both PCR replicates and problematic regions, from 61 unique samples were used (descriptive statistics shown in Table 4-7).

Breslow thickness, mitotic rate, ulceration, AJCC stage, melanoma-specific survival (MSS), age of the patient, sex, tumour site and primary type were tested for association with the read counts from each 10kb window.

Table 4-8 shows the number of windows that had an FDR<0.1. Fifty-one windows were significantly associated with age of the patient, 1,736 windows were significantly associated with Breslow thickness, 21,424 windows with mitotic rate, 5,909 with *BRAF* mutated tumours compared to non-*BRAF* mutated ones but only 1 window was associated with AJCC stage. The histograms of the p values for the associations between each 10kb window and mitotic rate, MSS and ulceration are shown in Figure 4-12. A flat histogram of p values is expected, as most windows will show no significant association with the tested variable, with a high peak at low p values for the significant associations. After regression analysis for MSS, an increased frequency of low p values was observed for some windows but these failed to reach statistical significance after multiple testing correction (Figure 4-12). On the contrary, after regression analysis for ulceration, no increased frequency of low p values was observed (Figure 4-12).

Table 4-7: Descriptive statistics of the sample set (n=61).

Variable	Sample set
Age at diagnosis (years), median (range)	59.67 (25.09-76.34)
Sex, n (%)	61
Male	25 (40.98)
Female	36 (59.02)
Breslow thickness, mm, median (range)	2.3 (1.15-12)
Mitotic rate, per $mm^{2,}$ n (%)	54 (88.52)
Mitotic rate, per mm^2, median (range)	3 (0-70)
AJCC stage, n (%)	61
I	21 (36.84)
II	33 (54.10)
III	7 (11.48)
Ulcerated tumours	22 (36.06)
Primary type, n (%)	61
SSM/LMM	33 (54.10)
NM	24 (39.34)
ALM/Unclassified	4 (6.56)
Tumour site	61
Head/Neck/Trunk	21 (34.43)
Limbs/Acral	40 (65.57)
Deaths, n (%)	23 (37.70%)
Deaths from melanoma, n (%)	20 (32.79%)
Follow-up period, years, median (range)	8.03 (0.89-11.76)
Mutation status, n (%)	60
BRAF	23 (38.33)
NRAS	21 (35.00)
None	16 (26.67)

Table 4-8: Associations between read counts in each window and tumour characteristics (genome-wide analysis).

It can be seen that the number of associated windows was very variable between clinical features being tested.

Variable	Statistical model and number of windows with FDR <0.1
Age (years)	Linear regression, 51
Gender (Male/Female)	Logistic regression, 0
Breslow thickness, mm	Linear regression, 1736
Mitotic rate, per mm2	Linear regression, 21424
AJCC stage (I/II/III)	Linear regression, 1
Ulceration (yes, no)	Logistic regression, 0
Primary type (SSM/LMM, NMM)	Logistic regression, 0
Tumour site (Head/Neck/Trunk, Limbs)	Logistic regression, 0
Melanoma-specific survival	Cox regression, 0
Mutation status (*BRAF*/*NRAS*)	
BRAF vs non-*BRAF*	Logistic regression, 5909
NRAS vs non-*NRAS*	Logistic regression, 0
BRAF vs no mutation in both	Logistic regression, 0
NRAS vs no mutation in both	Logistic regression, 0
BRAF vs *NRAS*	Logistic regression, 0

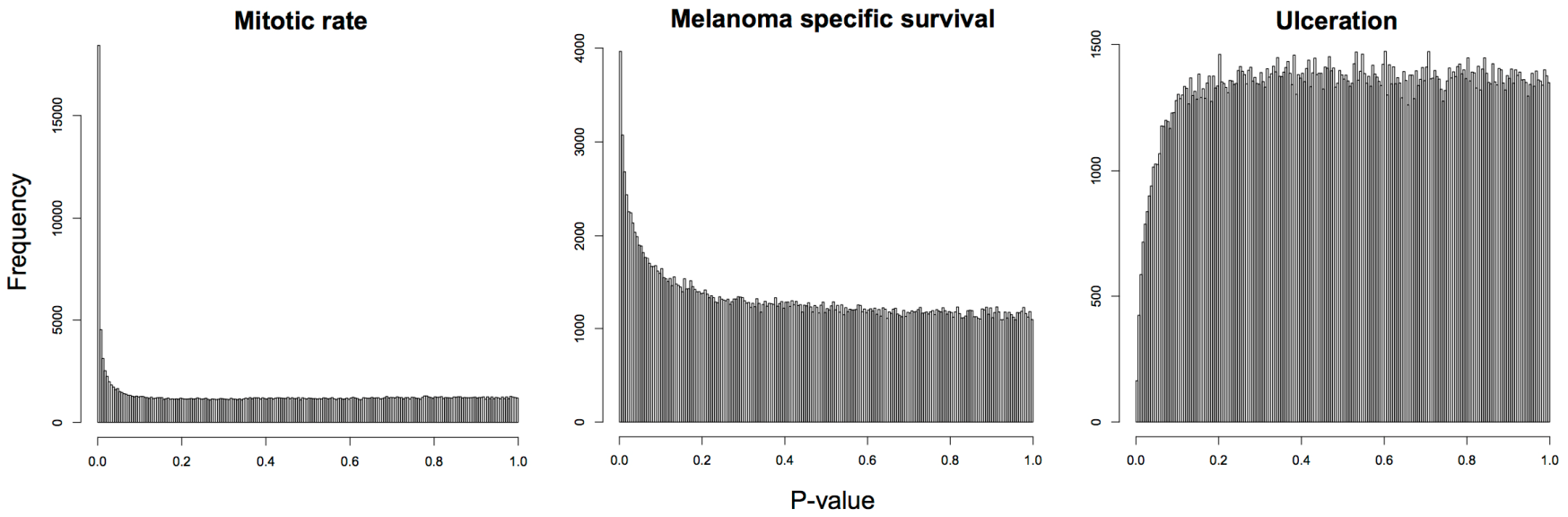


Figure 4-12: Histograms of the frequency of p values for the associations between each 10kb window and mitotic rate, melanoma-specific survival (MSS) and ulceration.

A flat histogram of p values is expected, as most windows will show no significant association with the tested variable, with a high peak at low p values for the significant associations as is seen above for associations with mitotic rate. After regression analysis for MSS, an increased frequency of low p values was observed for some windows but these failed to reach statistical significance after multiple testing correction. On the contrary, after regression analysis for ulceration, no increased frequency of low p values was observed. The histogram of p values for mitotic rate is used for comparison and shows the high number of windows that reached statistical significance.

The FDR values for each window's association with Breslow thickness, mitotic rate and *BRAF* mutated tumours are shown in Figure 4-13, Figure 4-14 and Figure 4-15 respectively. These graphs show that only individual windows showed a statistically significant association with Breslow thickness while windows adjacent to each other showed a statistically significant association with mitotic rate and *BRAF* mutation status. For example, read counts from the start of the chromosome 5p arm, chromosome 7q and 8q arms were significantly associated with mitotic rate (Figure 4-14). It was also observed that read counts from chromosome 10q arm were significantly associated with *BRAF* mutated tumours (Figure 4-15). As the loss of chromosome arm 10q has been previously associated with positive *BRAF* mutation status (Jonsson *et al*., 2007, Gast *et al*., 2010), I investigated whether I could identify if this observation represented a copy number loss or a gain by plotting the read counts from chromosome 10q arm in *BRAF* mutated tumours and non-*BRAF* mutated tumours (Figure 4-16). On the log scale read counts below zero represent copy number loss while read counts greater than zero represent copy number gain. Read counts below zero were seen in *BRAF* mutated tumours whereas read counts close to zero were observed for non-*BRAF* samples, showing that 10q arm loss was associated with *BRAF* mutated melanomas (Figure 4-16) as expected from the literature.

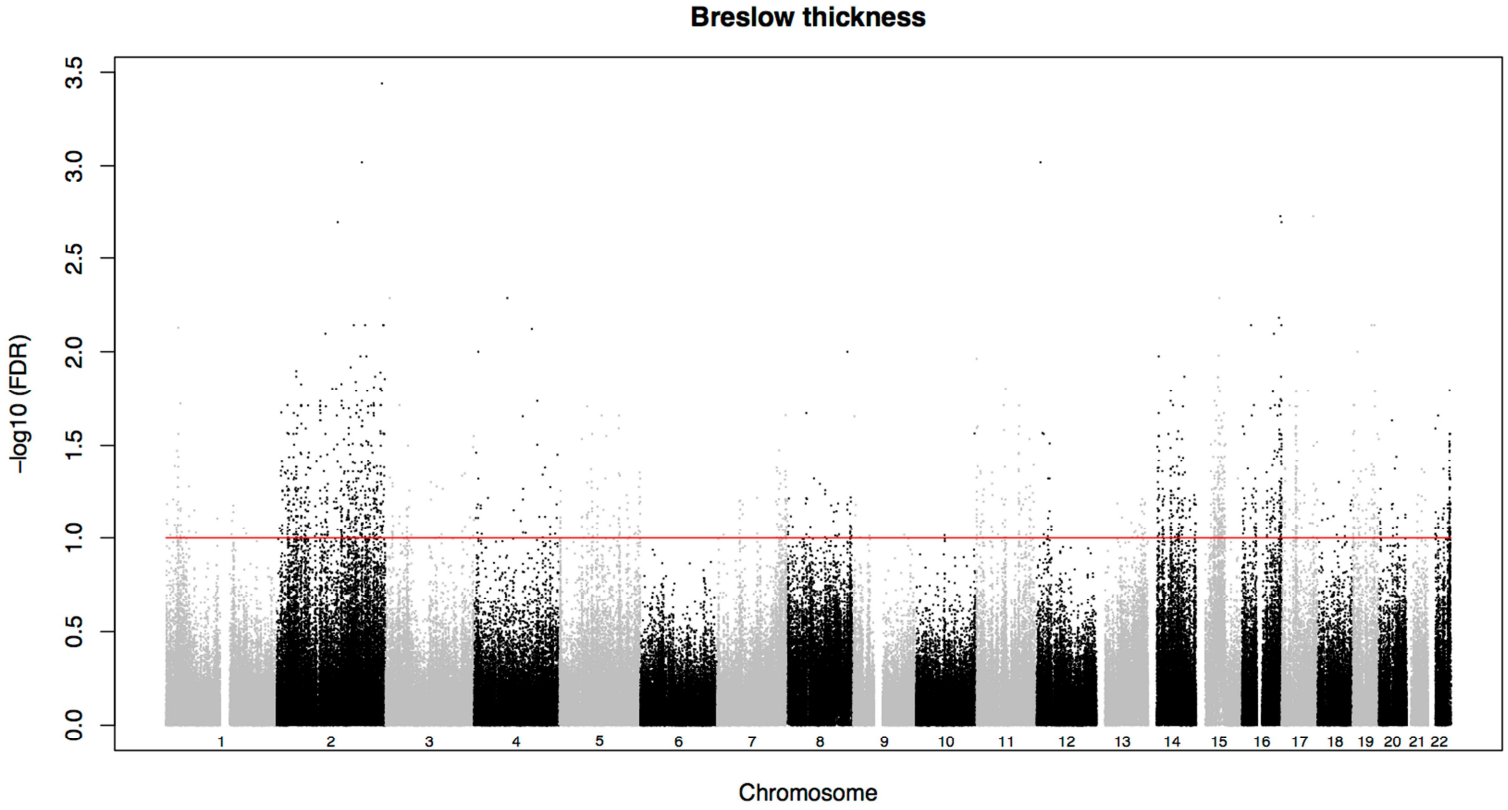


Figure 4-13: The statistical significance values (-$\log_{10}$ FDR) across the 22 chromosomes of each 10kb window's association with Breslow thickness.

The red line represents the FDR threshold. Mainly, individual windows and small regions such as in chromosome 1p were significantly associated with Breslow thickness. It is noted that very small regions in chromosome 2 are significantly associated with Breslow thickness.

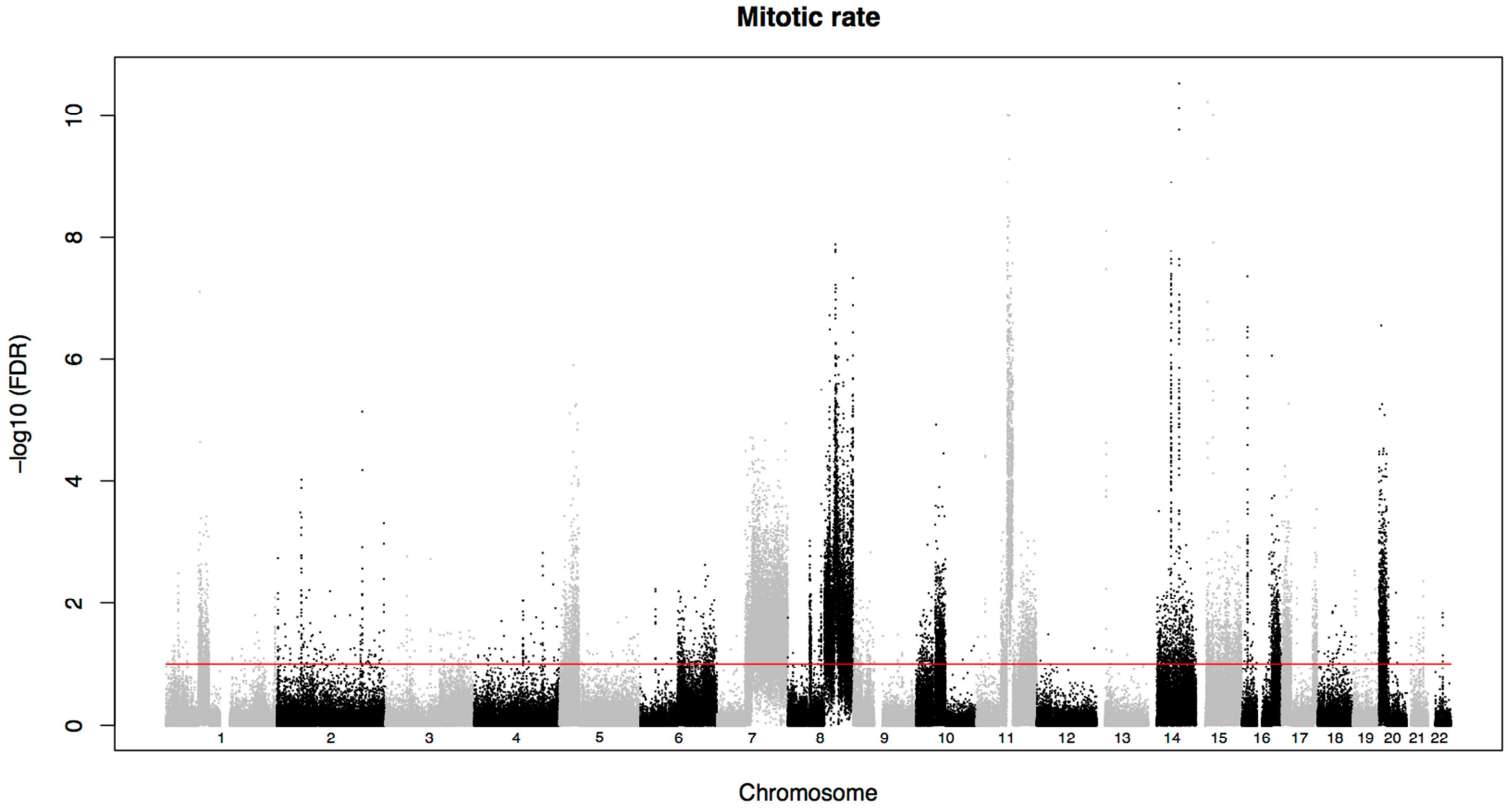


Figure 4-14: The statistical significance values (-$\log_{10}$ FDR) across the 22 chromosomes of each window's association with mitotic rate.

The red line represents the FDR threshold (0.1). Mitotic rate showed the most significant associations with copy number. A strong signal was observed in relatively small regions such as in chromosome arm 1p, 5p, and 11p but also in whole chromosome arms (7q, 8q, 20p).

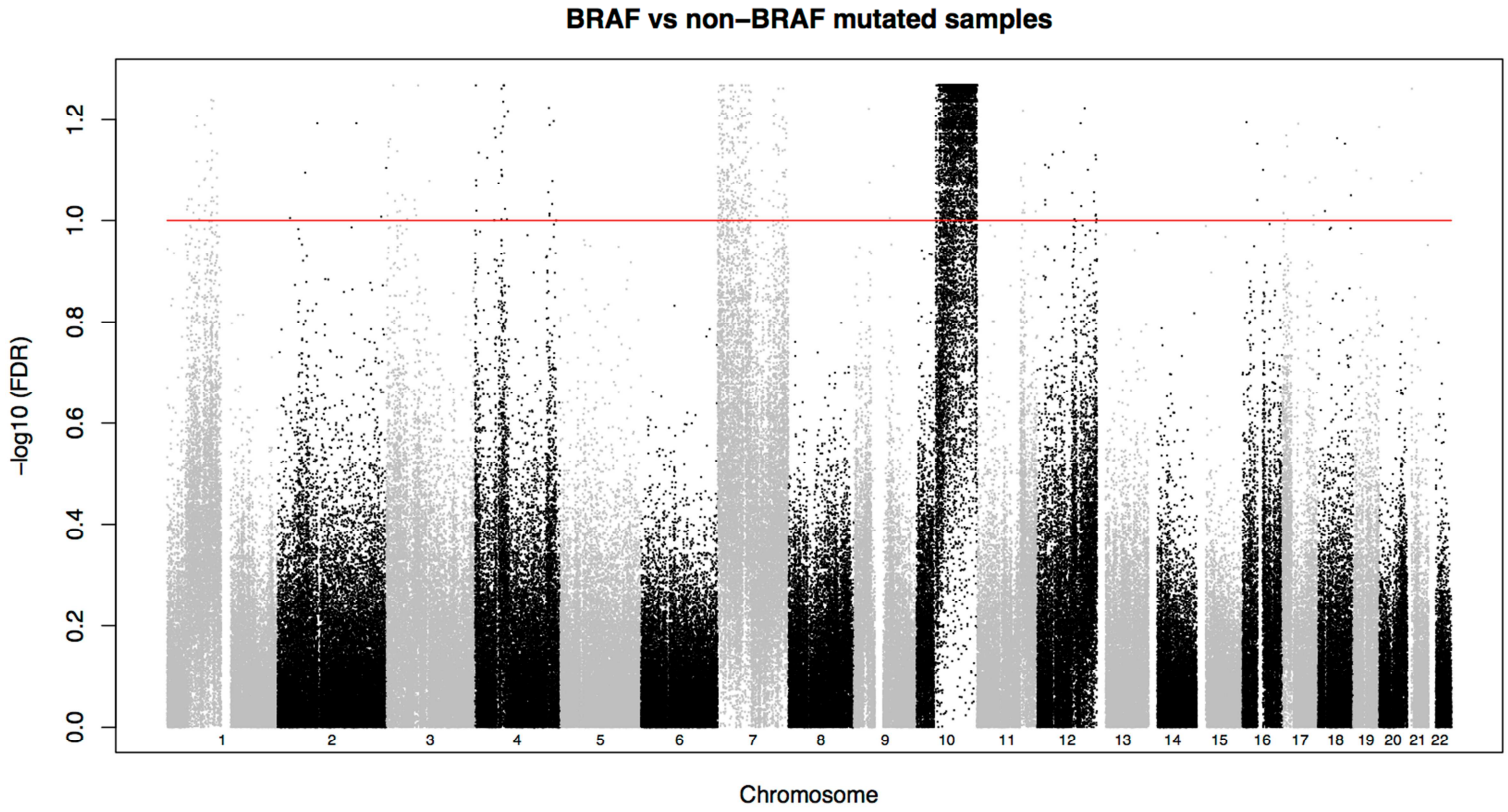


Figure 4-15: The statistical significant values (-$\log_{10}$ FDR) across the 22 chromosomes of each 10kb window's association with *BRAF* mutated samples.

The red line represents the FDR threshold of 0.1. It can be seen that a strong signal is from chromosome arm 10q where the tumour suppressor *PTEN* is coded, a gene which is known to be deleted in *BRAF* mutated melanomas.

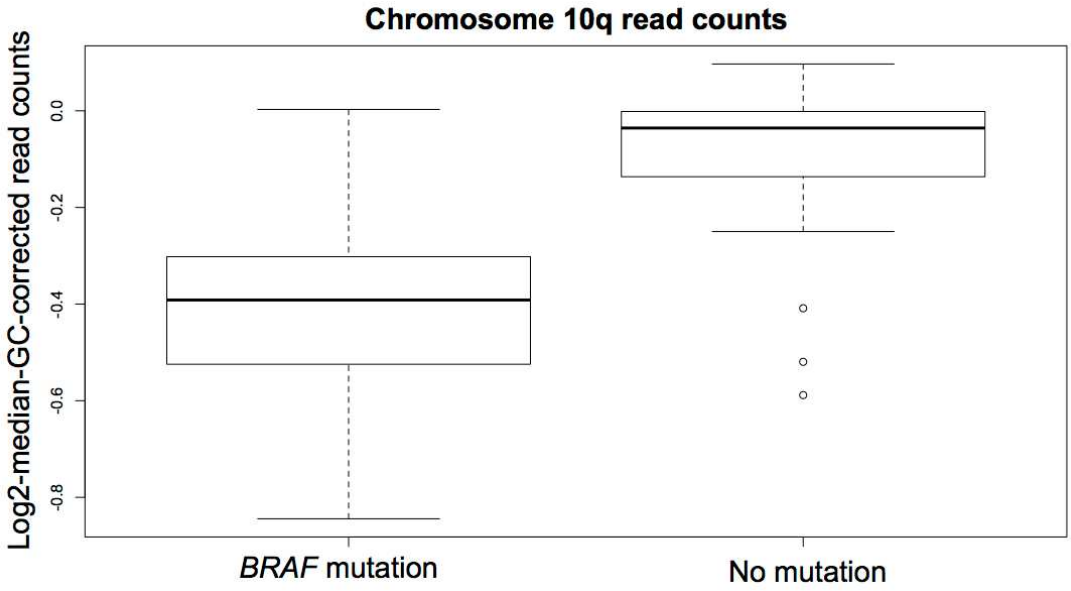


Figure 4-16: Chromosome 10q read counts in samples with *BRAF* mutation versus samples with no *BRAF* mutation.

On the log scale read counts <0 represent copy number loss while read counts >0 represent copy number gain. It is shown *BRAF* mutation was associated with loss of chromosome arm 10q, as has previously been reported.

To reduce noise, the coefficients from the regression models were smoothed using a bandwidth of 501 windows. The advantage of smoothing the coefficients instead of p-values was that the direction of the association could be identified. For example, for mitotic rate a positive coefficient for a specific region can be interpreted as higher copy number associated with higher mitotic rate or lower copy number is associated with lower mitotic rate for that region (Figure 4-17). Positive associations between copy number status of whole chromosome arms such as 7q, 8q and mitotic rate were identified. Negative associations between copy number status of whole chromosome arms such as 17p and 20p and mitotic rate were also found.

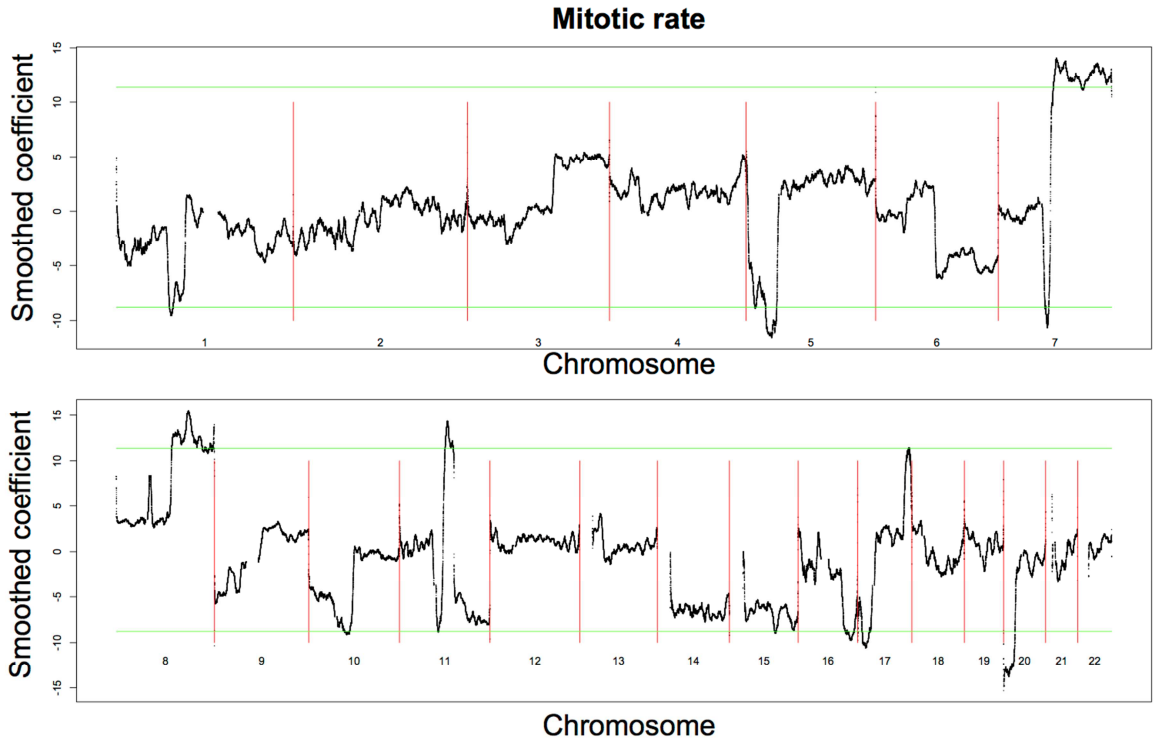


Figure 4-17: The smoothed coefficients (based on a 501 windows running median) from the regression model for mitotic rate.

The green lines represent the FDR threshold of 0.1. Positive associations between copy number status of whole chromosome arms such as 7q, 8q and mitotic rate were identified. Negative associations between copy number status of whole chromosome arms such as 17p and 20p and mitotic rate were also found. The effect of small regions is possibly masked due to smoothing of the coefficients using a bandwidth of 501 windows.

To identify if these chromosome arms 7q and 8q were gained or lost in association with mitotic rate the read counts of these chromosome arms for patients grouped based on mitotic rate tertiles (0-1, 1-5, >5) were plotted (Figure 4-18). Higher copy number in 8q was observed in patients with higher mitotic rate. This was not the case for 7q where similar read counts were observed in all patients except for a single sample which had the highest mitotic rate (70mm^2) and showed the highest copy number. Linear regression models rely on the assumption that the residuals are normally distributed. The sample with very high mitotic rate and increased copy number probably affected the results where 7q appeared to be significantly associated with mitotic rate (as shown in Figure 4-14 and Figure 4-17). This suggests that a more robust regression model which will not assume normally distributed data should be performed in the future to identify clear associations.

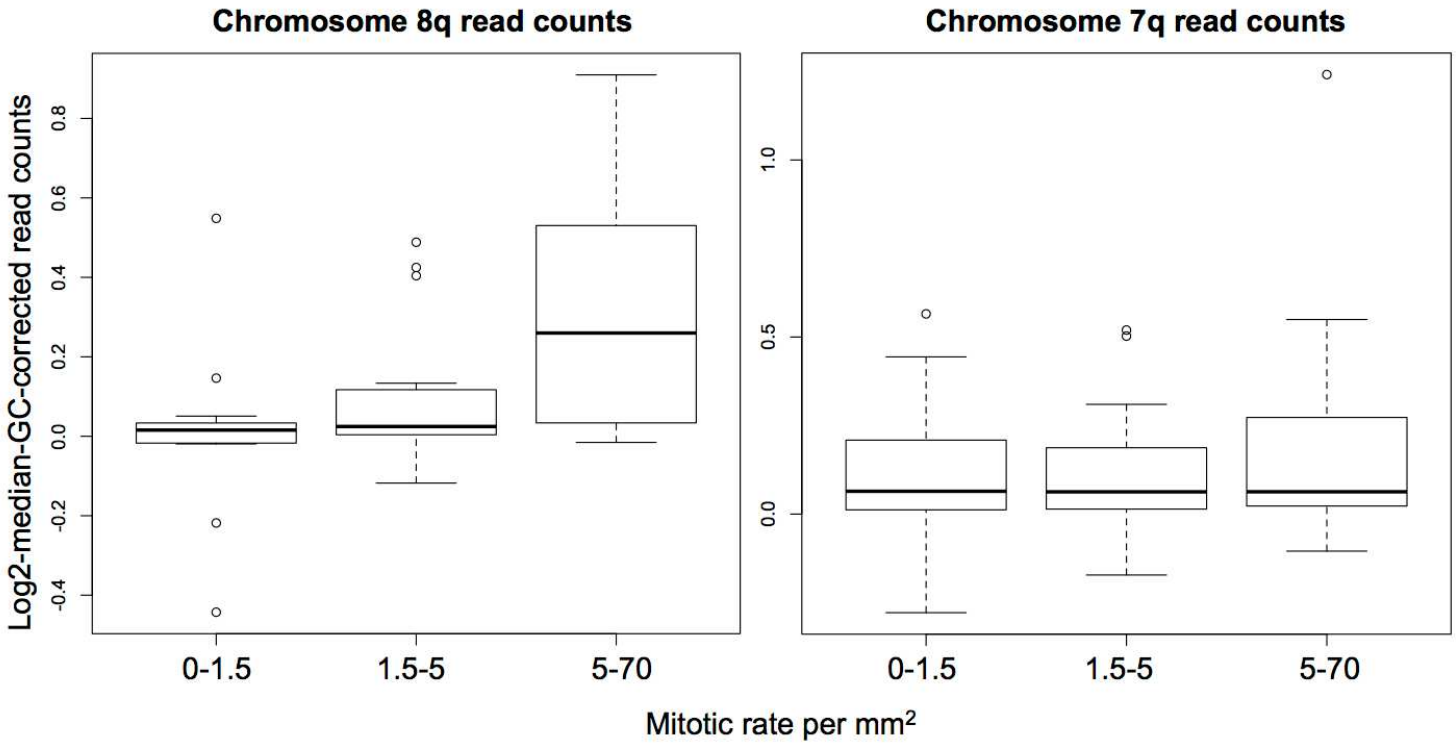


Figure 4-18: Chromosome 8q and 7q read counts in patients with different mitotic rate.

Patients were grouped in tertiles (with mitotic rate 0-1.5, 1,5-5, 5-70). A) Higher copy number in chromosome 8q was seen in patients with higher mitotic rate. B) Similar copy number in chromosome 7q was seen in all patients except for a single patient which has a mitotic rate of 70mm^2. Linear regression models rely on the assumption that the residuals are normally distributed. This observation suggests that a robust regression model should be used in the future to identify clear associations.

4.4.11 Copy number and gene expression data integration for the chromosomal regions found to be associated with mitotic rate

I report here a study of whether CNAs in regions shown to be associated with mitotic rate mirrored variation in expression levels of genes within those regions. Whole-genome DASL data were available from 47 out of 61 samples with copy number data. The windows significantly associated with mitotic rate based on the smoothed coefficients (see Figure 4-17) were merged with the DASL data based on the DASL probe coordinates as described in section 4.3.8. 1894 DASL probe/NGS window pairs were identified and Spearman's correlation coefficients between the signal of each DASL probe and the read count of the corresponding NGS 10kb window were calculated. Thirty-three significant positive and 3 negative correlations were identified. Twenty-two DASL probes whose expression was significantly correlated with copy number were located on chromosome 7, 12 on chromosome 8 and 1 on chromosome 16. The top 10 most significant correlations are listed in Table 4-9.

To reduce noise, the NGS 10kb window read counts were smoothed using a bandwidth of a 501 windows running median. Then, Spearman's correlation coefficients between the smoothed NGS read count and the matched DASL probe signal were calculated. This analysis revealed 159 significant positive and 5 negative correlations (FDR<0.1). Two probes whose signal was significantly correlated with copy number were located on chromosome 1, 4 on chromosome 5, 82 on chromosome 7, 39 on chromosome 8, 17 on chromosome 11, 1 on chromosome 15, 4 on chromosome 16, 6 on chromosome 17 and 9 on chromosome 20. The top 10 most significant correlations are listed in Table 4-10.

The 159 DASL probes (representing 131 genes) significantly correlated with the smoothed NGS read counts were analysed using MetaCore™ pathways analysis software in an attempt to identify enriched pathways. The top 3 significantly enriched pathways (IL-2 signalling, cell cycle and DNA damage pathways) are listed in Table 4-11.

Table 4-9: Top 10 significant correlations between the signal of a DASL probe and the read count of the corresponding NGS 10kb region.

For this analysis the NGS 10kb windows significantly associated with mitotic rate based on the smoothed coefficients were identified and matched to DASL probes (1894 DASL probe/NGS window pairs). The gene name that the DASL probe targets is presented. The smoothed coefficient from the regression model with mitotic rate is also presented to show the direction of the association with mitotic rate. The last column indicates whether the reported gene is located in the corresponding NGS 10kb region in the hg19.

Chr	Start	End	Gene	Spearman's rho	FDR	Smoothed coefficient	hg19
7	129560001	129570001	*KLHDC10*	0.56	0.07	18.10	×
16	74900001	74910001	*WDR59*	0.55	0.07	-12.71	✓
8	144970001	144980001	*PUF60*	0.53	0.09	17.98	×
7	130030001	130040001	*TSGA14*	0.53	0.09	18.10	×
7	149800001	149810001	*GIMAP8*	-0.52	0.09	17.69	×
7	123320001	123330001	*WASL*	0.51	0.09	17.24	✓
8	94810001	94820001	*RBM12B*	0.50	0.09	18.54	×
7	94020001	94030001	*CASD1*	0.49	0.09	19.85	×
7	100870001	100880001	*CLDN15*	0.49	0.09	17.61	✓
8	145060001	145070001	*GRINA*	0.49	0.09	18.04	✓

Table 4-10: Top 10 significant correlations between the signal of a DASL probe and the smoothed read count of the corresponding NGS 10kb region.

For this analysis the significant NGS 10kb windows associated with mitotic rate based on the smoothed coefficients were identified and matched to DASL probes. The read count of each NGS 10kb window was smoothed using a bandwidth of 501 windows running median (which results in 5Mb sliding windows). The gene name that the DASL probe targets is presented. The smoothed coefficient from the regression model with mitotic rate is also presented to show the direction of the association with mitotic rate. The last column indicates whether the reported gene is located in the corresponding NGS 5Mb region in the hg19.

Chr	Start	End	Gene	Spearman's rho	FDR	Smoothed coefficient	hg19
8	144470001	145480001	*PUF60*	0.67	0.001	17.99	✓
7	129530001	130540001	*TSGA14*	0.60	0.006	18.10	✓
11	77420001	78430001	*GAB2*	060	0.006	16.79	✓
7	111900001	112910001	*TMEM168*	0.61	0.006	17.72	✓
11	77090001	78100001	*INTS4*	0.60	0.006	16.59	✓
8	144560001	145570001	*GRINA*	0.58	0.006	18.04	✓
8	145010001	146020001	*DGAT1*	0.59	0.006	19.20	✓
7	86360001	87370001	*CROT*	0.58	0.007	18.50	✓
20	800001	1810001	*FKBP1A*	0.56	0.013	-19.85	✓
8	95820001	96830001	*C8orf37*	0.55	0.014	17.61	✓

Table 4-11: The top 3 overrepresented pathways identified using MetaCore™.

For this analysis the 159 DASL probes significantly associated with the smoothed read count of their corresponding NGS 10kb regions were used for pathway analysis.

Pathway	FDR	Genes (Chr arm)
Immune response_IL-2 activation and signalling pathway	0.069	*GAB2* (11q), CHUK (10q), SHPS-1 (20q)
Cell cycle_Role of 14-3-3 proteins in cell cycle regulation	0.096	*Tp53* (17p), *CDC25B* (20p)
DNA damage_ATM/ATR regulation of G2 / M checkpoint	0.096	*Tp53* (17p), *CDC25B* (20p)

4.4.12 Copy number and gene expression data for the chromosomal regions coding for genes known to be involved in melanoma biology: a candidate gene analysis

Here, I report the correlations between copy number and gene expression data of genes of interest. Fifteen genes which have previously been associated with melanoma in terms of gene expression (*RAD51*, *RAD52*, *RAD54B*, *SPP1*, *TOP2A*), mutation (*BRAF, MC1R, NRAS*, *KIT*, *TERT*) and copy number alteration (*CDKN2A*, *EGFR*, *MITF, MYC, PTEN*) were chosen. The correlations between the signal of DASL probes targeting each interesting gene and the NGS read counts of the region where each DASL probe is located were calculated (Table 4-12). Interestingly, there was a statistically significant correlation between copy number and gene expression measure for *CDKN2A, PTEN*, *MYC, RAD52 and TOP2A*. The median and the range of read counts for *CDKN2A* and *PTEN* indicate that this region was commonly lost. It was noted that *SPP1* expression was significantly negatively correlated with copy number status.

Table 4-12: Spearman correlation coefficients between DASL and NGS data for the candidate genes.

The Spearman's correlation coefficient between each DASL probe coding for a candidate gene and NGS read counts of the corresponding 10kb region where each probe is located was calculated. The gene name that the DASL probe targets is presented. The log base 2 GC corrected read count for each region and the range are also reported in an attempt to identify if the region is gained or lost. Counts >0 show that the region is gained. The significant associations ($p<0.05$) are highlighted in black. The rows highlighted in red represent regions where the reported gene is not in the corresponding region in the hg19.

Chr	Start	End	Gene	Spearman's rho	P value	Median read count (range)
3	69780001	69790001	*MITF*	-0.15	0.31	-0.29 (-1.45-1.42)
3	69810001	69820001	*MITF*	0.04	0.79	-0.01 (1.52-1.31)
3	70010001	70020001	*MITF*	-0.26	0.08	-0.11 (1.91-0.50)
4	**88900001**	**88910001**	***SPP1***	**-0.35**	**0.02**	**-0.08 (-0.79-0.37)**
4	**88900001**	**88910001**	***SPP1***	**-0.29**	**0.05**	**-0.08 (-0.79-0.37)**
7	55180001	55190001	*EGFR*	-0.17	0.26	0.12 (-0.52-0.74)
7	55190001	55200001	*EGFR*	-0.09	0.54	0.15 (-0.70-0.65)
7	55200001	55210001	*EGFR*	0.05	0.72	0.07 (-0.69-0.86)
7	55230001	55240001	*EGFR*	-0.19	0.20	0.32 (-0.71-1.13)
8	95460001	95470001	*RAD54B*	-0.08	0.58	0.11 (-0.57-1.11)
8	95480001	95490001	*RAD54B*	0.04	0.79	0.21 (-0.57-0.99)
8	**128820001**	**128830001**	***MYC***	**0.28**	**0.06**	**0.11 (-0.42-1.04)**
8	**128820001**	**128830001**	***MYC***	**0.33**	**0.02**	**0.11 (-0.42-1.04)**
9	**21960001**	**21970001**	***CDKN2A***	**0.44**	**0.002**	**-0.39 (-2.95-0.25)**
9	**21970001**	**21980001**	***CDKN2A***	**0.30**	**0.04**	**-0.54 (-4.37-0.21)**
9	21990001	22000001	*CDKN2A*	0.02	0.91	-0.74 (-4.82-0.28)
10	**89690001**	**89700001**	***PTEN***	**0.35**	**0.02**	**-0.34 (-3.10-0.38)**
12	**1020001**	**1030001**	***RAD52***	**0.35**	**0.01**	**0.02 (-0.48-0.87)**
17	**38540001**	**38550001**	***TOP2A***	**0.36**	**0.01**	**0.16 (-0.93-0.92)**
1	115050001	115060001	*NRAS*	-0.10	0.51	-0.13 (-1.01-0.74)
4	55300001	55310001	*KIT*	0.26	0.08	0.02 (-0.59-0.46)
4	55300001	55310001	*KIT*	0.04	0.77	0.02 (-0.59-0.46)
5	1300001	1310001	*TERT*	0.12	0.40	0.08 (-0.59-2.44)
5	1300001	1310001	*TERT*	-0.13	0.37	0.08 (-0.59-2.44)
5	1300001	1310001	*TERT*	0.12	0.40	0.08 (-0.59-2.44)
7	140080001	140090001	*BRAF*	0.21	0.16	0.15 (-0.91-1.21)
15	38790001	38800001	*RAD51*	-0.01	0.93	-0.02 (-1.23-0.33)
15	38810001	38820001	*RAD51*	0.05	0.74	-0.05 (-0.92-0.78)
15	38810001	38820001	*RAD51*	-0.11	0.46	-0.05 (-0.92-0.78)
16	88510001	88520001	*MC1R*	0.49	0.001	-0.21 (-1.12-0.60)

4.5 Discussion

In recent years, high-throughput technologies have provided researchers with useful tools to address diverse biological problems.

Next generation sequencing technology is driving advances in cancer research in that genomic, transcriptomic and epigenomic data can be generated to offer a more complete view of the cancer cell (Hawkins *et al*., 2010). In melanoma, whole-genome and whole-exome sequencing data have provided new insights in the pathogenesis of the disease (reviewed by (Hill *et al*., 2013)). For example in 2012, Hordis *et al* generated whole-exome data from 121 melanoma tumours and reported that the high prevalence of C>T transitions attributed to UVB exposure was linked to an increased rate of driver mutations (e.g. *RAC1*), providing evidence that UVB mutagenesis has a direct effect in the pathogenesis of melanoma (Hodis *et al*., 2012).

In a different study using already available whole-exome data, the authors identified a recurrently amplified region affecting *PHGDH (*phosphoglycerate degydrogenase - involved in cell metabolism) in melanoma samples showing further evidence that NGS data can provide new evidence to explore the importance of other than the currently known pathways related to melanoma development (Locasale *et al*., 2011).

Most NGS studies have used DNA extracted from cell lines or frozen tumours. Melanoma primaries are so small that cryopreservation of tissue is rarely possible: staging of the disease is primarily dependent upon histopathological features, so that freezing of even a small part of the primary is not possible lest that staging process be compromised. In order to use NGS to study sufficient numbers of tumours to give statistical power to address important clinical questions, then it is necessary to develop a protocol for use of formalin-fixed tumour. It is also clearly necessary to establish that data generated are of sufficient quality to give reproducible valid data.

Recent developments in protocols have made it possible to use low amounts of DNA to generate libraries for NGS. Studies have reported the use of NGS of material from FFPE tumours (Wood *et al*., 2010, Conway *et al*., 2012, Adams *et al*., 2012). In Chapter 3 I reported the successful library preparation using DNA from FFPE primary melanoma tumours, which enabled us to select patients from the Leeds Melanoma Cohort study to identify somatic CNAs in melanoma tumours moving towards the aim to study CNAs as potential prognostic biomarkers. This is

an on-going project of which the first results were presented here. These results are discussed as approaches to determining the validity of use of this technique for formalin fixed tumours.

4.5.1 DNA samples

The samples were selected for this study in a case control design where cases were patients who have died and controls were sex and Breslow thickness matched participants who have not died. Here, sequencing data from the first 75 samples are presented to test the feasibility of NGS technology using FFPE-derived DNA. Seventy-five samples is the largest number of FFPE primary melanoma tumours subjected to whole-genome NGS in the world. It is as yet a small proportion of the final number of samples which will be sequenced for this study and therefore the data reported here are as yet underpowered.

The Leeds Melanoma Cohort is a population ascertained cohort but the samples analysed here and in the study overall are inevitably selected as sufficient DNA to generate libraries is dependent upon the thickness of the primary tumour. Thus, my study is biased towards thicker poorer prognosis melanomas. If a potential biomarker is identified however I would plan to test such a biomarker in the smaller tumours if a specific test can be devised to test that biomarker.

A limitation of this work is the sampling of tumours using Tissue MicroArray (TMA) cores. As discussed in Chapter 2, using TMA cores does not allow confirmation of the tumour content throughout the core. There is potentially greater contamination with normal cells that in microdissected tumours. However, by using this method a greater range of tumours will be examined than in previous research based upon frozen tumours. Both DNA and RNA is extracted using a TMA core which preserves architecture of the block for clinical use which is considered crucial. The use of microdissection could possibly address some of these issues but it would be difficult to extract sufficient yields of DNA for whole-genome NGS studies and would be more time consuming for large-scale studies. My work is however designed to identify biomarkers for use in clinical practice and it is conceivable that sampling tumours using a TMA needle would be a more valid option than microdissection. This approach therefore seems reasonable as a standard approach to sampling was adopted.

4.5.2 Libraries

Libraries from 269 samples have been prepared to date (September 2013) and 92% of them were successful and produced enough yield to be sequenced. Although melanin score was inversely associated with DNA concentration, melanin score was not a predictor of successful library preparation. This result supports the view that BSA addition in the enzymatic reactions of library preparation can reduce the inhibitory effect of melanin. This is in accordance with the literature where addition of BSA or other proteins in PCR reactions is able to reverse the inhibitory effect of melanin and let DNA polymerase amplify the sequences (Eckhart *et al*., 2000, Mani *et al*., 2001). The only predictor of successful library preparation was DNA input. We managed to prepare libraries from DNA samples where the amount used was as low as 10.5ng but also from samples where DNA concentration was detected as 0ng/μl using the broad range picogreen assay (range of detection 2-1000 ng/μl). Age of the block and day of the week that the block was processed were not statistically significant predictors of successful library preparation. The median age of blocks of which successful libraries were produced was lower (10.13 years) than the median age of blocks of which the libraries failed to be generated (11.19). However, there was a single case where the age of the block was approximately 20 years and a successful library was produced. The above observations, unsurprisingly, suggest that DNA yield is a successful predictor of library success, however a cut-off value of DNA yield under of which we were not able to generate a library could not yet be established. A recent study has shown that a library could not be produced from a single 40ng FFPE-derived DNA sample that appeared to have organic contamination (high 260/230 ratio) (Simbolo *et al*., 2013) but this approach was not tested in this study. My work has also shown that PCR replication was greater in libraries produced from lower quantities of DNA, (see below) and as this work proceeds, an approach to defining the lower limits of DNA concentration which will generate reliable data must be explored.

During my PhD, a new library preparation protocol became available which replaced the manual protocol as discussed in Chapter 3. To explore the possibility that the different protocols might affect analysis of sequencing data, 16 samples with DNA samples available from two cores (biological replicates) were chosen and one library was prepared using the manual protocol and one library using the NEB protocol. Similar DNA input was used for both protocols to allow direct comparison of the sequencing data when these become available. It was shown that library yields were better when the NEB protocol was used which supported our decision

to replace the old protocol as it is much quicker and uses less PCR cycles than the manual protocol. As the NEB protocol uses fewer PCR cycles it is likely that fewer PCR replicates will be sequenced. This will be tested by direct comparison of the sequencing data generated from the 16 libraries prepared using the manual and their matched libraries using the NEB protocol.

4.5.3 Quality control of sequencing data

Library quality control was not the only aspect tested to assess the feasibility of using NGS technology with FFPE samples. A number of quality control steps were performed to exclude low quality data from subsequent analysis. Up to the submission of this thesis, 100bp paired-end sequencing data were available from 75 libraries (61 unique samples and 14 replicates) generated using the manual protocol. This protocol uses adapters with a 5bp unique barcode and a T always present at the 6th position to allow multiplexing. Every base is assigned a quality score, which reflects the probability of an incorrect call at that base for that specific read. The maximum quality score a base can be assigned is 40, which reflects the probability of 0.0001 that this base has been wrongly called. Typically, a quality score of greater than 28 is taken to indicate high quality data (Babraham Bioinformatics, 2010). The quality score assigned to the 6th base was dramatically low while quality scores were low for the subsequent 4 bases. The low quality score at the 6th position is expected, as there is no complexity across the whole sample at the position. However, the low quality scores observed in the subsequent few cycles is not easily explained (the issue was discussed with Illumina technical support). The signal intensity files from the sequencer revealed a drop in intensities of the other 3 bases for both reads at that position. It is unknown why a low quality score was assigned for the second pair only. The NEB protocol utilises a different barcoding system (indexing) where a 6bp index is present in the middle of the primer sequence. This 6bp index barcode is sequenced using a separate sequencing primer (as a completely different third sequencing run) and therefore, the issue described above will be not an issue in the data generated from the NEB libraries. For this analysis, we decided to remove the initial 4 bases (after the six of the barcode) from all sequences to remove these low quality calls in these data generated using the manual library method.

Lower quality scores were also assigned at the 3' end of the reads. The source of the lower quality scores at the 3' end is known as a phasing artefact (Kircher *et al*., 2011). Phasing occurs due to imperfections in the sequencing chemistry (incomplete incorporation of a nucleotide, or errors in bridge PCR yielding multiple

DNA species in a single spot) leading to the individual DNA molecules in the cluster getting “out of step” with each other during sequencing. Due to lower quality scores, the last 10 bases were removed from subsequent analysis.

Finally, we searched for and removed Illumina adapter sequences at the 3’end of the sequenced reads. Adapter sequences will interfere with mapping efficiency of the DNA sequences (Minoche *et al*., 2011).

In summary, based on the above observations it was considered essential to trim the first 10 and the last 10 bases as well as bases at the 3’ end of the sequences which matched part of the Illumina adapters to improve accuracy of the sequencing data. The trimming procedure is essential to remove low quality data and improves mapping efficiency (Minoche *et al*., 2011). The data were then aligned to the human reference genome and data were checked to identify other possible biases.

4.5.4 PCR amplification

PCR amplification is essential before NGS analysis to enrich for the fragments to be sequenced. After fragmentation and adapter ligation, 18 PCR cycles were used to amplify the ligated fragments using the manual protocol. The high number of PCR cycles used to generate libraries was essential due to the low DNA input used. We did observe a high number of PCR replicates associated with lower DNA input (predominantly where DNA input was less than 200ng) and increased age of the block. There were samples though that showed low level of sequenced PCR replicates even when DNA input was as low as 22.9ng or even for the DNA sample which could not be measured using the broad range picogreen assay as discussed above (0ng/µl). Two samples which were relatively recent (under 8 years) had high levels of replicated sequences. As above then DNA low yields are viewed as potentially problematic. It would be cost-effective to sequence these low DNA yield samples at lower depth to avoid sequencing of PCR replicates, which are likely to be uninformative. My intention is to revisit this when analysing the data generated using the NEB library protocol.

Comparison between raw and dereplicated data (after removal of PCR replicates) showed an even pattern of reads removed per chromosome (Figure 4-5), except for chromosome 16 due to “problematic” regions in that chromosome, which suggests that removing PCR replicates might not adversely affect subsequent analysis of the data, but this has not been addressed. PCR replicates are defined based on an identical mapping position. A single nucleotide variation seen inside the defined PCR replicate read would suggest that these might not be real PCR replicates and

could be further explored. In the context of this work, PCR replicates were removed as is widely performed in NGS studies (Ding *et al*., 2010). To reduce the frequency of PCR replicates, we tried to use more than 200ng DNA whenever possible for subsequent library preparations. It is possible that more samples should be multiplexed and sequenced in a single lane when lower DNA input is used to address the high level of PCR replicates observed. When more data are analysed we might be able to distinguish samples which will perform well from samples which will not to allow efficient sampling selection for future studies.

4.5.5 Problematic regions

During the exploration to identify biases that might affect the signal of interest in our data, a published study reporting problematic regions in the reference genome sequence came to our attention (Pickrell *et al*., 2011). These problematic regions were identified to produce false-positive copy number peaks in the ENCODE ChIP-sequencing data. Our data were in agreement with their study where increased coverage was seen in regions where there are multiple copies in reality but only one is reported in the reference sequence. Therefore, these regions were excluded from statistical analysis.

4.5.6 GC bias

Due to the massively parallel sequencing process, PCR amplification, cluster amplification and sequencing reactions (in which biases such as uneven coverage of GC-rich or AT-rich regions (Aird *et al*., 2011) can be easily introduced) should be standardised as much as possible. In the context of our work different relationships between GC content in 10kb regions and coverage were observed (Figure 4-10). This shows that the effect of GC content on coverage is not simple. These effects have also been reported by other studies and a correction is considered essential as these effects can dominate the signal of interest (Dohm *et al*., 2008, Benjamini and Speed, 2012). We decided to adopt a linear correction strategy where the chromosome was included as a covariate while other proposed models do not allow for additional covariates (Gusnanto *et al*., 2012).

4.5.7 Coverage and reproducibility

The median sequencing coverage was 1.14x (range, 0.37x-1.64x). The previous study by Wood *et al* and the most recent from the same group by Belvedere *et al* using tissue from oral or lung squamous cancers have reported CNAs in FFPE-

derived tumours using NGS data with much lower coverage (range 0.004x-0.06x) (Wood *et al*., 2010, Belvedere *et al*., 2012). These studies were multiplexing more samples per lane and therefore less sequencing per sample was performed. We expect that the relatively high coverage NGS data generated here will enable us to detect smaller CNAs. However, this has not yet been addressed.

Not surprisingly, the only predictor of coverage was DNA input. Neither age of the block or melanin content was associated with coverage. The correlation coefficients between non-replicate pairs was much lower for non-problematic regions than for the problematic regions which provides further evidence that the previously identified problematic regions by Pickrell *et al* are indeed unreliable. For non-problematic regions the median correlation coefficient between replicate pairs was 0.89 compared to 0.49 for non-replicate pairs, which suggests that the assay is suitable for FFPE samples. There was no difference between raw and dereplicated data, which further supports our view that PCR replicate removal might not adversely affect subsequent analysis. The correlation coefficients between replicate pairs of the same libraries, replicate pairs of libraries generated with different DNA input, and replicate pairs of libraries generated using different PCR cycles was high (Spearman's rho range 0.81-0.94) and no difference could be observed between the different groups. The equivalent good correlation between replicate libraries generated with different DNA input (25ng versus 250ng or 25ng versus 100ng) shows that comparable data can be generated when even very low DNA input is used which is important as only small amounts of DNA are available from some FFPE samples. However, this is a case where an FFPE tumour produced sufficient DNA to run 3 independent libraries. This might not be a representative example of primary melanomas overall which commonly will only produce very small amounts of DNA sufficient for one library. As already discussed, further work will be needed when more sequencing data will be available to try and predict which samples need to be sequenced and at what depth.

4.5.8 Copy number alterations

Biases previously reported to affect NGS data analysis (GC bias, PCR duplication, problematic regions in the reference) were explored and correction was performed as already discussed. As more and more NGS data become available it is possible that these biases will be better understood or more appropriate tools and methods will arise to correct for them. Also there are other as-yet unknown errors that we might find.

Following the correction of biases observed and assessment of the reproducibility of NGS using FFPE samples, data were analysed to identify CNAs in tumours of patients with well-documented characteristics, such as Breslow thickness and mitotic rate. This analysis aimed to explore the feasibility of NGS technology to produce high quality data from FFPE samples. Univariable analysis was performed to identify CNAs associated with MSS and other tumour characteristics known to be predictors of survival. As matched normal reference samples were not available, it is not possible to accurately detect the number of alleles that had been gained or lost. Median normalised data on the log 2 scale though can be interpreted as read counts >0 depict copy number gain and <0 depict copy number loss. Regression analysis between each 10kb region and mitotic rate showed the most significant associations. CNAs of whole chromosome arms such as 7q and 8q were significantly associated with mitotic rate as observed in Figure 4-14 and Figure 4-17. Frequent copy number gains of 7q and 8q have been previously identified in melanoma (Bastian *et al.*, 2003, Jonsson *et al.*, 2007, Rakosy *et al.*, 2013). We confirmed that chromosome 8q gain was associated with mitotic rate but this was not the case for 7q where similar copy number was seen in all patients by looking at read counts. The linear regression model relies on the assumption that the residuals are normally distributed and as mitotic rate values were not normally distributed some of the significant associations resulted from the regression analysis could be erroneous. A robust regression model, which will not assume that the data are normally distributed should be applied in the future, and I will explore that after submission of this thesis. To reduce noise the coefficients were smoothed using a bandwidth 501 windows running median. The limitation of this approach is that associations between the phenotype and small regions (<<5Mb) will not be picked up.

The strongest associations between CNAs and a clinico-pathological factor were seen for mitotic rate although regression analysis between each 10kb region and Breslow thickness showed individual windows and relatively small chromosomal regions to be associated with Breslow thickness. Regression analysis between each 10kb region and MSS, ulceration, AJCC stage, tumour site and clinico-pathological subtypes of melanoma showed no significant associations. The above observations are hypothesised to be explained at least in part as these features are more complex than tumour mitotic rate. Mitotic rate is representative of how proliferative a tumour is while, for example, Breslow thickness depends on the stage of diagnosis (how readily a patient seeks advice and how effectively he is managed by medical services). Mitotic rate is postulated to be a much simpler

measure reflecting tumour biology rather than a composite of biology and health care. It is possible that when we increase the sample size more associations between copy number and tumour characteristics will be identified.

Regression analysis was also performed between each 10kb region and *BRAF*/*NRAS* mutation status. The most significant association was observed between *BRAF* mutated tumours and loss of chromosome arm 10q, where the tumour suppressor *PTEN* is located. This finding has been previously identified in melanoma cell lines and frozen tumours (Jonsson *et al*., 2007, Gast *et al*., 2010) which is pleasing, suggesting validity of the data generated from small primary FFPE samples.

In summary, here an agnostic simple approach was taken to explore whether copy number alterations were associated with tumour characteristics. This approach was able to validate previous findings in melanoma, such as 8q gain and 10q loss in a subset of tumours. While more associations where found such as between chromosome 1p, 6p, 17p and 20p and mitotic rate I haven't explored if these regions were gained or lost. More robust methods will be explored using this small data set and applied to the larger dataset when available to increase power.

4.5.9 Gene expression and copy number data integration

It has become apparent that melanoma inter-tumour heterogeneity may be due to both genomic and transcriptional differences (Jonsson *et al*., 2007, Bastian *et al*., 2003, Jonsson *et al*., 2010, Rakosy *et al*., 2007, Jewell *et al*., 2010). Therefore, integration of genomic and transcriptomic data from the same tumour sample might allow us to identify pathways and genetic alterations that influence the biology of melanoma and facilitate molecular classification of tumours. Copy number changes would only be expected to be strongly associated with gene expression changes in a proportion of genes as gene expression might also be influenced by epigenetic mechanisms. Recently, a study reported integration of copy number, gene expression and methylation data generated from 36 primary melanoma tumours (Rakosy *et al*., 2013). This study identified gene expression signatures associated with copy number changes in ulcerated tumours but also established inverse correlations between gene expression and methylation patterns for a proportion of genes, suggesting that transcriptional silencing of these genes is mediated through epigenetics (Rakosy *et al*., 2013).

In the context of this work, whole-genome DASL gene expression and NGS copy number data were available from 47 tumours. Chromosomal locations significantly

associated with mitotic rate were identified and Spearman correlation coefficients were calculated between DASL probe signals and NGS read counts of regions where the DASL probe(s) were located (10kb or smoothed regions). Correlations between gene expression measures and copy number were poorer when read counts of individual 10kb windows were used but stronger when the smoothed windows of 5Mb size were used. This might reflect the reduction of noise when smoothed windows were used to calculate correlations. Another important source of this observation might be the fact that the annotation file for DASL probes provided by the company was erroneous. Two weeks before the submission of this thesis I identified that some probe coordinates were based on the human reference genome version hg18 while some others on the hg19. The results presented here therefore should be treated with caution. However, the correlations between gene expression measures and smoothed windows might not reflect erroneous results if we think that the position of most of the genes on one genome assembly (hg18) will possibly be with 5Mb of the corresponding position on a more recent genome assembly (hg19). This integrated analysis revealed 159 DASL probes whose signal was significantly associated with copy number and therefore were used for pathway analysis using MetaCore™ tool. The top three significantly overrepresented pathways were IL-2 signalling, cell cycle and DNA damage response pathways. *TP53* (17p) and *GAB2* (11q) were among the genes identified in these pathways. *GAB2* amplification has been previously identified in melanomas and associated with metastatic potential (Horst *et al*., 2009) while *TP53* is a known tumour suppressor gene known to be silenced in melanoma (Meyle and Guldberg, 2009). However, I have not managed to look for the direction of the association between the expression of these genes and mitotic rate in my data yet. More importantly, the data need to be re-analysed using correct probe annotations in the hg19 and after identification of significantly altered regions using more robust models as described above to draw sound conclusions.

As an additional quality control measure, genes previously related with melanoma/melanoma progression were selected and correlations between gene expression measures and copy number were calculated. This analysis revealed that gene expression measures of genes previously identified to be deleted in melanoma (*CDKN2A* and *PTEN*) and a gene previously reported to be gained in melanoma (*MYC*) were significantly correlated with copy number. Interestingly, significant correlations were established for two DNA repair genes known to be associated with melanoma progression: *RAD52* and *TOP2A*. RAD52 is located in 12p region which has not been reported to be frequently altered in melanoma while

TOP2A is located in 17q, a region previously identified to be altered in melanoma (Bastian *et al*., 2003, Jonsson *et al*., 2007). Other genes failed to establish a significant correlation except for *SPP1* whose expression was significantly inversely correlated with copy number. The latter observations might depict the fact that other mechanisms alter the expression of these genes or that the copy number data for 10kb windows were quite noisy. Overall however the reported associations provided further support for the validity of the NGS data produced.

4.5.10 Conclusion and future work

Further work will help clarify which sample characteristics will predict performance with NGS assays and enable more extensive studies to assess CNAs in primary melanomas at the right depth. I have shown that low DNA concentrations are, not surprisingly, associated with lower probability of adequate library preparation and more PCR replicates sequenced and therefore potential data bias, but that there was no clear cut-off identifiable to date. I think it is likely that there is an optimal lower concentration which will vary according to method of library preparation (and therefore number of PCR cycles), and number of samples run per lane. I will address this in the months to come.

I have reported evidence that the CNAs detected were in regions previously described in melanoma and which are biologically credible (such as chromosome 10/ *PTEN* in *BRAF* mutated tumours) and this is evidence that my approach to analyzing the data is reasonable. The data reported here are addressed to methodology. In the future I hope to explore this further and by using more robust models and a larger sample set I will identify if CNAs might improve prognostication. Computational NGS data analysis is an active research area and therefore the combination of different analysis methods might be the best solution to accurately identify CNAs. A recent study in melanoma using aCGH reported that patients with bad prognosis showed more frequent small CNAs compared to patients with good prognosis (Hirsch *et al*., 2013). This suggests that a pattern of CNAs might prove of prognostic significance rather than a single CNA.

NGS has successfully been used in the clinic where pilot studies have assessed the practical challenges of this technology to identify useful information in a clinically relevant time frame (Rizzo and Buck, 2012). For example, a study has reported the use of both whole-genome and targeted exome NGS which allowed the identification of mutational landscapes in tumours from patients with colorectal cancer and melanoma. The identification of these mutations within 24 hours of

biopsy enabled patients enrolment in biomarker-driven clinical trials (Roychowdhury *et al*., 2011).

Using FFPE tissue for NGS is challenging, however these efforts enable analysis of large numbers of samples with mature survival data and well-annotated tumour and patient characteristics. Studies of this kind will undoubtedly improve our understanding of the biological processes associated with melanoma progression and allow identification of prognostic biomarkers.

Chapter 5
The effect of 1α,25$(OH)_2D_3$ on the growth and gene expression of cultured melanoma cells

5.1 Aims

The aims of this chapter are:

- To assess the reported anti-proliferative effect of 1α,25$(OH)_2D_3$ on cultured melanoma cells;
- To assess the vitamin D receptor (VDR) status of the melanoma cell lines using immunocytochemistry;
- To identify differentially expressed genes and altered pathways in response to 1α,25$(OH)_2D_3$ using microarray technology in cells showing evidence of an anti-proliferative effect when cultured with 1α,25$(OH)_2D_3$;
- To identify if VDR target genes are overrepresented in the significantly expressed genes in response to vitamin D_3 using an *in silico* approach.

5.2 Background

Vitamin D, as discussed in Chapter 1, is a fat-soluble hormone, which is very important to human health and has been implicated in reducing the risk of cancer initiation and progression (Holick, 2004). There is no doubt that vitamin D deficiency causes bone disease, but there are numerous reports showing that deficiency and insufficiency increase the risk of a multitude of health problems, many of which associations remain controversial. An independent association in the Leeds Melanoma Cohort has been reported between lower serum vitamin D levels at diagnosis and thicker primary melanomas and worse prognosis even when the data are adjusted for Breslow thickness (Newton-Bishop *et al.*, 2009). Vitamin D levels at diagnosis would therefore seem to be a prognostic biomarker.

The identification of an association between vitamin D levels and outcome from melanoma does not however establish a causal relationship between levels and

survival, and the group has been attempting to devise approaches to understand the significance of these associations. We hypothesised that if vitamin D was playing an active role on progression of melanoma, then we might be able to detect an association between gene expression in melanoma primaries stored from patients participating in the Leeds Melanoma Cohort and vitamin D levels in the patient at the time of diagnosis. A study was therefore designed to identify transcriptional patterns (using Illumina DASL arrays) in primary melanomas associated with the patients' vitamin D levels at diagnosis, comparing tumours from patients with high serum vitamin D levels and tumours with low vitamin D levels. This project has completed data accrual but results are not as yet available. When these data are available, the intent is to determine if the gene expression patterns mirror biological pathways reported to be related to signalling through the vitamin D receptor (Ramagopalan *et al*., 2010, Heikkinen *et al*., 2011).

Reported associations do not however prove causality and there are especial concerns with respect to vitamin D as levels are higher in fitter, healthier people, and some argue that reported associations between higher vitamin D levels and better outcomes might mean that the levels are merely acting as a marker of better lifestyles. Therefore, the aim of the work described in this chapter was to investigate the transcriptional changes of added $1\alpha,25(OH)_2D_3$ (henceforth referred to as vitamin D_3) to a range of cultured melanoma cells. In turn, these observations will be compared and contrasted with the forthcoming primary tumour expression data when it becomes available. The cell culture work described here was performed in collaboration with Professor Jorg Reichrath at Saarland University, in Homburg, Germany.

An anti-proliferative effect of vitamin D_3 *in vitro* has been reported in many studies of different cancers reviewed by Deeb *et al* (Deeb *et al*., 2007) and in studies using melanoma cells reviewed in (Table 1-4). The concentrations of vitamin D_3 used in the studies reported here (10^{-10}-10^{-7}M) have previously been used to show a dose-dependent anti-proliferative effect and furthermore distinguish between vitamin D sensitive and resistant cultured melanoma cells (Seifert *et al*., 2004). These data were used to inform the experimental design of the work described here.

A number of studies have previously used microarray platforms to identify gene expression patterns associated with the addition of vitamin D to culture media in different cancers but not in melanoma. A non-genomic (rapid) response through

Ca^{2+} influx and G-protein coupled receptors but also a genomic response through the classical VDR pathways have been reported upon vitamin D_3 stimulation *in vitro* and *in vivo* (Deeb *et al*., 2007). As a result, in the microarray experiment described in this chapter, melanoma cells were treated with vitamin D_3 for 6, 24 and 48 hours as a means to identify both early and late transcriptional changes. Differentially expressed genes were identified by comparing the expression profiles of vitamin D_3 treated, and untreated samples. Here, I have explored the use of bioinformatics databases to try to better understand the significance of these differentially expressed genes in response to vitamin D_3.

Gene Ontology (GO), is one such database that assigns genes to functional categories, and this has been reported to be useful in understanding the biology behind extensive lists of differentially expressed genes (Ashburner *et al*., 2000, Werner, 2008). Genes can be assigned into a category based on their location in the cell within 'cellular component' terms, based on their activity under 'molecular function' terms and based on events using one or more 'molecular function' terms from within 'biological processes' terms (Ashburner *et al*., 2000). Recent studies have used GO terms to understand biological functions of differentially expressed genes generated after statistical analysis. For example, it has been reported that genes involved in vitamin D metabolism, cell cycle, apoptosis, cell adhesion and immune response are differentially expressed in response to vitamin D_3 (Deeb *et al*., 2007, Fleet *et al*., 2012).

Pathway analysis has also been used to identify chains of biological events when microarray analysis is conducted (Werner, 2008). A number of databases, such as KEGG (Kyoto Encyclopaedia of Genes and Genomes) (Kanehisa *et al*., 2012), have constructed pathways based on literature research and differentially expressed genes derived from microarray data can be mapped onto these databases to unravel particularly over-represented pathways (Werner, 2008). A recent study in breast cancer cells has shown that vitamin D_3 treatment deregulates genes involved in the TGF-beta pathway (Milani *et al*., 2013). Another study in cultured podocytes (highly differentiated kidney cells) has shown that the NF-κB pathway is targeted in response to vitamin D_3 treatment (Cheng *et al*., 2013).

It is important to rank the associations found with GO terms and pathways using microarray data (Werner, 2008), and therefore web-based tools have been designed, such as MetaCore™, which uses hypergeometric distribution tests,

representing the probability that the reported association is due to chance (Ekins *et al*., 2007).

Here, I report the first study designed to identify genes and pathways altered in response to 10^{-7}M vitamin D_3 treatment of melanoma cells using a microarray approach. The Affymetrix U133 plus 2 microarray platform was used as described in section 2.8. Generation of large-scale gene expression data was then processed using the MetaCore™ web-based tool to identify over-represented biologically relevant annotations related to vitamin D_3 treatment.

5.3 Methodology

The anti-proliferative and microarray work were described in detail in sections 2.5.6 and 2.8 respectively. In summary, two different primary melanoma cell lines, SkMel28, MeWo and a cell line reputed to be MelJuso (which was later established not to be MelJuso and which is here referred to as 'MelJuso') were treated with 10^{-10}M, 10^{-8} and 10^{-7}M vitamin D_3 or ethanol controls. The SkMel5 cell line was also used but data are not presented as these cells were established to be SkMel28. The failure to establish the authenticity of the 'MelJuso' cells was described in section 2.5.2. In an attempt to identify if this cell line is of melanocytic origin immunocytochemistry (ICC) was performed using antibodies against melanocyte markers (HMB45, Melan-A, S100, Pan Melanoma Cocktail). Cell proliferation was measured using the crystal violet assay after 24, 48, 72 and 144 hours in order to confirm sensitivity to vitamin D_3. I attempted to do this experiment initially in Leeds, but had difficulties managing delivery of the fat-soluble vitamin D_3 in ethanol to the cells and therefore established a collaboration with Professor Jorg Reichrath of Saarland University, Homburg, Germany. Two months were spent performing the cell culture experiments within the laboratory in Saarland University. To identify the VDR status of the three cell lines reported here, VDR protein expression was measured using ICC as described in section 2.5.8.

To identify differential gene expression an independent experiment was performed and melanoma cells were treated with the concentrations of vitamin D_3 mentioned above or ethanol controls, and RNA was extracted after 6, 24 and 48 hours. The RNA extracted from the cell lines whose proliferation was reduced in vitamin D treated cells, SkMel28 and MeWo melanoma cell lines after treatment with 10^{-7}M vitamin D_3 and the ethanol controls was sent for microarray

analysis. Microarray data generation and data pre-processing are described in detail in sections 2.7 and 2.7.1.

5.3.1 Statistical analysis

5.3.1.1 Proliferation assay

The proliferation was measured using crystal violet staining as described in section 2.5.6. Each cell line was assayed 3 times, with 4 biological replicates at each time point. Three measurements of crystal violet cellular uptake were taken for each of the biological replicates. The mean of these 3 measurements for each biological replicate was used for statistical analysis. One way analysis of variance (ANOVA) test was performed using log to the base 2 transformed values to assess statistical significance in proliferation between treated cells with different concentrations of vitamin D_3 compared to controls at all time points. Then, stratified analysis was performed using ANOVA at each time point independently. Finally, pairwise comparisons between treated cells and controls using the Mann-Whitney test were done to reassure the first analysis. A statistically significant difference was considered established when the Bonferroni-adjusted p value was <0.017 (multiple testing correction for the 3 cell lines analysed). Statistical analysis was done in STATA v.11 (StataCorp., 2007).

5.3.1.2 Gene expression assay

The analysis was performed by bioinformatician Dr Alastair Droop, using the LIMMA package (Smyth, 2004) available in Bioconductor (Gentleman *et al.*, 2004), in R (R Development Core Team, 2012). RMA-transformed probe set intensity values were used to identify differentially expression signatures between vitamin D_3 treated samples at 6, 24 and 48 hours and corresponding controls. Briefly, LIMMA fits linear models and borrows information about the within group variance, calculates t and F-statistics for each probe set and assigns p values to identify if a probe set is statistically significantly expressed (Smyth, 2004). It has been shown that LIMMA outperforms other variance modelling strategies in the analysis of gene expression microarray data in terms of false-positive rate, execution time and its ability to perform well even in small sample sizes (Jeanmougin *et al.*, 2010). The log fold changes (relative expression values) for each probe set (targeting a specific gene), reported by LIMMA, are the differences between the log expression value in the treated

sample and the control sample. Multiple testing correction was performed by applying false discovery rate (FDR) to the p values (Benjamini and Hochberg, 1995). Significant difference was considered when FDR<0.1.

For simplicity, henceforth a probe set will be referred to as a probe. As described in section 2.7 multiple probes target a specific gene and because the statistical analysis was performed at the probe level the terms ‘differentially detected probes’ and ‘differentially expressed genes’ are used interchangeably.

For each cell line, LIMMA produced a list of probes whose signal was statistically significantly differentially detected when vitamin D_3 treated cells were compared to controls at each time point independently. LIMMA also produced a list of probes whose signal was statistically significantly differentially detected between vitamin D_3 treated cells and controls overall at all time points for each cell line and this list is referred to below as the “complete list”. Then, a number of different approaches were used using the above lists of probes to identify biologically interesting genes altered in response to vitamin D_3. The steps are shown schematically in Figure 5-1.

Firstly, unsupervised hierarchical clustering for each cell line irrespective of time was performed using the top 400 significant probes from the complete lists and heatmaps were plotted. Then, a number of analyses using various bioinformatic approaches were carried out as described in the following section.

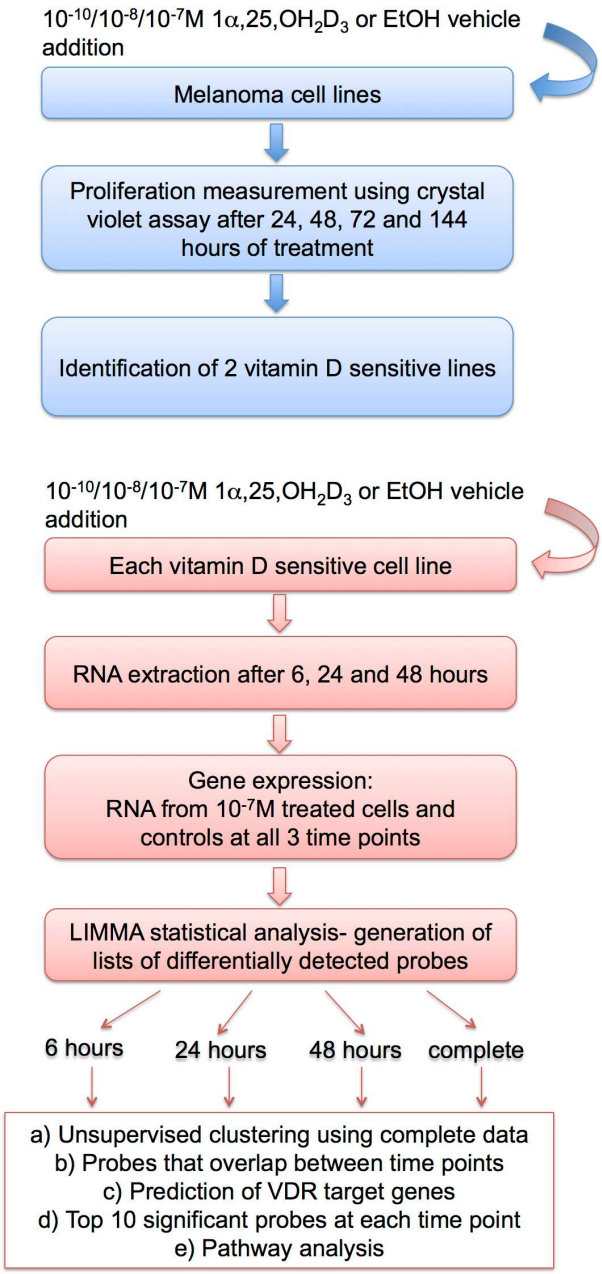


Figure 5-1: Schematic of proliferation and gene expression experiments using cultured melanoma cells.

Three melanoma cell lines were tested to identify the effect of added vitamin D_3 on proliferation. From the 3 lines tested, 2 lines showed evidence of inhibition of proliferation after treatment for 144 hours and these lines are referred to as “vitamin D sensitive” lines. The vitamin D sensitive cell lines were selected for gene expression studies. RNA extracted from vitamin D_3 treated and control cells was analysed using the Affymetrix U133 plus 2 whole genome array and statistical analysis was performed using LIMMA (Smyth, 2004) to identify differentially expressed genes.

5.3.1.3 Bioinformatics

5.3.1.3.1 Prediction of VDR target genes

Computational analysis was performed to predict which of the genes that were differentially expressed in response to vitamin D_3 are direct VDR targets. VDR targets were defined as all genes with a VDR binding sequence known as a vitamin D response element (VDRE) in their promoters. The "complete lists" were used and the promoters of the genes were defined as 5kb upstream or downstream from the transcription start site. The matrix for the RXRA-VDR binding sequence from JASPAR database (ID:MA0074.1) was used to predict VDREs in the promoters of genes (Bryne *et al.*, 2008) (Figure 5-2). The 10 known binding sequences were used to create a position weight matrix (PWM) for the RXRA-VDR binding site. The defined promoters were matched against this PWM with a minimum score of 75% (Wasserman and Sandelin, 2004). This procedure was also performed for all genes (all annotated genes in the human reference genome version hg19) in the human genome and a one way Fisher's exact test was performed to identify if VDRE enrichment was seen in the list of significant genes after vitamin D_3 treatment. Dr Alastair Droop performed the bioinformatics analysis using R (R Development Core Team, 2012) and Bioconductor (Gentleman *et al.*, 2004) and I performed the statistical analysis using Stata (StataCorp., 2011).

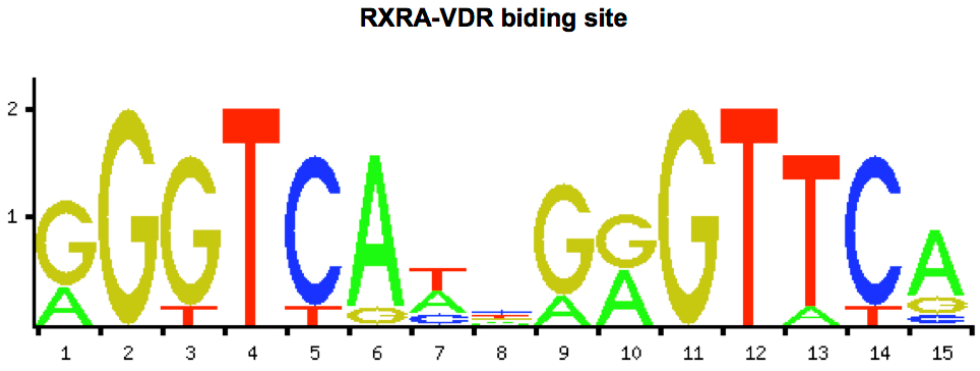


Figure 5-2: RXRA-VDR heterodimer binding site from the JASPAR CORE database

Visual representation of the position weight matrix (PWM).

5.3.1.3.2 Enrichment analysis using MetaCore™

MetaCore™, a commercially available web-based tool (Ekins *et al*., 2007), was used to identify biological processes and pathways that were enriched in response to vitamin D_3. Genes that were attributed a significant FDR<0.1 by LIMMA at each time point for each cell line individually were considered for this analysis. Overrepresented GO processes and pathways were identified based on p values and FDRs. An FDR<0.1 was used to determine statistical significance.

5.4 Results

5.4.1 Proliferation assay

The effects of vitamin D_3 on the proliferation of MeWo and SkMel28 cells were in accordance with previously published data (Seifert *et al*., 2004). No difference in proliferation was observed when all cell lines were at low cell density (at 24, 48 and 72 hours). MeWo and SkMel28 but not 'MelJuso' melanoma cells showed a statistically significant decrease in proliferation after 144h hours when treated with 10^{-8}M and 10^{-7}M vitamin D_3 (ANOVA stratified analysis at 144 hours, Bonferroni-adjusted p<0.001). Figure 5-3 shows the absorbance of the crystal violet dye at 550nm which is relative to the number of cells per treatment group at each time point for all three cell lines. Henceforth, MeWo and SkMel28 cells will be referred to as vitamin D sensitive and 'MelJuso' cells as vitamin D resistant. Figure 5-4 shows images of vitamin D sensitive SkMel28 and vitamin D resistant 'MelJuso' cells after vitamin D_3 treatment or vehicle treatment after 144 hours.

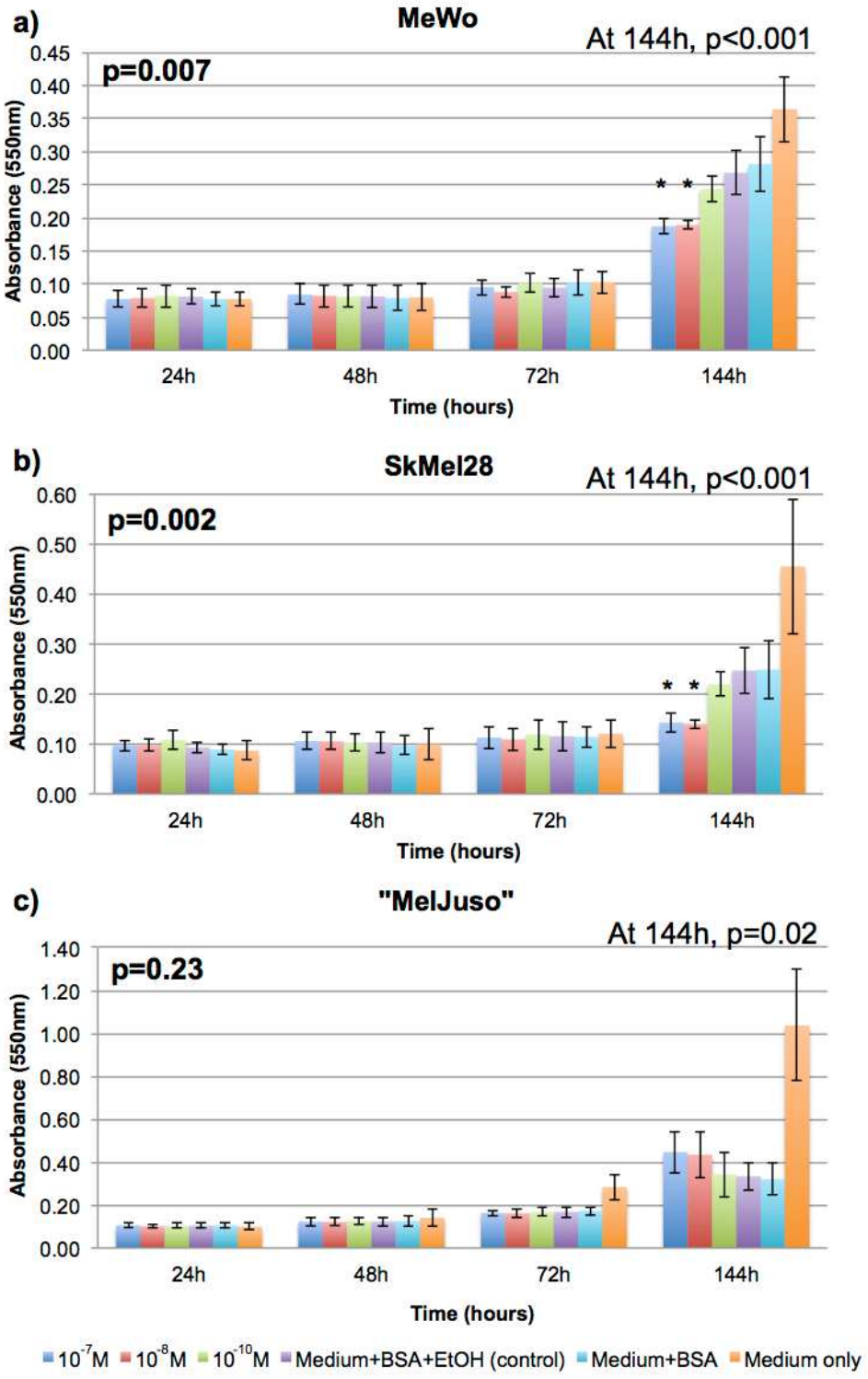


Figure 5-3: Crystal violet dye uptake by vitamin D_3 treated and vehicle-treated cells after 24, 48, 72 and 144 hours for a) MeWo, b) SkMel28 and c) "MelJuso" cell lines.

(Figure description on the next page)

(Figure 5-3 description): Each column represents the mean ± standard deviation of 12 biological replicates from 3 independent experiments. Statistical significance was established when the Bonferroni-adjusted p value was <0.017 (multiple testing correction for the 3 cell lines analysed). The p values reported in the figures represent the Bonferroni-adjusted p values. The highlighted p values on the left corner were reported from the ANOVA analysis and represent the overall effect of vitamin D_3 treated cells compared to controls irrespective of time. The stratified ANOVA analysis performed at each time point showed a statistically significant difference at 144 h for MeWo and SkMel28 cells ($p<0.001$) but not for 'MelJuso' ($p=0.02$). The asterisk refers to the statistically significant difference between treated and control values when pairwise comparisons were performed using the Mann-Whitney test to re-assure the first analysis.

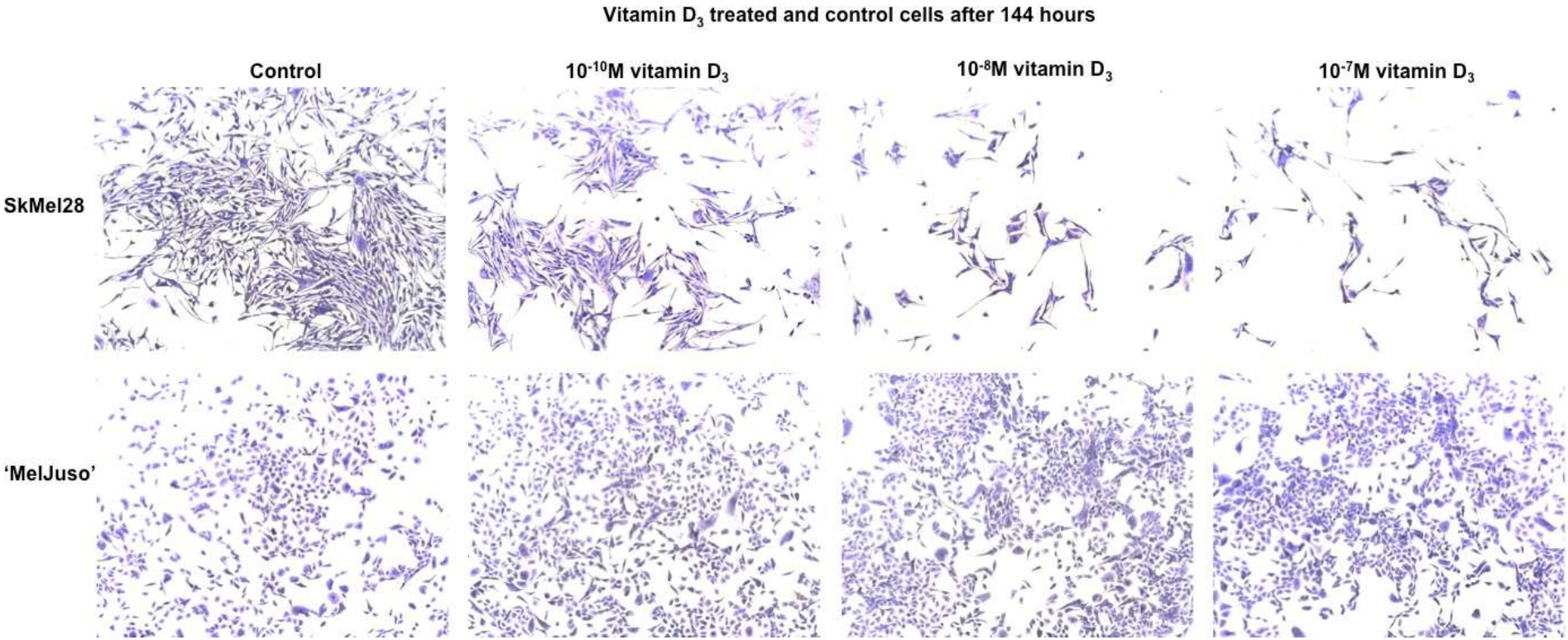


Figure 5-4: SkMel28 and 'MelJuso' cells after 144 hours of vitamin D_3 and vehicle-control treatment.

Inhibition of cell proliferation observed at the end of the proliferation experiment for SkMel28 cells treated with 10^{-10}M, 10^{-8}M and 10^{-7}M vitamin D_3 compared to controls. On the contrary, no inhibition of cell proliferation was observed for 'MelJuso' cells treated with vitamin D_3 compared to controls (10x magnification).

5.4.2 Immunocytochemistry using melanocyte markers

As described in section 2.5.2, ‘MelJuso’ cells could not be authenticated. Protein expression was demonstrated by ICC using antibodies raised against well-described melanocyte markers: HMB45, Melan-A and S100. Also a Pan Melanoma Cocktail was used consisting of antibodies against HMB45, Melan-A and Tyrosinase. Non-specific binding was observed in negative controls but only in those cells that were rounded up or demonstrating reduced cell adherence. The “MelJuso” cells did show expression of the melanocyte markers tested compared to negative controls suggesting that these cells are of melanocytic origin (see Figure 5-5).

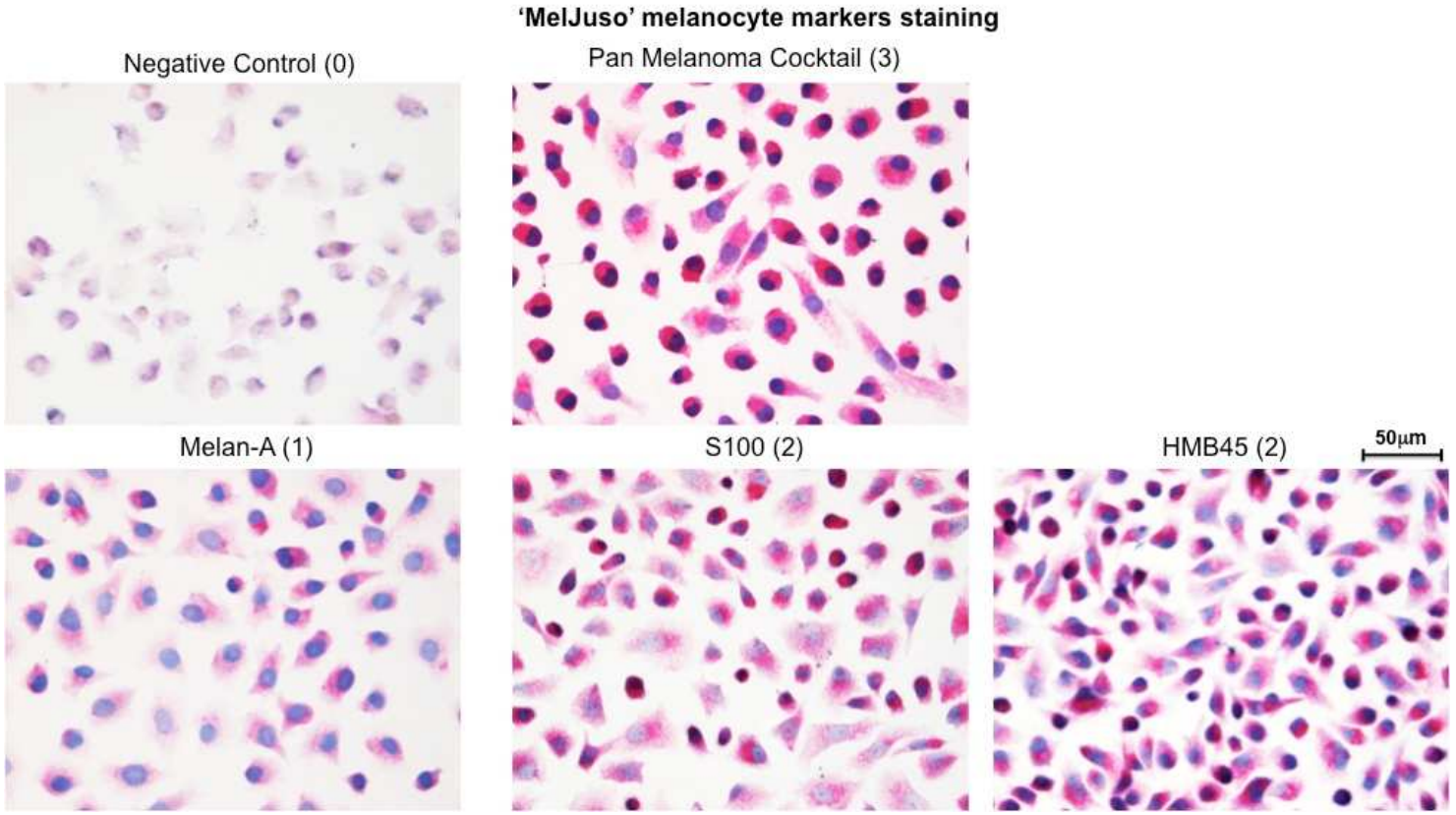


Figure 5-5: Melanocyte markers immunocytochemistry in 'MelJuso' cells.

The 'MelJuso' cells showed expression of the melanocyte markers tested compared to negative controls suggesting that these cells are of melanocytic origin (40x magnification). The numbers in brackets represent the staining intensity where 0 represents no staining and 3 strong staining.

5.4.3 VDR immunocytochemistry

Protein expression was demonstrated by ICC using an antibody raised against VDR which has been used previously in the immunohistochemistry of melanoma tumours (Brozyna *et al*., 2011). A titration experiment was initially performed to identify the optimal dilution of the primary antibody. Figure 5-6 shows the VDR staining of SkMel28 cells using 3 dilutions of the primary antibody (1:2000, 1:1000, 1:500) and the negative control. The 1:500 dilution was selected as the most appropriate and it was used to stain all 3 melanoma cell lines used in the proliferation assay. Non-specific staining was observed in all three cell lines in negative controls but only in those cells that were rounded up or demonstrating reduced cell adherence. VDR was shown to be expressed in all three cell lines irrespective of their sensitivity to vitamin D_3 (Figure 5-7) although the MeWo cells appeared to show less VDR expression than SkMel28 and 'MelJuso' cells.

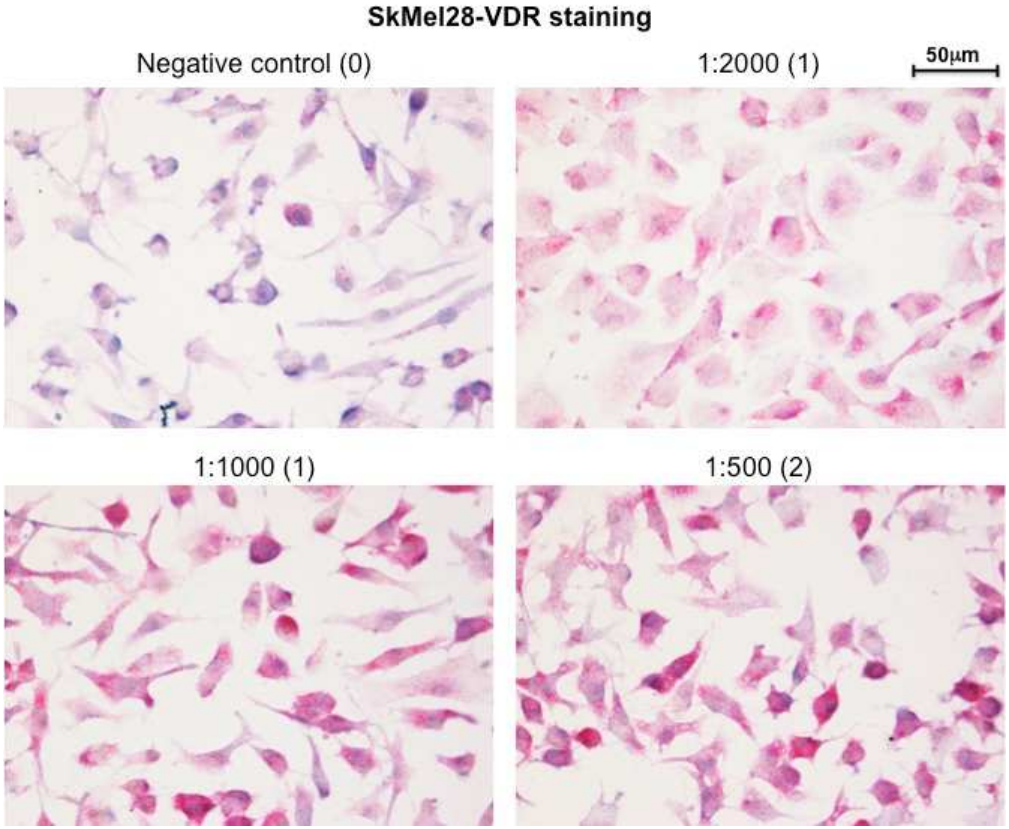


Figure 5-6: Titration of anti VDR antibody in SkMel28 cells.

Representative images of antibody dilutions of 1:2000, 1:1000 and 1:500 compared to negative control (40x magnification). The 1:500 dilution was selected as the optimal concentration to stain all 3 cell lines. The numbers in brackets represent the staining intensity where 0 represents no staining and 3 strong staining.

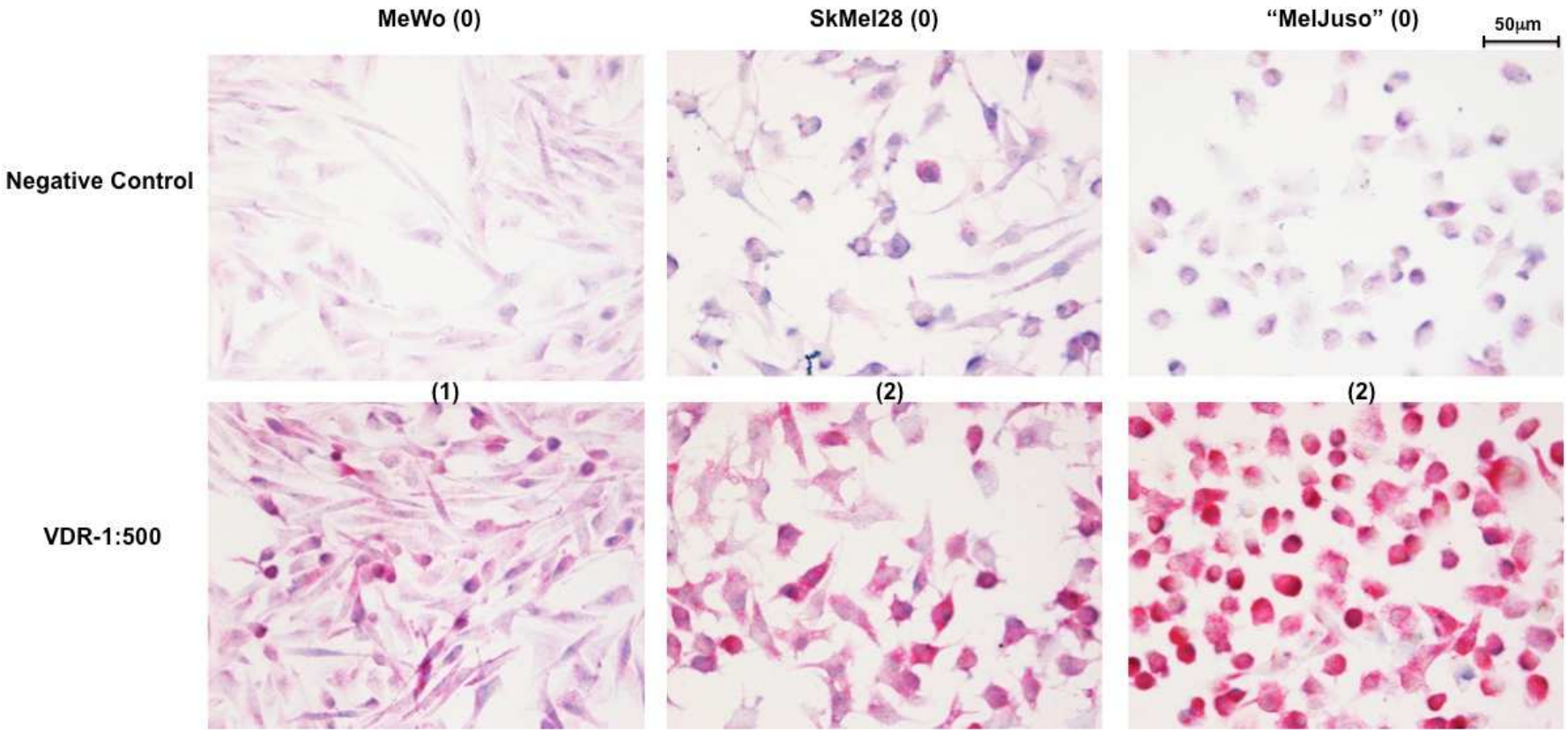


Figure 5-7: VDR immunocytochemistry using 1:500 dilution of anti-VDR antibody for MeWo, SkMel28 and "MelJuso" cells.

All cell lines showed VDR expression compared with their controls irrespective of their sensitivity to vitamin D_3 (40x magnification). The numbers in brackets represent the staining intensity where 0 represents no staining and 3 strong staining.

5.4.4 Identification of differentially expressed genes in response to vitamin D_3

SkMel28 and MeWo cell lines were treated with vitamin D_3 (10^{-7}M) or vehicle and harvested at 6, 24 and 48 hours. RNA was extracted and assayed to identify genes differentially expressed. One sample (a 6-hour vehicle-treated MeWo replicate) was degraded (RIN<7) and was not run on the array. Quality control measures during data pre-processing as described in section 2.7.1 showed that the assay was successful. Only one sample failed to pass the quality control analysis and was therefore excluded. Three lists of significantly detected probes were obtained: one for each time point, and a complete list irrespective of time as described in 5.3.1.2 (the full lists are shown in Appendix A.4.1). Unsupervised clustering was performed using the top 400 significantly detected probes from the complete lists for each cell line individually and heatmaps were plotted (Figure 5-8).

The heatmaps show that replicates at each time point clustered together highlighting the reproducibility of the data. In addition, the control and treated samples cluster independently of one another demonstrating quite different profiles. The marked exception is the treated samples harvested at 6 hours which demonstrate gene expression profiles that are more closely related to the control groups.

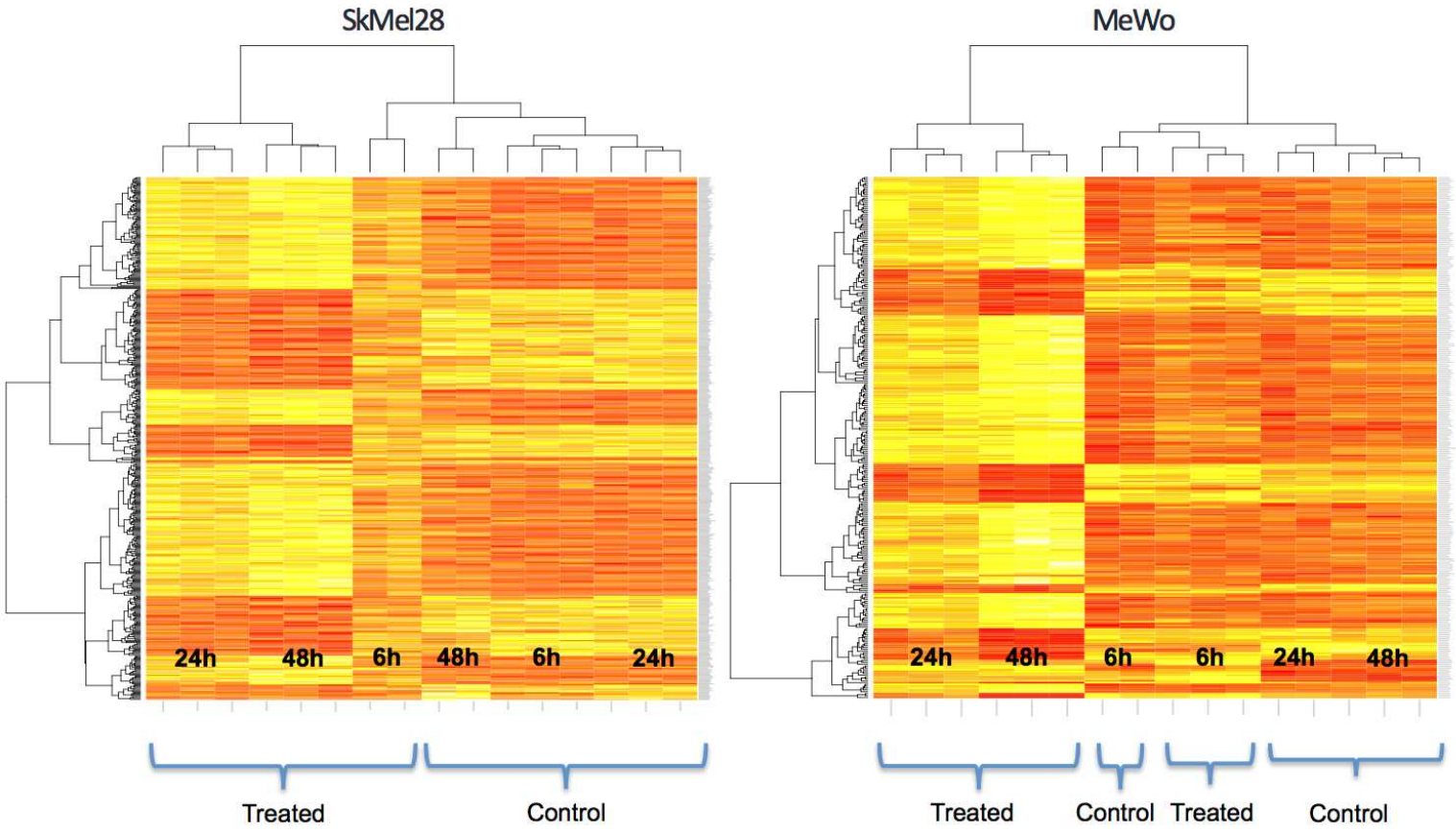


Figure 5-8: Heatmaps of the top 400 significant probes from the complete lists for SkMel28 and MeWo cell lines.

(Figure description on the following page)

(Figure 5-8 description): The signal for each probe is represented by coloured panels with lighter (yellow) colours representing very low signal and darker (red) colours representing high signal. The heatmaps were plotted after unsupervised hierarchical clustering was performed irrespective of time for each cell line independently. In each cell line two large clusters were evident representing the vitamin D_3 treated and the control group. In contrast, in both cell lines the 6-hours treated samples cluster with the control group. All replicates at each time point cluster together, highlighting the reproducibility of the data.

At 6 hours, 129 probes were significantly differentially detected in SkMel28 cells and only 19 probes in Mewo cells (Figure 5-9). Analysis of the probes revealed that the gene *CYP24A1* encoding for the vitamin D_3 degradation enzyme was the top overexpressed gene in both cell lines at 6 hours and remained overexpressed at later time points too. The number of probes that were differentially detected at more than 1 time point in SkMel28 and MeWo cells independently are shown in Figure 5-9. Eight common probes were significantly detected in both cell lines after 6 hours, 298 at 24 hours and 544 at 48 hours (Figure 5-10).

The top 10 probes differentially detected at each time point were identified and the genes that these probes target, the fold changes and the ranking at all time points are listed in Table 5-1 and Table 5-2 for SkMel28 and MeWo cells respectively. All probes in these tables relate to genes that demonstrated upregulation rather than downregulation after vitamin D_3 treatment. *CYP24A1*, *PRSS3*, *CLMN* and *PTGES* are common genes upregulated in both cell lines at least at 1 time point.

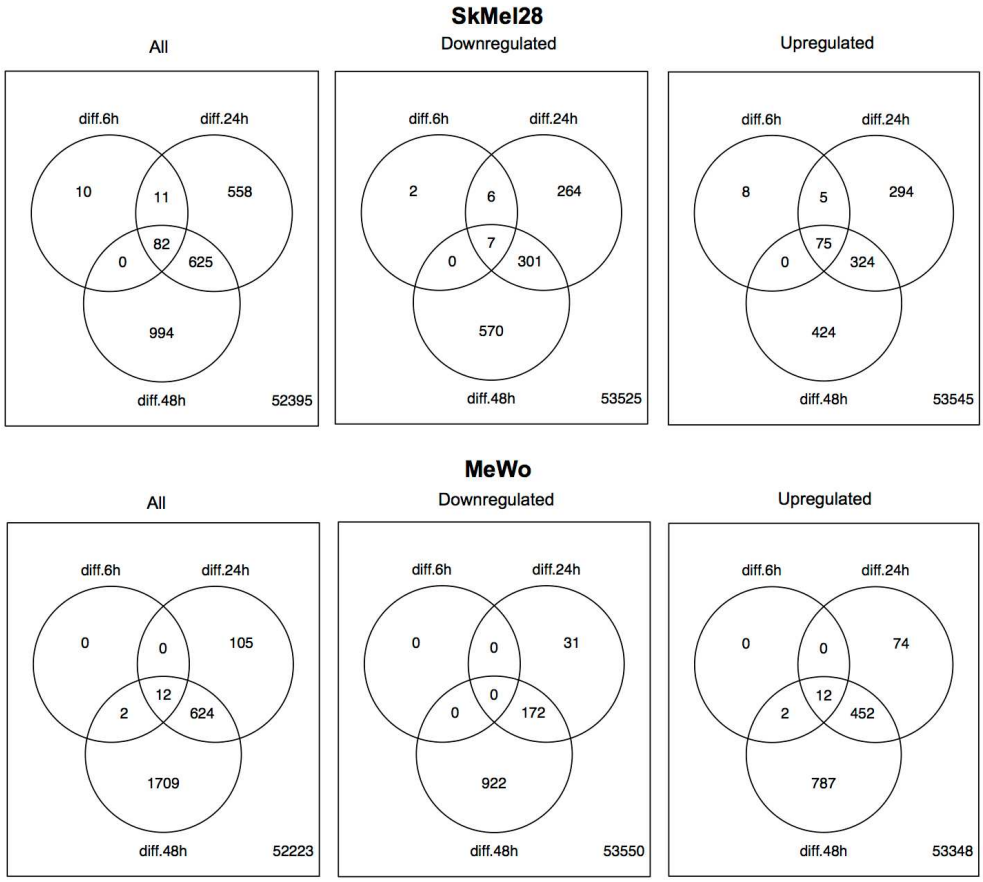


Figure 5-9: Probes differentially detected after vitamin D_3 treatment between time points for each cell line.

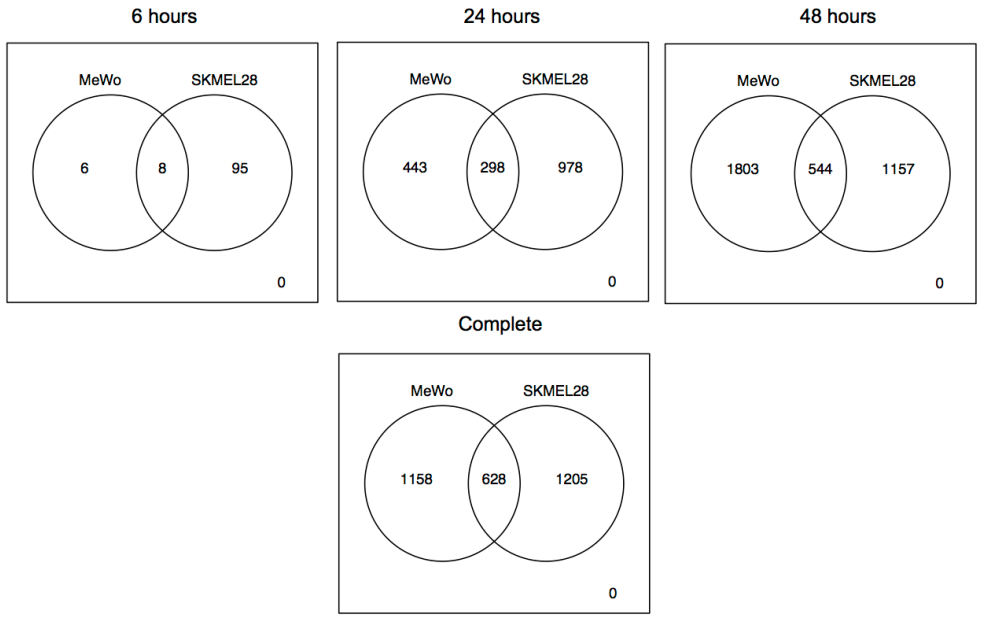


Figure 5-10: Overlapping probes differentially detected after vitamin D_3 treatment between MeWo and SkMel28 cells.

Table 5-1: The top 10 significant probes at each time point (targeting 17 genes) in SkMel28 treated cells compared to controls and their fold changes (FC) at all time points.

The number in superscript indicates the ranking at each time point. When a predicted VDRE is located in the gene promoter it is indicated in the corresponding column.

SkMel28				
Gene	**FC at 6h**	**FC at 24h**	**FC at 48h**	**VDRE**
CYP24A1	55.4 [1]	34.6 [2]	80.4 [1]	✓
LOC100127888	4.4 [2]	4.3 [12]	2.5 [127]	✓
KIAA0226L	4.0 [3]	3.1 [85]	3.6 [71]	✓
PRSS33	16.3 [4]	59.9 [6]	51.0 [10]	✓
TRPM1	4.0 [5]	4.53 [35]	3.1 [141]	✓
PI15	2.7 [6]	2.9 [33]	2.4 [116]	✓
SLCO4A1	3.8 [7]	3.3 [73]	-	✓
CLMN	2.4 [8]	3.0 [20]	2.6 [59]	✓
HEXIM1	2.7 [9]	1.8 [348]	-	✓
TPST1	2.9 [10]	2.6 [112]	2.8 [104]	✓
ATP6V0D2	-	6.1 [1]	6.8 [2]	✓
PTGES	3.8 [11]	15.9 [3]	11.0 [7]	✓
SULT1C2	4.5 [13]	22.7 [4]	43.3 [3]	✓
ATP6V0D2	-	6.9 [5]	6.4 [8]	✓
ADORA3	4.4 [16]	13.7 [7]	18.9 [6]	✓
MB	-	18.8 [8]	16.9 [11]	✓
C9orf135	-	7.9 [9]	6.3 [20]	✓
SULT1C2	-	29.6 [10]	57.4 [5]	✓
MFSD6	-	4.1 [11]	6.4 [4]	✓
PRSS33	-	50.5 [14]	82.0 [9]	✓

Table 5-2: The top 10 significant probes at 6, 24 and 48 hours (targeting 16 genes) in MeWo treated cells compared to controls and their fold changes (FC) at all time points.

The number in superscript indicates the ranking at each time point. When a predicted VDRE is located in the gene promoter it is indicated in the corresponding column.

	MeWo			
Gene	**FC at 6 h**	**FC at 24h**	**FC at 48h**	**VDRE**
CYP24A1	16.7 [1]	34.6 [6]	25.0 [19]	✓
EFTUD1	1.8 [2]	2.7 [10]	2.6 [15]	
DUSP10	1.7 [3]	1.4 [338]	1.4 [478]	✓
DUSP10	1.8 [4]	-	1.3 [1316]	✓
HEXIM1	2.3 [5]	2.1 [140]	1.5 [1252]	✓
MITF	1.8 [6]	1.5 [411]	1.4 [1225]	✓
HEXIM1	1.8 [7]	1.8 [130]	1.8 [202]	✓
NT5DC3	1.6 [8]	2.7 [12]	2.8 [8]	✓
CLMN	2.1 [9]	4.6 [9]	4.4 [14]	✓
MITF	1.8 [10]	-	1.4 [1142]	✓
STARD5	1.7 [14]	4.5 [1]	4.0 [6]	
PRSS33	-	10.9 [2]	43.0 [2]	✓
PRSS33	-	7.6 [3]	30.8 [1]	✓
CLMN	-	3.9 [4]	3.2 [17]	✓
PTGES	-	7.6 [5]	22.4 [3]	✓
MICAL3	-	3.5 [7]	3.4 [9]	✓
KLK6	-	2.9 [8]	2.4 [28]	✓
ST6GALNAC1	-	4.8 [17]	8.1 [4]	✓
SLC5A4	-	6.0 [21]	12.8 [5]	✓
SPP1	-	2.0 [203]	5.6 [7]	✓
TMEM164	-	2.4 [20]	3.0 [10]	

5.4.5 Prediction of VDR target genes

Computational analysis revealed that of the genes differentially expressed in SkMel28 cells as a response to vitamin D_3 treatment, 83.3% related to genes predicted to be direct VDR targets (930/1116 genes). Similarly, 83.5% of genes differentially expressed in MeWo cells were associated with VDR target genes (860/1030 genes). These data can be compared with an incidence of 77.5% of all annotated genes in the hg19 which were predicted to be direct VDR targets. This represents a statistically significant enrichment of VDR target genes in our data (Fisher's exact $p<0.001$ for both lines). All regions with predicted VDREs are listed in Appendix A.4.2.

5.4.6 Enrichment analysis using Metacore™

Enrichment analysis was performed using the list of significantly expressed genes at each time point to identify overrepresented GO processes and pathways. The top 5 overrepresented GO processes in response to vitamin D_3 for SkMel28 and MeWo cells are listed in Table 5-3 and Table 5-4 respectively. For SkMel28 cells, overrepresented biological processes such as cell adhesion and differentiation were among the top 5 at earlier time points, while processes involved in immune response such as IFN-γ, IL-4 and IL-10 regulation were among the top 5 at 48 hours. For MeWo cells, overrepresented biological processes such as cell proliferation, migration and regulation of transcription were among the top 5 at all time points. No distinguishable difference was observed in overrepresented GO processes between the 3 time points for MeWo cells. GO processes related to immune response were enriched in SkMel28 but not in MeWo cells when the top 5 overrepresented processes were compared. However, GO processes related to immune response such as IL-2, IL-3 regulation were also enriched in MeWo cells but were less significant (data not shown).

Table 5-3: The top 5 overrepresented GO processes in SkMel28 cells after 6, 24 and 48 hours of treatment with vitamin D_3 identified using MetaCore™.

	SkMel28-6 hours	FDR	Number of genes
1	Muscle cell fate determination	3.07×10^{-5}	3/3
2	Positive regulation of mucus secretion	5.32×10^{-4}	3/7
3	Negative regulation of ossification	8.11×10^{-4}	4/49
4	Regulation of norepinephrine secretion	8.11×10^{-4}	3/30
5	Histamine secretion by mast cell	8.11×10^{-4}	3/11
	SkMel28-24 hours	**FDR**	**Number of genes**
1	Cell adhesion	2.30×10^{-11}	77/981
2	Homophilic cell adhesion	4.79×10^{-08}	29/174
3	Osteoblast differentiation	6.80×10^{-05}	16/110
4	Neutrophil homeostasis	6.80×10^{-05}	6/11
5	Negative regulation of fibroblast growth factor receptor signalling pathway	1.80×10^{-04}	7/15
	SkMel28-48 hours	**FDR**	**Number of genes**
1	Antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	3.84×10^{-12}	17/137
2	Interferon-gamma-mediated signalling pathway	2.49×10^{-08}	25/94
3	Regulation of interleukin-4 production	6.10×10^{-08}	8/42
4	Regulation of interleukin-10 secretion	6.10×10^{-08}	8/14
5	Nervous system development	1.48×10^{-07}	71/2500

Table 5-4: The top 5 overrepresented GO processes in MeWo cells after 6, 24 and 48 hours of treatment with vitamin D_3 identified using MetaCore™.

	MeWo-6 hours	FDR	Number of genes
1	Regulation of interleukin-6 biosynthetic process	1.91×10^{-4}	2/19
2	Mammary gland epithelial cell differentiation	0.002	2/22
3	Mammary gland epithelial cell proliferation	0.002	2/28
4	Embryonic placenta development	0.005	2/126
5	Response to endoplasmic reticulum stress	0.005	2/152
	MeWo-24 hours	**FDR**	**Number of genes**
1	Negative regulation of transcription from RNA polymerase II promoter	1.95×10^{-09}	53/762
2	Positive regulation of transcription, DNA-dependent	3.49×10^{-08}	50/1381
3	Positive regulation of cell migration	1.97×10^{-07}	23/362
4	Negative regulation of ossification	7.09×10^{-07}	10/49
5	Intracellular protein kinase cascade	7.09×10^{-07}	18/521
	MeWo-48 hours	**FDR**	**Number of genes**
1	Outflow tract morphogenesis	7.24×10^{-11}	25/70
2	Regulation of cell proliferation	2.98×10^{-09}	46/1731
3	Negative regulation of cell proliferation	3.49×10^{-09}	84/792
4	Nervous system development	3.90×10^{-09}	87/2500
5	Positive regulation of transcription, DNA-dependent	3.90×10^{-09}	102/1381

The top 10 pathways that were overrepresented in response to vitamin D_3 for SkMel28 and MeWo cells are listed in Table 5-5 and Table 5-6 respectively. No pathways were overrepresented for the 6-hour time point for MeWo and SkMel28 cells using a threshold of FDR<0.1. Wnt and TGF-beta signalling pathways were enriched when both melanoma cell lines were treated with vitamin D_3. Interestingly, a VDR-related pathway was overrepresented at both time points in MeWo cells only. No overrepresentation of VDR signalling was observed in SkMel28 cells when the list of all enriched pathways was checked. The PGE2 signalling pathway was among the top 10 overrepresented at both time points in SkMel28. In MeWo cells it was overrepresented but less significantly (ranked 24^{th} at 24 hours and 35^{th} at 48 hours, data not shown). The NOTCH signalling pathway was overrepresented at both time points in MeWo cells but not overrepresented in SkMel28 cells at all.

Table 5-5: The top 10 overrepresented pathways in SkMel28 cells after 24 and 48 hours of treatment with vitamin D_3 identified using MetaCore™.

	SkMel28-24 hours	FDR	Number of genes
1	Development_WNT signalling pathway. Part 2	$8.97x10^{-4}$	10/53
2	Cell adhesion_Ephrin signalling	0.008	8/45
3	Cell adhesion_Chemokines and adhesion	0.018	11/100
4	Signal transduction_Calcium signalling	0.030	7/45
5	G-protein signalling_Regulation of RAC1 activity	0.047	6/36
6	Cell adhesion_ECM remodelling	0.050	7/52
7	G-protein signalling_Regulation of p38 and JNK signalling mediated by G-proteins	0.050	6/39
8	PGE2 pathways in cancer	0.050	7/55
9	G-protein signalling_RhoB regulation pathway	0.050	4/16
10	Cytoskeleton remodelling_TGF, WNT and cytoskeletal remodelling	0.057	10/111
	SkMel28-48 hours	**FDR**	**Number of genes**
1	Cell adhesion_ECM remodelling	$7.21x10^{-4}$	11/52
2	Development_TGF-beta-dependent induction of EMT via SMADs	$7.21x10^{-4}$	9/35
3	Cardiac Hypertrophy_NF-AT signalling in Cardiac Hypertrophy	$7.21x10^{-4}$	12/65
4	Cell adhesion_Chemokines and adhesion	$7.21x10^{-4}$	15/100
5	PGE2 pathways in cancer	$7.21x10^{-4}$	11/55
6	Development_Regulation of epithelial-to-mesenchymal transition (EMT)	0.002	11/64
7	Development_WNT signaling pathway. Part 2	0.002	10/53
8	DNA damage_ATM / ATR regulation of G2 / M checkpoint	0.003	7/26
9	Cell adhesion_Ephrin signalling	0.003	9/45
10	Immune response_Antigen presentation by MHC class II	0.003	5/12

Table 5-6: The top 10 overrepresented pathways in MeWo cells after 24 and 48 hours of treatment with vitamin D_3 identified using MetaCore™.

	MeWo-24 hours	FDR	Number of genes
1	Development_TGF-beta-dependent induction of EMT via SMADs	0.001	7/35
2	Development_Notch Signalling Pathway	0.003	7/43
3	Transcription_CREB pathway	0.003	7/47
4	Development_PIP3 signalling in cardiac myocytes	0.003	7/47
5	Development_TGF-beta receptor signalling	0.003	7/50
6	Cytoskeleton remodelling_TGF, WNT and cytoskeletal remodelling	0.003	10/111
7	Transcription_Role of VDR in regulation of genes involved in osteoporosis	0.009	7/61
8	Development_Thromboxane A2 pathway signalling	0.016	6/49
9	Cytoskeleton remodelling_Cytoskeleton remodelling	0.030	8/102
10	Regulation of lipid metabolism_Regulation of lipid metabolism via LXR, NF-Y and SREBP	0.030	5/38
	MeWo-48 hours	**FDR**	**Number of genes**
1	Signal transduction_PKA signalling	3.59×10^{-5}	14/51
2	Development_WNT signalling pathway. Part 2	3.59×10^{-5}	14/53
3	Development_PIP3 signalling in cardiac myocytes	3.59×10^{-5}	13/47
4	Development_Regulation of epithelial-to-mesenchymal transition (EMT)	3.59×10^{-5}	15/64
5	Development_TGF-beta-dependent induction of EMT via SMADs	4.96×10^{-5}	11/35
6	Development_TGF-beta-dependent induction of EMT via MAPK	0.001	11/47
7	Regulation of CFTR activity (norm and CF)	0.001	12/58
8	Transcription_Role of VDR in regulation of genes involved in osteoporosis	0.002	12/61
9	Development_Notch Signalling Pathway	0.002	10/43
10	Development_Hedgehog and PTH signalling pathways in bone and cartilage development	0.002	9/36

5.5 Discussion

Despite many clinical trials that have been conducted to test the effect of various adjuvant treatments for melanoma the only FDA-approved adjuvant regimen for patients with melanoma is interferon alpha IFN-α (Davar *et al*., 2012). This is used in the US and in Europe but not in the UK as the toxicity is thought to outweigh the small survival difference reported in meta-analyses (Mocellin *et al*., 2010). More recently the reports of survival benefit from targeted therapies in stage IV patients (anti-CTLA-4 and BRAFi) give hope that these regimens might also be useful in the adjuvant setting (Davar *et al*., 2012, Eggermont and Robert, 2011). As discussed in section 1.12.2, vitamin D has been studied as an anti-cancer agent both *in vitro* and *in vivo*, showing inhibition of proliferation and apoptosis in various cancer models (Deeb *et al*., 2007). In patients with melanoma, there are ongoing clinical trials which aim to study the effect of vitamin D as an adjuvant treatment alone or in combination with other regimens (U.S. National Institutes of Health, 2013). Promising results have been reported in phase II and a single phase III clinical trial that vitamin D might be useful as anti-cancer treatment for patients with prostate cancer (Deeb *et al*., 2007). It is therefore noteworthy that the Leeds group has reported an observational study in the Leeds Melanoma Cohort in which higher serum vitamin D levels at diagnosis were associated with lower Breslow thickness and better prognosis even when adjusted for thickness (Newton-Bishop *et al*., 2009). Observational studies are useful, but must be viewed with caution: associations do not necessarily prove causality. This is especially a concern for measurements of serum vitamin D, as levels are known to be higher in leaner, fitter people. The observations of Newton-Bishop *et al* might simply reflect vitamin D levels acting as a marker of healthier lifestyles. Furthermore, the optimal serum levels of vitamin D for cancer patients are unknown; the Leeds group has continued to explore the relationship between vitamin D levels in the blood and melanoma. Despite a wealth of data referring to the potential of vitamin D as an anti-cancer agent very few studies have explored the transcriptional changes in melanoma in response to vitamin D_3 (as reviewed in section 1.12.3) with a view to understanding the underlying biology.

A two-fold study was therefore designed: firstly to identify transcriptional patterns associated with a cellular response to vitamin D_3 in melanoma cell lines *in vitro*; and then secondly investigating to what extent such transcriptional profiles were reflected in primary melanomas from patients presenting with high serum vitamin

D levels and those patients who presented with low vitamin D serum levels. At this point the *in vivo* primary tumour results are not yet available but our intent ultimately is to determine if the gene expression patterns mirror those noted in the *in vitro* cell line model and to what degree these represent biological pathways reported to be related to signalling through the vitamin D receptor (Ramagopalan *et al*., 2010, Heikkinen *et al*., 2011). Here I discuss the data derived from the *in vitro* models.

5.5.1 Anti-proliferative effect of vitamin D_3

The anti-proliferative effect of vitamin D_3 has been well studied using various *in vitro* and *in vivo* cancer models (Deeb *et al*., 2007) including (as described in section 1.12.3) melanoma cell lines. Two previously identified vitamin D sensitive (MeWo, SkMel28) and two vitamin D resistant (MelJuso, SkMel5) melanoma cell lines were originally selected and tested to confirm their sensitivity to vitamin D using a proliferation assay. These initial experiments in the Leeds laboratory were unsuccessful: in particular, no effect was observed after addition of vitamin D_3 to the cells in the previously reported vitamin D sensitive cells using WST-1 or ^{3}H-thymidine proliferation assays (data not shown). I postulated that the addition of a fat soluble hormone (vitamin D) in ethanol (which was unstable) required optimisation. As a result, a collaboration with Professor Reichrath of Saarland University, in Homburg was established as he has worked extensively with *in vitro* melanoma models in response to added vitamin D_3 (Seifert *et al*., 2004, Reichrath *et al*., 2007). Consequently, I spent two months under his supervision in Homburg successfully performing the cell culture experiments. A significant divergence from my previous protocol was the stipulation that 1% BSA was added to the vitamin D_3 treated cell culture: this appears to reduce vitamin D_3 adherence to the plastic wells but also aids the cellular uptake of vitamin D_3.

The proliferation assays were purely performed to confirm the sensitivity of the melanoma cells to vitamin D_3 prior to the gene expression experiments. Therefore, these experiments were designed and executed based on the standard proliferation assay utilised by the Reichrath laboratory without further optimisation. These assays are designed to incorporate at least 7 days of growth and so require initial seeding densities that are very low. As a result, growth curves showed that the cells were still within a lag phase during the first days of the experiments. After 144 hours (7 days) the cells were sufficiently confluent to undergo exponential growth at which point a difference in proliferation could be readily observed both statistically and visually when vitamin D_3 treated MeWo

and SkMel28 cells were compared to controls confirming previously published data (Seifert *et al*., 2004). Significantly, no reduction was observed in the growth rate of vitamin D_3 treated 'MelJuso' cells when compared to controls with a slight stimulatory effect apparent, although this failed to prove statistically significant (see Figure 5-3 and Figure 5-4). The data on the SkMel5 cell line are not presented as they were reported to be SkMel28 after cell line authentication and it is this issue I discuss next.

5.5.2 Cell line authentication

Much emphasis has been placed upon the importance of cell line authentication to be implemented as a routine practise in experimental work in recent years, especially in the case of cell lines derived from other than the original source (American Type Culture Collection Standards Development Organization Workgroup, 2010). Therefore, DNA was extracted from the four original melanoma cell lines assayed in Homburg and once back in Leeds cell line authentication was performed. The MeWo and SkMel28 cell lines were correctly authenticated. Alarmingly, one cell line assayed in Homburg and referred to as the vitamin D_3 resistant cell line SkMel5 (as reported by Seifert *et al*., 2004), was actually shown to be the sensitive SkMel28 after cell line authentication, thereby explaining the growth retardation effect observed after treatment (data not shown). Even more worrying was the discovery that the cell line reputed to be MelJuso could not be authenticated, as the DNA profile did not match any known cell lines and therefore they had to be considered of unknown origin. On request, immunocytochemistry was performed in Homburg to demonstrate the expression of melanocytic markers in the 'MelJuso' cells as a further means of demonstrating their authenticity. The melanocytic markers tested are routinely used in the clinic to identify the presence of metastasised melanoma cells in the excised lymph nodes (Marsden *et al*., 2010); being robust and highly specific they do not therefore require negative controls. The 'MelJuso' cells expressed all the melanocytic markers tested and therefore were considered of melanocytic origin.

5.5.3 VDR expression using ICC

The classical pathway via which vitamin D_3 is reported to act is through binding to the VDR-RXR complex and regulation of transcription though binding to target genes which contain VDREs in their promoters (Deeb *et al*., 2007). This is in

contrast to the non-canonical pathway via which vitamin D_3 is speculated to act through a membrane VDR or cytosolic VDR which promotes Ca^{2+} influx and further activation of the protein kinase C (PKC) pathway. In addition, vitamin D_3 is also speculated to act through binding to G-protein coupled receptors (Deeb *et al*., 2007) which in turn activate phospholipase C gamma (PLCγ), Ras, PI3K and protein kinase A (PKA) pathways to induce MAPK signalling which itself might introduce cross-talk with the classical VDR pathway to promote transcriptional changes (Deeb *et al*., 2007).

In melanoma, it has been reported that VDR expression is lost in advanced stage tumours (Brozyna *et al*., 2011). However, it is unknown whether lower VDR expression in advanced stage tumours is due to low serum vitamin D levels or a defective VDR. Furthermore it has been reported though, that the anti-proliferative effect of vitamin D is still active even when VDR is knocked out in vitamin D sensitive breast cancer cells (Costa *et al*., 2009), which suggests that vitamin D_3 might be able to act differently in tumours with low or no VDR expression via a non-canonical pathway. Both MeWo and SkMel28 cells clearly expressed VDR, but intriguingly, so too did 'MelJuso'. This suggests a further level of complexity in which, despite the presence of VDR within a melanoma cell, vitamin D_3 fails to induce a growth inhibitory effect. This is not surprising as tumour cells may well have molecular changes downstream of the VDR. A high percentage of SkMel28 and 'MelJuso' cells grown on the slides for ICC were rounded up or demonstrated reduced cell adherence and therefore I believe that this experiment should be repeated to draw firm conclusions.

5.5.4 Gene expression work

Previous studies have explored the transcriptional changes in melanoma cells in response to vitamin D_3 using either a candidate gene approach or studying a specific mechanism elicited by vitamin D_3 and these were reviewed in section 1.12.3.

Whole-genome microarray studies of various cancer cells have been conducted and genes involved in cell proliferation, apoptosis, angiogenesis and immune response were reported to be differentially expressed in vitamin D_3 treated cells compared to controls (Deeb *et al*., 2007, Fleet *et al*., 2012).

Here, I report for the first time a microarray approach to identify transcriptional changes specifically in vitamin D sensitive melanoma cells (SkMel28, MeWo) treated with vitamin D_3. Three different time points (6, 24 and 48 hours) were

selected to identify both early and late changes in response to the hormone. The data reported here demonstrate clear transcriptional changes between treated and control cells (Figure 5-8). A progressive response to vitamin D_3 was noted with an increase in transcriptional changes observed through time in both cell lines. Predictably, some genes underwent transcriptional upregulation in both cell lines, but significant differences were also observed, potentially reflecting varying tumour biology (as might be expected in tumour cell lines). *CYP24A1* (Cytochrome P450, member 24A1) encoding for the vitamin D_3 degradation protein was the top upregulated gene at 6 hours in both cell lines and an increase of its expression was observed after 24 hours in treated MeWo cells and after 48 hours in treated SkMel28 cells. It was observed that *VDR* was overexpressed at later time points (at 48 hours in SkMel28 and both at 24 and 48 hours in MeWo treated cells). I would suggest that the upregulation of these genes provides evidence to support the view that the vitamin D_3 treatment experiment had induced specific transcriptional changes in the tumour cells. Particularly, *CYP24A1* has been reported to be the most significantly upregulated gene in response to vitamin D_3 treatment in most microarray studies using various cancer cell lines as reviewed by Kriebitzsch *et al* (Kriebitzsch *et al*., 2009).

Examination of the top most differentially expressed genes (presented in Table 5-1 and Table 5-2) revealed four genes (in addition to *CYP24A1*) that were commonly upregulated in both cell lines at least at one time point. These were *PRSS3* (Protease serine 3), *CLMN* (Calmin), *HEXIM1* (Hexamethylene bis-acetamide inducible 1) and *PTGES* (Prostaglandin E synthase). *PRSS3* is a serine protease currently with no published direct link to vitamin D_3. However, vitamin D_3 treatment revealed a large induction of this gene at all time points (FC>40 at 48hours in both lines) and its function is therefore discussed. The induction of several proteases has been linked to tumour progression and metastasis until recently, where studies revealed that proteases also have tumour suppressive effects (Lopez-Otin and Matrisian, 2007). For example, a protease known as seprase was found to suppress melanoma tumourigenicity through regulation of cell proliferation and survival (Ramirez-Montagut *et al*., 2004). PRSS3 has been reported to be silenced in oesophageal and gastric cancer (Lopez-Otin and Matrisian, 2007), which suggests a putative tumour suppressive role of PRSS3 induced by vitamin D_3.

CLMN, a retinoic acid-regulated gene, has been previously identified as upregulated in response to added vitamin D_3 in breast cancer cultures (Milani *et*

al., 2013). It was recently reported that *CLMN* ectopic overexpression in murine neuroblastoma cells plays a role in G_1/S cell cycle exit which supports its involvement in the anti-proliferative effect of vitamin D_3 (Marzinke and Clagett-Dame, 2012).

HEXIM1, is mostly known for its role in inhibition of RNA polymerase II through the positive transcription elongation factor b (p-TEFb) but it has also been implicated to play a role in inhibition of metastasis though effects on cell invasion and angiogenesis (Ketchart *et al*., 2013) and promotion of apoptosis through positive regulation of p53 (Lew *et al*., 2012). As a result *HEMIX1* might be involved in the pro-apoptotic and anti-angiogenic effects of vitamin D_3.

Finally, *PTGES* is involved in the synthesis of prostaglandin E_2. It has been reported that vitamin D_3 inhibits the prostaglandin signalling pathway through downregulation of prostaglandin E_2 which is in accordance with the anti-cancer effects of vitamin D_3 as prostaglandin signalling stimulates cancer cell growth and progression (Fleet *et al*., 2012). In our data, *PTGES* was overexpressed in response to vitamin D_3, which might be explained by the reported involvement of PTGES in the promotion of apoptosis after induction of its expression by p53 (Faour *et al*., 2006).

Within the top ten differentially expressed genes in SkMel28, *TRPM1* (Transient receptor potential cation channel subfamily member 1) and *ADORA3* (Adenosine A3 receptor) were observed at all 3 time points. Both genes were also upregulated in MeWo cells after 24 and 48 hours of treatment but they were not ranked within the top 10. *TRPM1,* otherwise called *Melastatin* is reported to be associated with numerous cellular and homeostatic functions but it has also been reported as a marker in melanoma (Guo *et al*., 2012) in which its expression correlates well with melanocytic terminal differentiation. Significantly, loss of *TRMP1* expression is a poor prognostic marker and is associated with metastatic potential. *TRPM1* generates two transcripts: exonic *TRPM1* codes for the protein whilst intron 6 codes for mir-211 which has been linked to tumour suppression (Guo *et al*., 2012). It is intriguing to speculate that the observed correlation between increased tumour thickness and low serum vitamin D_3 at presentation (Newton-Bishop *et al*., 2009), may reflect a direct tumour response in which vitamin D_3 inducible genes such as *TRPM1* are responsible for increased proliferation. *ADORA3* is a G-coupled transmembrane protein and when activated induces intracellular signalling. It has been reported that its stimulation through agonists in cancer cells suppresses cell proliferation (Stagg and Smyth,

2010). In melanoma, it has been reported that treatment of the melanoma cell line B16-F10 with an ADORA3 agonist resulted in b-catenin ubiquitination causing downregulation of c-Myc and cyclin D1 transcription and thereby suppressing tumour growth (Fishman *et al*., 2002).

MITF (micropthalmia transcription factor), the master regulator of the melanocytic lineage, was upregulated after exposure to vitamin D_3, although only in MeWo cells. MITF has various roles including cell cycle regulation, differentiation and apoptosis (Levy *et al*., 2006). It has been reported that MITF overexpression in melanoma promotes its anti-proliferative effects and that its expression is suppressed by *BRAF* mutation to overcome its growth-inhibitory effects (Wellbrock and Marais, 2005). It is noteworthy that MeWo cells are *BRAF* wild type and SkMel28 are *BRAF* mutant. It seems feasible that MITF is not overexpressed in the BRAF mutant SkMel28 cells because its expression is controlled through oncogenic MAPK signalling.

5.5.4.1 Prediction of VDR target genes using an in silico approach

A computational approach was followed to identify how many genes significantly expressed in response to vitamin D_3 were direct VDR targets. A statistically significant enrichment of VDR target genes was observed in our data compared to the predicted VDR targets in all genes annotated in human reference genome version hg19. This approach was limited to predict genes the promoters of which contained the classical DR3-type VDREs using the JASPAR consensus sequences of the VDR-RXR complex as described in section 5.3.1.3.1. Recent ChIP-sequencing data in human monocytes have identified a clear shift of VDR binding from non-DR3-type sites to DR3 type sites when the cells are stimulated with vitamin D_3 (Heikkinen *et al*., 2011). However, 10% of the genes identified to contain a VDRE after stimulation with vitamin D_3 did not contain a DR3-type site (Heikkinen *et al*., 2011). *EFTUD1* (elongation factor Tu GTP binding domain containing 1*)*, which was among the top 10 upregulated genes in MeWo cells was predicted not to contain a VDRE by our approach. However, this gene was reported to possess a non-DR3 type VDRE by Heikkinen *et al* (Heikkinen *et al*., 2011), which suggests that it is indeed a VDR target gene. This suggests that in the future, the identification of putative VDREs can be performed more accurately using the ENCODE database which includes data from ChIP-sequencing experiments (Consortium *et al*., 2012). The two recent ChIP-sequencing studies by Ramagopalan *et al* and Heikkinen *et al* reported only

18.2% common VDREs in vitamin D_3 stimulated cells (Ramagopalan *et al*., 2010, Heikkinen *et al*., 2011). This might reflect the different cut-off values used to determine significant VDREs and the different duration of vitamin D_3 treatment but it might also suggest that VDR utilises different binding sites in different cells (Heikkinen *et al*., 2011). As a result, further investigation of the vitamin D_3 transcriptional response is required within the context of ChIP-sequencing data to identify VDR target genes in melanoma cells.

5.5.5 Pathway analysis

Pathway analysis has helped to categorise the biological functions of genes significantly differentially expressed in response to vitamin D_3. Pathway analysis performed using MetaCore™ highlighted the overrepresentation of GO processes, such as cell proliferation, migration and adhesion and differentiation in cells in response to vitamin D_3. Overrepresented pathways included Wnt and TGF-beta signalling in both cell lines. Vitamin D_3 has previously been reported to alter both pathways (Deeb *et al*., 2007, Fleet *et al*., 2012). In particular, it has been reported that Wnt signalling is inhibited through induction of *CDH1* (E-cadherin) which promotes translocation of b-catenin from the nucleus to the plasma membrane, inhibiting b-catenin-TCF4 signalling pathway which induces cell proliferation (Deeb *et al*., 2007). It was noted that NOTCH signalling was overrepresented in vitamin D_3 treated MeWo cells only which suggests that vitamin D_3 acts in different ways in different tumours. This pathway has been reported to play a role in the vitamin D_3-mediated anti-tumour effects in a prostate cancer cell line model (Kovalenko *et al*., 2010).

In summary, identification of the genes most significantly overexpressed in response to vitamin D_3, their epigenetic control elements and the pathways associated with these genes offers credible support for the view that vitamin D_3 has anti-tumour effects. Future work should include validation of these results and in particular detailed examination of the pending patient tumour data set. It is vital to explore whether transcriptional changes identified *in vitro* are mirrored *in vivo*, and to what extent these effects are moderated by complex tumour biology.

Chapter 6
Osteopontin plasma concentrations in cutaneous melanoma

6.1 Aims

This chapter describes work carried out to assess plasma osteopontin levels as a potential prognostic biomarker in melanoma.

The aims are:

- To evaluate the suitability of stored EDTA plasma samples from the Leeds Melanoma Cohort to measure osteopontin concentrations using a commercial immunoassay (R&D systems);
- To evaluate the performance of the osteopontin assay using stored EDTA plasma samples;
- To assess the difference in osteopontin plasma concentrations between patients with melanoma and healthy controls;
- To correlate osteopontin plasma concentrations with tumour characteristics;
- To identify the prognostic value of osteopontin plasma concentrations.

6.2 Background

AJCC staging is the gold standard to assess prognosis in melanoma incorporating tumour characteristics such as Breslow thickness and ulceration, lymphatic metastasis, distant metastasis and lactic dehydrogenase (LDH) serum level (Balch *et al*., 2009). Features, such as age, sex and tumour site reviewed by (Homsi *et al*., 2005) have also shown prognostic significance but are not included in the AJCC staging system. Although the predictive value of the AJCC staging system is improving, there is still variance in outcome that cannot be explained with the current prognostic factors (Ross, 2006). There is a need therefore to improve available prognostic biomarkers.

Circulating markers tested so far for their prognostic significance or use in disease monitoring have proven of little clinical utility in early stage patients (Hofmann *et al*., 2009, Paschen *et al*., 2009). Potential serologic markers that have been investigated so far in melanoma were discussed in the introduction (section 1.10). In this chapter, I report an assessment of the prognostic and disease monitoring utility of plasma osteopontin levels in patients with melanoma. Increased osteopontin gene expression levels in primary melanomas were reported by the Leeds group to be of independent prognostic significance (Conway *et al*., 2009) using an agnostic approach: the DASL platform. We further wanted to investigate if plasma circulating osteopontin levels were of similar prognostic value and/or might detect asymptomatic disease recurrence.

Osteopontin, coded by the secreted phosphoprotein 1 (*SPP1*) gene, was first described as a secreted glycoprotein overexpressed after malignant cell transformation (Senger *et al*., 1983, Senger *et al*., 1989) and characterised as a bone matrix protein (Franzen and Heinegard, 1985). Osteopontin belongs to the small integrin-binding ligand N-linked glycoproteins (SIBLINGs) family (Fisher *et al*., 2001). SIBLINGs, which are secreted proteins, promote cell adhesion, motility and survival through binding to integrins, CD44 and metalloproteinases (Figure 6-1) (Bellahcene *et al*., 2008).

As a result of its involvement in these pathways, osteopontin has been extensively studied in cancer biology as a potential biomarker or a therapeutic target. It is expressed in different cell lineages and has been described as important in tumour progression and metastasis in many cancers (El-Tanani *et al*., 2006) including melanoma. In 2005, *SPP1* gene expression was reported to be greater in vertical growth phase primary and metastatic melanomas than in naevi and melanomas *in situ* (Smith *et al*., 2005) which suggested that the glycoprotein might be associated with disease progression. This association was also suggested by three immunohistochemical studies (Rangel *et al*., 2008, Zhou *et al*., 2005, Maier *et al*., 2011). Our group confirmed increased expression of *SPP1* in primary tumours to be of independent prognostic value for melanoma (Conway *et al*., 2009) indeed the gene whose expression was most strongly associated with outcome. Other groups reported supportive evidence using different gene expression assays (Jaeger *et al*., 2007, Soikkeli *et al*., 2010) or immunohistochemistry (Rangel *et al*., 2008, Kashani-Sabet *et al*., 2009, Alonso *et al*., 2007).

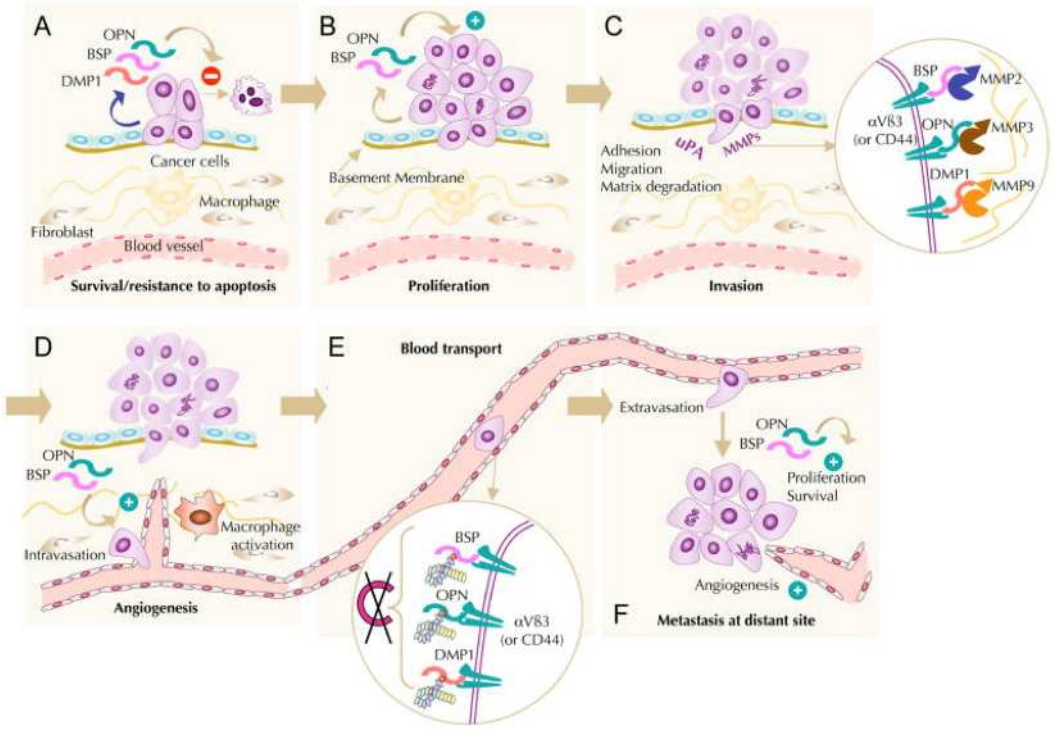


Figure 6-1: The role of osteopontin and other SIBLING proteins in cell adhesion and metastasis.

A, B) At the primary site, SIBLING proteins are secreted by cancer cells which promote their proliferation and survival, C) Cancer cells can detach from the primary site and invade the stroma. Osteopontin enhances cell motility and invasion into the surrounding tissue. The insert shows SIBLING proteins bound to their receptors (osteopontin binds to αvβ3 integrins and/or CD44). Local proteolysis is mediated through binding to specific metalloproteinases (MMPs). D) SIBLINGs promote the migration and adhesion of activated endothelial cells, E) The expression of SIBLINGs on the surface of cancer cells protects themselves from complement-mediated lysis. F) At distant sites, SIBLINGs proteins are secreted by cancer cells and promote their proliferation and survival as at the primary site. Reprinted by permission from Macmillan Publishers Ltd: [Nature Reviews Cancer] (Bellahcene *et al.*, 2008), copyright (2008).

Increased blood levels of osteopontin have been reported as associated with progression in many cancers (Rittling and Chambers, 2004), including breast cancer (Rodrigues *et al*., 2007), head and neck cancer (Li *et al*., 2011, Snitcovsky *et al*., 2009), pancreatic ductal adenocarcinoma (PDA) (Sullivan *et al*., 2009), hepatocellular cancer (Sun *et al*., 2010) and endometrial cancer (Cho *et al*., 2009). An increased osteopontin level was reported as a predictor of outcome in non-small cell lung cancer (NSCLC) (Mack *et al*., 2008, Isa *et al*., 2009) and to be reduced after tumour resection of NSCLC (Blasberg *et al*., 2010). No association was found between osteopontin levels and disease course in mesothelioma (Wheatley-Price *et al*., 2010). A small number of studies have reported increased osteopontin plasma levels in uveal melanoma (Reiniger *et al*., 2007, Haritoglou *et al*., 2009) which were highly correlated with the presence of liver metastasis (Kadkol *et al*., 2006).

Elevated osteopontin plasma concentrations have recently been reported in two studies in metastatic melanoma (Kluger *et al*., 2011, Maier *et al*., 2011). In both, comparison was made between samples from patients with and without concurrent metastatic disease.

Here, I report a pilot study of 185 patients with melanoma recruited to the Leeds Melanoma Cohort in which the prognostic value of osteopontin was tested. This is the first study looking at osteopontin plasma level as a prognostic marker in early stage disease patients (AJCC I to III) taking into account factors already known to be of prognostic value. The work was done collaboratively with Professor Rosamonde Banks at the University of Leeds. The Human Osteopontin Immunoassay kit (R&D systems), described in detail in section 2.11, was used to detect osteopontin levels in stored plasma samples from the Leeds Melanoma Cohort. Prior to use, the assay was evaluated based on guidelines of immunoassay development (DeSilva *et al*., 2003) and stored plasma samples were assessed for their suitability to measure osteopontin concentrations accurately using the immunoassay.

6.3 Detailed methodology

The Leeds Melanoma Cohort is a large population based cohort of melanoma patients recruited since 2000. Samples have been collected from participants but a proportion have been mailed to the laboratory. In order to limit variation between samples, all samples were therefore processed in the laboratory after a

delay of 24 to 72 hours: that is samples obtained in Leeds which were not mailed were left on the bench before processing resulting in variation in the time from venepuncture to processing with a median of 1 day (range, 0-4 days). There are no published data for the stability of osteopontin plasma levels in stored samples. Therefore, to investigate the potential impact of delays in processing we first investigated the stability of osteopontin levels over time. In order to do this, additional blood samples were obtained as described in section 2.2.3. In these samples osteopontin was measured to determine if there was change due to delays in processing. EDTA plasma samples were used in the assay as proteolytic cleavage of osteopontin by thrombin during the clotting process occurs in serum samples (Ayache *et al*., 2006). To confirm this, osteopontin concentration was measured in matched plasma and serum samples from patients with renal cell carcinoma.

A total of 185 patients were identified from participants bled at recruitment to the Leeds Melanoma Cohort for whom stored plasma samples were available (Newton-Bishop *et al*., 2009). Participants were recruited to the study within 3-6 months after diagnosis, when possible (Newton-Bishop *et al*., 2009). Samples were selected as follows: a) 76 samples from participants who were believed to be disease-free at venepuncture (53 treated stage I/II, 23 treated stage III), and who have not relapsed in the subsequent period of a median of 7.5 years (range, 1.1-11.2); b) 82 from participants who were believed to be disease-free at sampling but subsequently relapsed (57 treated stage I/II, 25 treated stage III); and c) 27 who had metastatic disease at sampling (17 untreated stage III, 10 untreated stage IV). A patient was defined as disease-free if they had had their primary excised or their lymph nodes removed and there was no known clinical evidence of further disease. A minimum period of 6 weeks between surgery and venepuncture was used based on a study in NSCLC patients which showed that osteopontin plasma levels were elevated in the period 6 weeks after surgery, possibly due to the involvement of osteopontin in wound healing (Blasberg *et al*., 2010). Thirty healthy controls were also included in the study to compare osteopontin levels with those in the normal population. No difference in age and gender was observed between controls and cases.

6.3.1 Enzyme-linked immunosorbent assay (ELISA) validation

A commercial ELISA assay kit (Quantikine; R&D Systems) was used to measure osteopontin levels as described in section 2.11. This kit was selected in conjunction with Professor Rosamonde Banks based on a study that showed it

has the highest sensitivity and specificity between other commercially available kits (Plumer *et al.*, 2008). Prior to use, the assay was validated examining intra- and inter-assay precision, parallelism and recovery using recombinant osteopontin protein purchased from Abcam, and interference as previously described (Wind *et al.*, 2011, Sim *et al.*, 2012). Intra-assay precision was assessed by running three quality control samples five times in one plate while inter-assay precision was assessed when two quality control samples were run in duplicate across nine different plates. The parallelism test (or otherwise demonstration of dilutional linearity) is critical to identify if any components of the sample interfere in the ability of the assay to detect the analyte (interfering matrix effects). Here, for the parallelism test, serial dilutions of 3 samples were run in duplicate. The recovery test also determines if anything in the matrix interferes with the analyte determination. Here, for the recovery test, a low, high and a blank spike were introduced into plasma samples and were run in duplicate, except for the blank spike into the plasma samples which was run thrice in duplicate across the plate to give a more accurate 'true value' of how much protein was spiked into the samples. Interference is the effect of a substance present in the sample that might affect the accurate determination of the analyte. No interference was found from the following substances tested: bilirubin, haemolysis, triglycerides or rheumatoid factor (data not shown) (Sim *et al.*, 2012).

Samples passed the intra- and inter- assay precision test, and parallelism test when coefficient of variation (CV)<15% (Food and Drug Administration, 2001). The recovery of the recombinant protein should be within 80-120% (Food and Drug Administration, 2001).

6.3.2 Statistical analyses

All samples were assayed in duplicate and the osteopontin concentrations (ng/ml) presented here are the mean of the two replicates. The differences in osteopontin levels between matched EDTA plasma and serum from four patients with renal cell carcinoma (RCC) were tested using the Wilcoxon matched-pairs signed-ranks test. Also, the potential effects of differences in sample processing time on osteopontin plasma concentrations were assessed using the results from the matched samples processed at different time points (same day versus four days after venepuncture) and analysed using the Wilcoxon matched-pairs signed-ranks test.

Following the successful validation assay, 215 plasma samples were screened. First, osteopontin differences between healthy controls and cases grouped according to AJCC stage at venepuncture were compared using the Kruskal-Wallis test and multiple linear regression. Second, we looked at osteopontin as a prognostic indicator for patients with no evident disease so that patients with metastatic disease at sampling were excluded from subsequent analysis to test for the prognostic significance/disease monitoring of osteopontin in apparently disease-free early stage patients.

Factors previously shown to be associated with survival in the Leeds Melanoma Cohort were assessed for association with osteopontin levels: Breslow thickness; body mass index (BMI); AJCC stage; tumour site and mitotic rate; age at diagnosis; sex; tumour ulceration (ulcerated, not ulcerated); sentinel node biopsy (SNB) status; and vitamin D serum levels (nmol/l, adjusted for season) (Newton-Bishop *et al*., 2009), using Mann-Whitney U tests, Spearman Correlations and Kruskal-Wallis tests where appropriate. Multiple Linear regression was used to identify possible independent predictors of osteopontin level.

Odds ratios (OR) and 95% Confidence Intervals (CI) were estimated from logistic regression models for the effect of osteopontin levels on risk of death from melanoma and death from any causes. Due to the skewed frequency distribution of osteopontin levels, log-transformed osteopontin level (to the base 2) was entered into the models so that the estimated OR would be interpreted as the OR associated with a doubling of osteopontin level at recruitment. Osteopontin level was also considered as a categorical variable by grouping into approximate tertiles (≤49.35, >49.35-≤64.34, >64.34). Both unadjusted and adjusted ORs were calculated for osteopontin; the adjustment variables were: age; sex; BMI; site of the primary; season-adjusted vitamin D level; and stage at sampling. Here, the vitamin D variable was grouped into six categories, to show the effect based on 20nmol/l increments (≤20, >20-≤40, >40-≤60, >60-≤80, >80-≤100, >100). Secondary analyses incorporated time-to-event data and Kaplan-Meier curves were plotted. Hazard ratios (HR) and 95% CI were estimated from Cox proportional hazards models for the effect of osteopontin level on melanoma specific survival (MSS) and overall survival (OS). OS was defined as the period between recruitment to the Leeds Melanoma Cohort and death, and MSS as the period between recruitment and death from melanoma. The median follow-up time was 4.3 years (range 0.6-10.7 years). An arbitrary significance level of $p<0.05$ was used. STATA version 10 (StataCorp., 2007) was used for statistical analyses.

6.4 Results

6.4.1 Validation assays

6.4.1.1 A comparison of levels in EDTA plasma versus serum

Stored matched EDTA plasma and serum samples from five patients with renal cell carcinoma were used to assess the difference in osteopontin levels between the two sample types. A significant difference in osteopontin concentrations between plasma and serum samples was observed (CV>20%) with serum samples having lower concentrations as expected (Table 6-1). Wilcoxon matched-pairs signed ranks test showed no statistical significance (p=0.12), but this could be explained due to small sample size. It was noted that osteopontin levels in serum were in the region of 50% lower than those measured in plasma, but only plasma samples were used to further validate the assay.

Table 6-1: Osteopontin concentrations in matched plasma and serum.

	Patient 1	***Patient 2***	***Patient 3***	***Patient 4***
Plasma Concentration (ng/mL)	45.11	71.79	36.73	286.65
Serum Concentration (ng/mL)	21.45	31.30	18.75	137.03
% CV	**52.5**	**56.4**	**48.9**	**52.2**

6.4.1.2 Sample processing time

Plasma samples processed immediately and samples processed four days after venepuncture from 4 healthy controls and 5 patients with melanoma were used to test for an effect due to delays in processing (Table 6-2). No significant difference in levels was observed between them (CV<20% and p=0.18, based on Wilcoxon matched-pairs signed ranks test). The samples stored from the Leeds Melanoma Cohort were therefore judged suitable for use in this evaluation of osteopontin as a disease marker for melanoma.

Table 6-2: Osteopontin plasma concentrations in day 0 versus day 4 processed samples.

	Normal 1	*Normal 2*	*Normal 3*	*Normal 4*	*Patient 1*	*Patient 2*	*Patient 3*	*Patient 4*	*Patient 5*
Day 0: Osteopontin level (ng/ml)	51.10	64.55	69.58	45.11	37.95	30.93	233.65	109.96	68.35
Day 4: Osteopontin level (ng/ml)	50.48	63.49	65.11	42.38	37.06	30.45	236.08	96.76	68.54
%CV	**1.2**	**1.6**	**6.4**	**6.1**	**2.3**	**1.5**	**1.0**	**12.0**	**0.3**

6.4.1.3 Parallelism test

The parallelism test was determined in three patients' samples processed after 4 days (Table 6-3). Serial dilutions of each sample were used to check for interfering matrix effects. All samples passed the test with CV<15%. However, the 1:12.5 dilution of sample 3 was above the upper limits of the assay range and therefore the concentration could not be calculated.

Table 6-3: Parallelism test for the serial dilutions of three samples

Sample	***Back calculated OPN concentration from each dilution (ng/ml)***				***%CV***
	1:12.5	**1:25**	**1:50**	**1:100**	
Sample 1	34.5	35.9	41.1	45.6	**12.9**
Sample 2	28.1	29.6	33.3	34.9	**10**
Sample 3	-	230.3	248.4	272.5	**8.4**

6.4.1.4 Recovery test

In the recovery test, low and high spikes were introduced into sample diluent and three sets of duplicates were run. The mean of the low spike in buffer was 8.7ng/ml, and the high spike in buffer was 15.6ng/ml. The recovery test was run twice using a recombinant protein purchased from R&D but it failed (data not shown). There was a concern that this was due to the form of protein used for recovery studies, which was tagged. As a result, a repeat of this assay was

performed using a different source of recombinant protein spike purchased from Abcam. The results of the final recovery test are shown in Table 6-4. Although one sample failed, this was due to a replicate. If the second replicate is removed, the recovery is 111%. All other samples passed the test.

Table 6-4: Recovery test

Sample	*Sample1*	*Sample 2*	*Sample 3*
Low spike concentration (ng/ml)	9.96	14.02	11.73
Low spike Recovery (%)	102.3	124.8	104.9
Pass / Fail	**Pass**	**Fail**	**Pass**
High spike concentration (ng/ml)	15.56	18.11	19.68
High spike Recovery (%)	93.1	96.0	109.6
Pass / Fail	**Pass**	**Pass**	**Pass**

6.4.1.5 Intra and inter-assay precision

The results of intra-assay precision are shown in Table 6-5 where the %CV between the 5 replicates of each QC sample was always <15%. The results of inter-assay precision is shown in Table 6-6 where the CV for each QC sample run across 8 or 9 different assays was always <15%.

Table 6-5: Intra-assay precision measured using 3 QC samples run 5 times on a single assay.

Assay	Low QC sample	Mid QC sample	High QC sample
1	1.88	4.78	9.61
1	1.76	5.00	10.00
1	1.91	5.02	9.46
1	1.82	5.07	9.49
1	1.82	4.79	9.29
Mean	1.84	4.93	9.57
SD	0.06	0.14	0.27
Intra Precision (%CV)	**3.1**	**2.8**	**2.8**

Table 6-6: Inter-assay precision measured using 1 QC sample measured across 8 assays and 1 QC sample across 9 assays.

Assay	Low QC sample	High QC sample
1	2.19	-
2	1.84	9.57
3	1.91	9.25
4	1.92	9.75
5	1.9	9.86
6	1.97	10.69
7	1.78	9.05
8	1.78	9.61
9	1.76	9.75
Mean	1.89	9.69
SD	0.13	0.49
Inter Precision (%CV)	**7.1**	**5.0**

6.4.2 Cross-sectional analysis of plasma osteopontin in all cases and healthy controls

The normal range of osteopontin level, as it was measured in 30 healthy controls, was 28.6-118.8ng/ml with a median level of 59.2ng/ml. Of the 185 melanoma patients 158 had treated stage I-III, 17 had untreated stage III and 10 untreated stage IV disease (median osteopontin levels 54.7ng/ml, 54.6ng/ml, and 74.0ng/ml respectively) (Figure 6-2). A statistically significant difference in osteopontin levels was observed between the 4 groups (Kruskal-Wallis test x^2=8.69, p=0.03; Figure 6-2). To explore this further, we performed multiple linear regression which showed that untreated stage IV patients had significantly higher osteopontin levels than the controls and the treated stage I-III patients in age-adjusted models (p=0.004 for untreated stage IV versus controls, p<0.001 for untreated stage IV versus treated stage I-III patients; Table 6-7). No significant differences were seen between controls and the treated stage I-III or the untreated stage III patient groups (p=0.59 and 9=0.92 respectively; Table 6-7).

In healthy controls 95% of samples had osteopontin levels <103.14ng/ml. This cut off was therefore taken as the upper end of normal. We found 2.5% of patients with treated stage I-III (4/158), 17.6% of patients with untreated stage III (3/17) and 30% of patients with untreated stage IV disease (3/10) had levels higher than this cut off (Fisher's exact=0.001). The cut off that had been

previously reported is 76 ng/ml (95th centile) (Sennels *et al.*, 2007), which is the 80th centile in our control group.

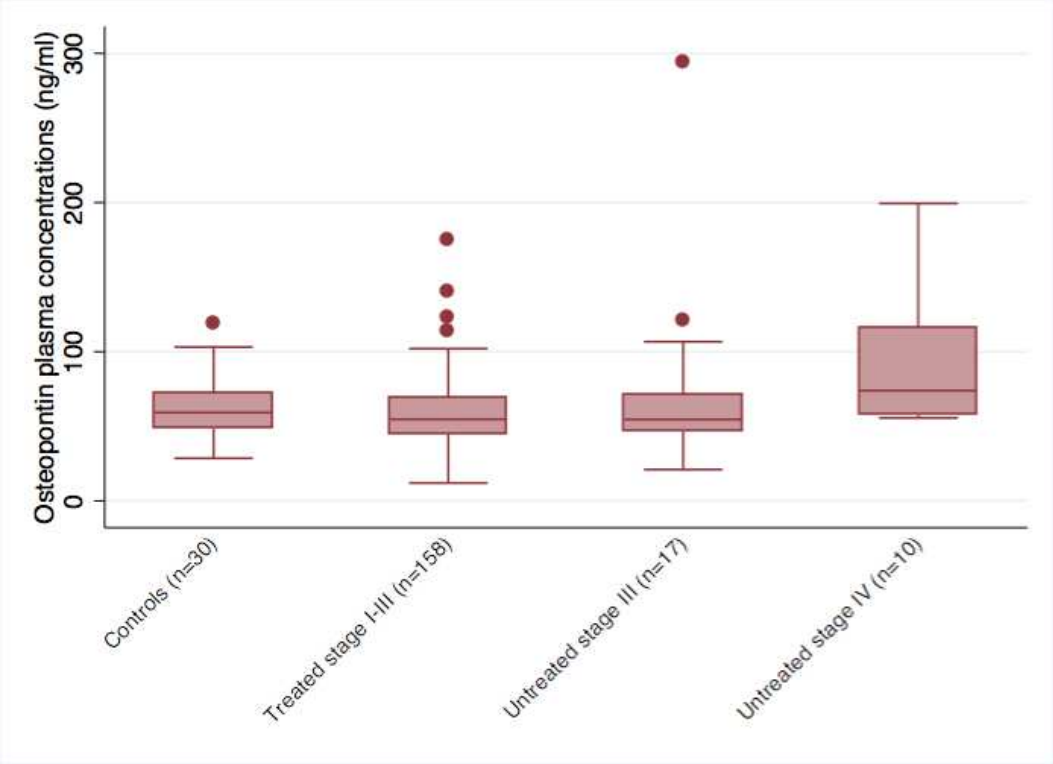


Figure 6-2: Box plots of osteopontin plasma levels in healthy controls and cases (grouped according to AJCC stage).

The edges of the box represent the 25th and 75th centiles and the whiskers represent the 5th and 95th centiles. Adapted from (Filia *et al.*, 2013).

Table 6-7: Age-adjusted linear regression coefficients for the relationship between AJCC stage and log transformed osteopontin levels when either controls or treated I-III groups were used as the baseline group.

	Observations	Controls (baseline group)		Treated I-III (baseline group)	
		Coeff (95% CI)	P value	Coeff (95% CI)	P value
Controls	30	5.50 (5.10-5.91)		-	
Treated I-III	158	-0.06 (-0.27-0.15)	0.59	5.44 (5.10-5.79)	
Untreated III	17	0.02 (-0.30-0.34)	0.92	0.07 (-0.20-0.33)	0.63
Untreated IV	10	0.57 (0.19-0.96)	0.004	0.62 (0.28-0.97)	<0.001

6.4.3 Osteopontin in patients free of disease at sampling

Age was positively correlated with osteopontin level in the disease-free patient group (Spearman's rho=0.2, p=0.02, Table 6-8; and overall Spearman's rho=0.21, p=0.004, data not shown). AJCC stage at sampling was borderline associated with osteopontin levels (p=0.06; Table 6-8) with higher median level seen in patients with treated stage III melanomas compared to those with treated stage I-II disease (64.3ng/ml and 54.1ng/ml respectively). Neither age nor stage at sampling were independent predictors of osteopontin level in a multiple linear regression model. There was no difference in osteopontin levels between SNB positive (n=38) and SNB negative (n=41) participants (Table 6-8).

Logistic regression showed a trend for increased risk of death from any cause with increasing osteopontin level but this was not statistically significant: adjusted OR was 1.24 (95%CI 0.52-2.96) for the middle versus lowest tertile, and 1.39 (95%CI 0.56-3.42) for the highest versus lowest tertile, adjusted for age, sex, BMI, site of primary, season-adjusted vitamin D level, and stage at sampling (Table 6-9). Time-to-event analyses showed support for that trend (Figure 6-3 and Figure 6-4 for MSS and OS respectively). When the 95th centile cut off was used 4/158 of the treated stage I-III patients had higher levels. Three out of four relapsed early (<1.6 years) and the remaining patient after 6 years. In total 154/158 of the treated stage I-III patients had osteopontin levels below the cutoff. Of these 79/154 were subsequent relapsers, 60 of which relapsed <1.6 years.

Table 6-8: The relationship between osteopontin levels and other characteristics of the participants who were disease-free at sampling (univariable analysis).

Adapted from (Filia *et al*., 2013).

Variable	N=	Median osteopontin (range)	Test statistic and p value
Osteopontin (ng/ml)	158	54.7 (27.9-140.0)	
Age	158		Spearman's rho=0.2, p=0.02
Sex	158		
Male	85	54.6 (27.9-122.4)	Mann-Whitney, z=-0.4, p=0.67
Female	73	54.8 (12.0-174.4)	
BMI	155		
<18.5	1	49.5	Kruskal-Wallis, χ^2=0.9, p=0.82
≥18.5 - <25	62	56.1 (28.4-174.4)	
≥25 - <30	57	53.3 (27.9122.4)	
≥30	35	54.4 (12.0-113.9)	
Breslow thickness (mm)	156		
≤1	11	47.7 (30.8-102.2)	Kruskal-Wallis, χ^2=2.5, p=0.47
>1 - ≤2	49	54.6 (31.2-174.4)	
>2 - ≤4	57	54.5 (12.0-98.4)	
>4	39	55.2 (27.9-113.9)	
Tumour site	158		
Trunk	71	53.9 (29.2-122.4)	Kruskal-Wallis, χ^2=5.7, p=0.13
Head/Neck	16	53.0 (29.3-102.0)	
Limbs	55	55.2 (12.0-174.4)	
Acral/Rare	16	68.0 (27.9-98.9)	
Mitotic rate (mm^{-2})	128		
<1	20	55.6 (31.2-93.0)	Kruskal-Wallis, χ^2=3.2, p=0.20
1-6	69	52.4 (12.0-140.0)	
>6	39	56.6 (35.2-98.8)	
Ulcerated tumours	158		
Not ulcerated	98	54.6 (28.4-122.4)	Mann Whitney, z=-1.0, p=0.32
Ulcerated	60	56.6 (12.0-174.4)	
Vitamin D (nmol/l)	150		Spearman's rho=-0.1, p=0.30
Stage at sampling	156		
Treated I/II	110	54.1 (27.9-122.4)	Mann Whitney, z=-1.9, p=0.06
Resected III	48	64.3 (28.4-93.2)	
SNB status	79		
Positive	38	55.2 (28.4-93.2)	Mann Whitney, z=-0.1, p=0.96
Negative	41	57.6 (30.8-122.4)	

Table 6-9: Association of osteopontin plasma levels with risk of death in participants who were disease-free at sampling.

Adjusted models include age, sex, BMI, tumour site, season-adjusted vitamin D level and stage at sampling. Adapted from (Filia *et al*., 2013).

		Alive, Died from melanoma			Alive, Died from any cause		
			OR (95%CI)	P value		OR (95%CI)	P value
Continuous osteopontin							
Unadjusted model		83, 75	1.11[1] (0.60-2.07)	0.74	79, 79	1.33[1] (0.71-2.49)	0.37
Adjusted model		82, 65	0.85[1] (0.42-1.70)	0.64	78, 69	1.05[1] (0.52-2.12)	0.88
Categorical osteopontin (tertiles)							
Unadjusted model	*≤49.35*	83, 75	1		79, 79	1	
	>49.35 - ≤64.34		1.41 (0.65-3.04)	0.38		1.52 (0.70-3.29)	0.29
	>64.34		1.46 (0.68-3.15)	0.33		1.83 (0.85-3.97)	0.12
Adjusted model	*≤49.35*	82, 65	1		78, 69	1	
	>49.35 - ≤64.34		1.14 (0.48-2.69)	0.77		1.24 (0.52-2.96)	0.63
	>64.34		1.02 (0.41-2.52)	0.96		1.39 (0.56-3.42)	0.48

[1] OR is interpreted as the OR associated with doubling the original osteopontin variable.

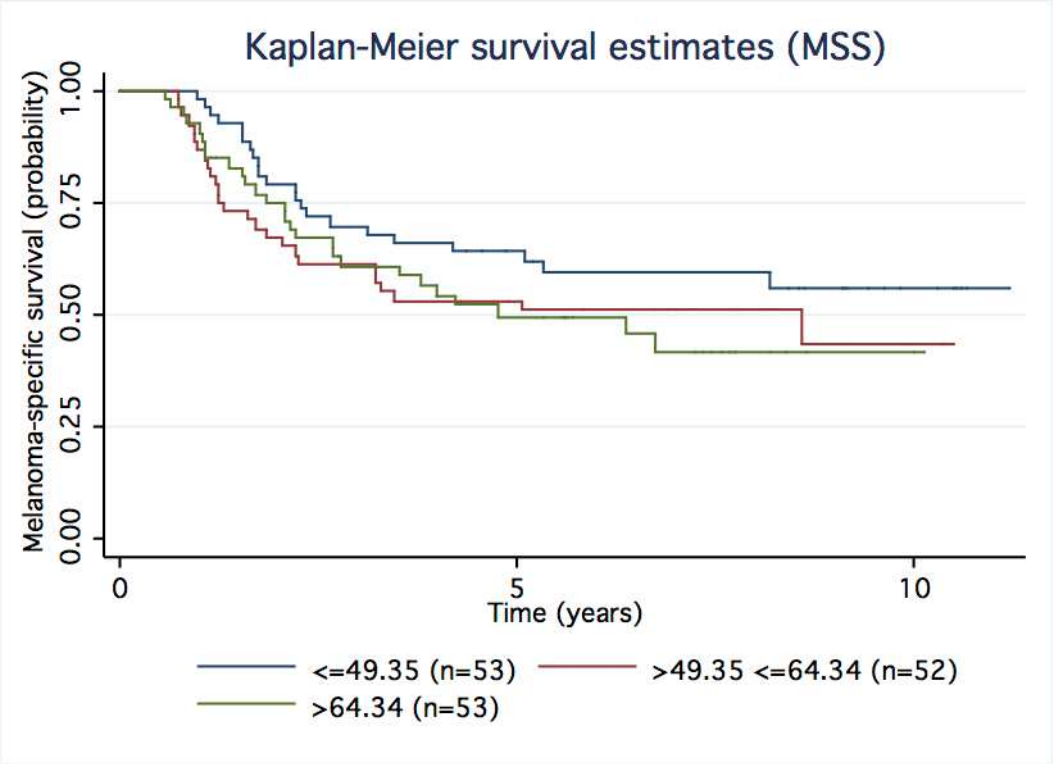


Figure 6-3: Kaplan-Meier analysis survival (MSS) estimates for osteopontin plasma concentrations in participants who were disease-free at sampling.

Adjusted HR 1.26 (95%CI 0.67-2.37), p=0.48 for middle versus low osteopontin tertile. Adjusted HR 1.19 (95%CI 0.62-2.28), p=0.61 for high versus low osteopontin tertile. Adjusted models include age, sex, BMI, tumour site, season-adjusted vitamin D level and stage at sampling. Adapted from (Filia *et al*., 2013).

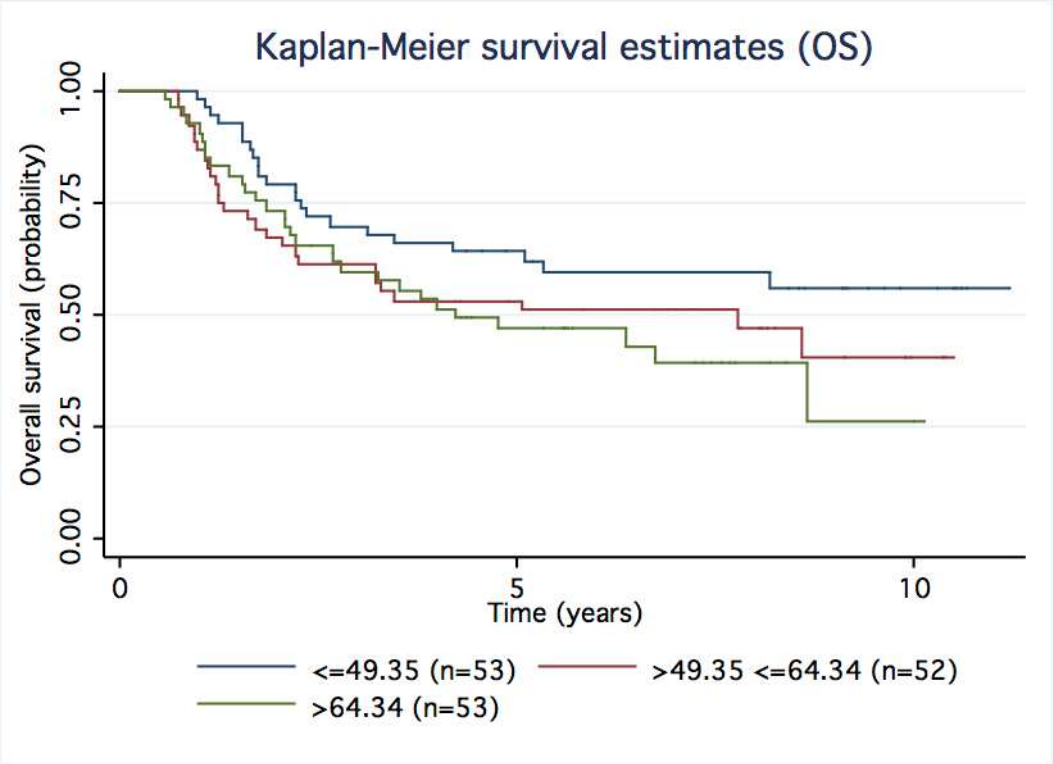


Figure 6-4: Kaplan-Meier analysis survival (OS) estimates for osteopontin plasma concentrations in participants who were disease-free at sampling.

Adjusted HR 1.34 (95%CI 0.72-2.52), p=0.36 for middle versus low osteopontin tertile. Adjusted HR 1.44 (95%CI 0.77-2.72), p=0.26 for high versus low osteopontin tertile. Adjusted models include age, sex, BMI, tumour site, season-adjusted vitamin D level and stage at sampling. Adapted from (Filia *et al*., 2013).

6.5 Discussion

6.5.1 Osteopontin assay validation

Differences in osteopontin concentrations between plasma and serum with serum having lower levels due to osteopontin cleavage during clotting had previously been described (Ayache *et al*., 2006). Here, we validated that finding, and therefore EDTA plasma samples were used in the osteopontin ELISA assay. It was noted that osteopontin levels in serum were generally about 50% lower than those observed in plasma samples, which might suggest that serum could also possibly be used to measure osteopontin as a potential biomarker, but this was not addressed.

The commercial assay was validated successfully having optimal intra- and inter-assay precision, passing parallelism and recovery tests but also having shown no interference from other analytes (Sim *et al*., 2012). The stored plasma samples from the Leeds Melanoma Cohort were also shown as suitable to be used in the assay measuring osteopontin concentration accurately without any difference due to the delayed processing after venepuncture which had occurred in that study.

6.5.2 Osteopontin plasma levels in cutaneous melanoma

There were previously no published data on osteopontin plasma levels and their prognostic significance in cutaneous stage I to III malignant melanoma. The Leeds group had previously reported a study which showed that higher tumour osteopontin gene expression levels in patients with melanoma were predictive of both relapse and death (Conway *et al*., 2009). I have therefore carried out a study of osteopontin levels as a prognostic biomarker measured in plasma samples, after assay validation, taken from melanoma patients at recruitment to the Leeds Melanoma Cohort.

The strength of this study is that we looked for the first time at osteopontin levels in patients with melanoma soon after excision of their primary tumour and for whom good follow up data were available. We also investigated for the first time if osteopontin levels could predict outcome from melanoma. The weaknesses are that this study is underpowered to look at outcome, only a single test sample was available and that no comparisons were made between osteopontin levels and

other blood markers known to have some prognostic significance, such as LDH or S100.

Currently, Breslow thickness, tumour ulceration, mitotic rate, lymph node metastasis, site of distant metastasis and serum LDH levels are the prognostic markers which are included in the most recent version of AJCC staging system. Age, tumour site and sex are also powerful prognostic factors (reviewed by Homsi *et al*. (Homsi *et al*., 2005)). Even if all these factors are considered, the variance in survival within stage is still significant; thin tumours for example might progress to advanced disease (Slingluff *et al*., 1988). Therefore there is an urgent need for identification of new prognostic biomarkers. There is also a need for a blood based screening test to detect early recurrence, especially now that more effective drug treatments are emerging (Flaherty *et al*., 2010, Robert *et al*., 2011).

A clinically useful biomarker should be measured easily, reliably and at low cost, by a sensitive and specific assay (Kulasingam and Diamandis, 2008). The impact of NGS in characterisation of tumours on the conduct of clinical trials but also in the clinical practise is very promising as already discussed in Chapter 4. However, it requires the use of tumour tissue. In the context of melanoma, tumour tissues are small and therefore it is crucial the tissues to be used effectively in case more clinical tests are required in the future after diagnosis. There is a great interest in developing blood tests which can give information on tumour cells indirectly (Sawyers, 2008). Instead of studying the tumour cells, it is possible to study the proteins secreted by tumour cells in the serum or plasma. A potential prognostic blood biomarker could also easily be used to monitor disease progression at low cost.

A number of serological prognostic biomarkers have been studied in melanoma as it was discussed in section 1.10. LDH serum level is most widely used (Balch *et al*., 2009) but is a poor marker of early recurrence and is commonly somewhat raised in otherwise healthy individuals. Serum levels of S100, the melanoma inhibitory activity protein (MIA) and amyloid A have been identified as potential prognostic biomarkers in advanced disease (Gogas *et al*., 2009). However, they failed to predict outcome in early-stage disease-free disease (Utikal *et al*., 2007) having a low sensitivity in a subgroup of patients (Hofmann *et al*., 2009). MIA serum level was tested for its prognostic significance in 1079 stage I/II patients with melanoma suggesting the usefulness of measuring MIA level in the follow-

up of patients but with low sensitivity in a subgroup of patients (Hofmann *et al*., 2009).

Osteopontin plasma levels have been reported to be much increased in metastatic melanoma in two studies since I started my study (Kluger *et al*., 2011, Maier *et al*., 2011). The authors of one of these suggested that osteopontin levels might best be used in a panel of plasma markers that were elevated in untreated stage IV compared to treated stage I/II patients (Kluger *et al*., 2011). Highly elevated osteopontin plasma levels were observed in metastatic compared to non-metastatic patients and an increase in sensitivity for the detection of metastatic disease was observed when S100 plasma levels were combined with osteopontin (Maier *et al*., 2011). The extent of metastatic disease was not discussed in either of these studies. Our study was not designed to test these observations but the data provide some support, in that higher osteopontin levels were observed in our small number of samples from individuals known to have untreated stage IV disease at sampling when compared to individuals with treated stage I-III disease. There are good biological reasons to support that osteopontin is involved in tumour progression and metastasis (Rittling and Chambers, 2004) through binding to integrins and metalloproteinases, and through activation of Ras, protein kinase C or PI3K/Akt signalling pathways (Das *et al*., 2005).

A significant difference in osteopontin levels was observed between healthy controls and patients with melanoma grouped according to AJCC stage, but most of this difference was explained by high levels in patients with stage IV disease. Using the 95th centile measure in healthy controls as a cut off however showed evidence of a trend to increased rates of results above that level with disease progression.

When only the disease-free patients were analysed in logistic regression models adjusted for known prognostic factors we saw a trend for increased risk of death with increasing osteopontin level. This provides support for the view that increased osteopontin levels might predict occult disease. Medical services in Leeds comply with the published UK melanoma guidelines, which state that there is no need for routine screening (imaging) in stage I to III patients. So, the possibility that there may have been occult disease in our disease-free patients should be considered. Therefore we did not look at time to relapse in this study.

Osteopontin plasma levels did not correlate with tumour levels; however, there were only 34 samples with matched plasma and tumour levels from the DASL

microarray experiment (data not shown). The plasma and tumour levels also failed to show a strong correlation using qRT-PCR gene expression data from 37 tumours (data not shown). These numbers are too small to investigate if there is a strong correlation. However, it is known that gene expression levels and protein levels do not always correlate (Vogel and Marcotte, 2012). Another possible explanation is that this single measurement in the blood taken at recruitment to the study might not reflect the tumour levels at diagnosis although recruitment in the study was performed within 3-6 months when possible.

Serial measurements of osteopontin plasma levels at frequent intervals after diagnosis may prove to be more informative than a single measurement. A study in uveal melanoma (Barak *et al*., 2011) showed a significant increase in osteopontin level from 12-18 months to 6-12 months prior to clinical confirmation of metastasis. The trend of the increase of osteopontin level was not significantly different compared to the osteopontin measurements in disease-free patients but this could be due to the small study size. When we used a measure of osteopontin level higher than 95th centile, relapse occurred in significant numbers of melanoma patients with results within the normal range sampled within 19 months of relapse. It seems unlikely therefore that a single measurement of osteopontin will have sufficient sensitivity and specificity for use in clinical practice. It may however be the case that biologically different tumours are associated with increased levels of different tumour markers and that a panel of markers will be necessary. Therefore the possibility of studying the longitudinal monitoring utility of osteopontin level in conjunction with other markers should also be assessed in future studies with imaging of patients with high levels to detect occult metastases.

Chapter 7
Final discussion

7.1 Melanoma genomics and prognostic biomarkers

The genetic changes, which "drive" melanoma, were unknown until recently. Progress was slower than for many cancers, partly because melanoma tumours are small and fresh/frozen tissue hard to come by. Some of this delay occurred because the most common "driver" mutant *BRAF* was not recognised as a proto-oncogene till systemic large-scale screening became possible. *BRAF* as an oncogene was reported in 2002 (Davies *et al.*, 2002). Although much more is known now about *BRAF* and other drivers, the rapid emergence of resistance in patients treated with BRAF inhibitors, and lack of treatment for patients without *BRAF* mutations means that much more information is needed about genomics to inform treatment decisions. In this thesis I describe attempts to evaluate next generation sequencing (NGS) as a tool to further investigate melanoma genomics with particular reference to primary tumours.

The American Joint Committee on Cancer (AJCC) staging scheme is the gold standard to assess prognosis in melanoma (Balch *et al.*, 2009). However, there is still variance in outcome that cannot be explained with the current prognostic factors (Ross, 2006), which highlights the need to identify new prognostic biomarkers. Increasing efforts have been made to identify new, more accurate molecular prognostic markers. In this thesis I have evaluated osteopontin plasma levels as a potential biomarker in early stage disease (AJCC stage I-III) patients with melanoma.

7.1.1 Identification of tumour prognostic biomarkers using next generation sequencing

Primary melanoma tumours are small and therefore rarely cryopreserved. To investigate somatic alterations in primary melanoma tumours, the use of formalin-fixed paraffin-embedded (FFPE) tissue is therefore necessary. The principal aim of this thesis was to assess the use of whole-genome next generation sequencing (NGS) technology to identify copy number alterations (CNAs) in melanoma using FFPE-derived DNA. NGS offers substantial benefits over array-based approaches

for copy number analysis: more accurate estimation of high CNAs as counting sequences does not suffer from signal saturation (Meyerson *et al*., 2010); the potential of higher resolution to identify small CNAs and precise estimation of the breakpoints of CNAs (Meyerson *et al*., 2010); it can be performed using low amounts of DNA while array-based methods usually require large amounts of FFPE-derived DNA (greater than 500ng).

There are only two studies reporting CNAs using FFPE primary melanoma tissue on the aCGH platform and both have used greater than 500ng DNA input (Bastian *et al*., 2003, Curtin *et al*., 2005). Recent studies have demonstrated the successful performance of low-coverage NGS (range 0.004x-0.06x) for copy number analysis using small amounts of FFPE-derived DNA from oral tumours and lung squamous cell lines and tumours (Wood *et al*., 2010, Belvedere *et al*., 2012). Reproducible results between FFPE and matched frozen tumours on the NGS platform have been reported using small data sets (Wood *et al*., 2010, Schweiger *et al*., 2009). These results are promising but further effort is needed to assess the feasibility of this technology to allow effective use of melanoma tumour tissue in future large-scale studies. In this thesis I report such a study using relatively higher coverage than reported by Wood *et al* by multiplexing fewer samples per lane.

I have shown that low DNA concentrations are, not surprisingly, associated with lower probability of adequate library preparation and more PCR replicates sequenced but that there was no clear cut off identifiable to determine a lesion not worth sequencing or worth sequencing at lower depth. It is possible that a more accurate quantification method (use of the higher sensitivity picogreen kit) could serve as a better indicator of what is worth sequencing, but this was not assessed. Efficient and cost-effective use of FFPE samples for biomarker discovery and testing will require the identification of the depth of sequencing that can be performed based on the concentration of DNA input, which is a requisite for future studies.

Previously reported biases such as sequencing of PCR replicates, GC-bias and errors in the human reference genome sequence (Dohm *et al*., 2008, Pickrell *et al*., 2011) were assessed and corrected by applying statistical models or removal of data to minimise false positive copy number changes. Future work on the human reference genome assembly may resolve problems such as the errors in the reference and allow the identification of CNAs in these regions that we might miss at the moment. Improvements in sequencing technology will also reduce sequencing errors such as those observed at the 3'-end of the fragments (phasing

artefact). Currently, quality filtering is suggested at the expense of removing part of the data (Minoche *et al*., 2011). This means that we might miss some changes which we might identify by re-analysing the data in the future.

The median coverage for chromosomes 1-22 was 1.01x (range, 0.33x-1.46x). With this coverage smaller CNAs are likely to be identified than those reported by other whole-genome studies (Wood *et al*., 2010, Belvedere *et al*., 2012), but the exact resolution has not yet been determined. Some of these samples could be re-sequenced to increase depth, however this would only be possible when large amounts of DNA was used to generate the libraries (high complexity libraries) to avoid sequencing of PCR replicates. This suggests that the identification of translocations would be limited using these samples.

I have reported evidence that the CNAs detected were in regions previously described to be associated with melanoma and which are biologically credible (such as chromosome 10/ *PTEN* in *BRAF* mutated tumours). This provides evidence to justify this sequencing approach. Correlations between whole-genome DASL expression and NGS copy number data from the same tumours further supported the validity of these data. There are still 265 samples which need to be analysed which will increase our statistical power to identify reliable CNAs associated with outcome or other tumour characteristics.

The conclusion from my work is that NGS should provide sufficiently robust data for assessing/applying a proposed genomic biomarker, but the larger data set which is pending will increase the confidence with which I can assess its likely value. If a potential biomarker is identified, its applicability would need to be tested in a more representative sample set (i.e. including smaller tumours) using specially designed tests.

NGS can also be used to identify transcriptional changes (known as RNA-sequencing) (Meyerson *et al*., 2010). In Chapter 2, I report the evaluation of an Illumina protocol designed to prepare libraries using FFPE-derived RNA. The advantage of RNA-sequencing over traditional microarray approaches is that alternative splice variants and transcripts from gene fusion events can be identified (Meyerson *et al*., 2010). The sequencing data produced from these libraries have not been analysed yet. The intent was to assess the consistency across matched FFPE and frozen tissue, and with analyses conducted on microarrays. This would allow a design of a large-scale study as described above to assess the feasibility of NGS using RNA from FFPE primary melanoma tumours.

7.1.2 Osteopontin as a potential plasma biomarker

Primary melanoma tumour tissues should be used efficiently in case clinical tests are needed in the future after diagnosis. Therefore there is a great interest in studying the proteins secreted by tumour cells in the serum or plasma instead of studying the tumour cells directly (Sawyers, 2008). A prognostic blood biomarker allowing detection of early relapse in the clinic is desirable especially as targeted therapies come on line. In Chapter 6 I present an assessment of osteopontin plasma levels using a commercial ELISA assay as a potential prognostic biomarker in early-stage melanoma patients. Following the validation of the assay, higher osteopontin levels were observed in our small number of samples from individuals known to have untreated stage IV disease when compared to individuals with treated stage I-III disease. A trend for increased risk of death with increasing osteopontin level was observed when only disease-free patients were analysed, but this did not reach formal significance. The greatest limitation of this work is that osteopontin levels were not compared with other well-studied serum markers, such as LDH.

My conclusion from this work, which has been published (Filia *et al.*, 2013), is that measurement of the osteopontin plasma level is not likely to be useful as a prognostic biomarker for early-stage disease patients. It may however be the case that biologically different tumours are associated with increased levels of different tumour markers and that a panel of markers will be necessary. A study in uveal melanoma has reported a significant increase in osteopontin level from 12-18 months to 6-12 months prior to clinical confirmation of metastasis (Barak *et al.*, 2011). It would therefore be interesting to study the longitudinal monitoring utility of osteopontin level in conjunction with other markers in the future. In the future serum proteomic profiling studies may serve as a means to discover new prognostic biomarkers in melanoma which could be tested alone or in combination with other markers to assess their prognostic significance.

7.2 Biomarkers in response to vitamin D

Most of the adjuvant therapies trialled in melanoma patients have not proved very promising (Molife and Hancock, 2002, Eggermont *et al.*, 2009) except for IFN-α and Ipilimumab (anti-CTLA4) (Davar *et al.*, 2012, Eggermont and Robert, 2011). Vitamin D has been reported as a potential new adjuvant therapy with promising results in

prostate cancer (Deeb *et al*., 2007). In patients with melanoma, there are ongoing clinical trials which aim to study the effect of vitamin D as an adjuvant treatment alone or in combination with other regimens (U.S. National Institutes of Health, 2013). In Chapter 5 I presented an *in vitro* study designed to examine the anti-tumour effects of vitamin D in melanoma. It is intriguing that these findings could ultimately lead to the identification of predictive biomarkers that would show likely benefits to vitamin D supplementation. The up-regulation of *VDR* and *CYP24A1* after vitamin D treatment in both cell lines suggests that the detected changes are specific to vitamin D. These have been the most widely reported changes in response to vitamin D in various cancer models (Kriebitzsch *et al*., 2009, Deeb *et al*., 2007).

The pathways significantly overrepresented in my data included Wnt, TGF-β, cell adhesion and proliferation pathways, which have been reported to be altered in other cancer models in response to vitamin D (Deeb *et al*., 2007, Fleet *et al*., 2012) which increased confidence that the results are vitamin D specific. Enrichment of different pathways such as NOTCH signalling only in MeWo cells might reflect tumour heterogeneity. The overrepresentation of pathways such as G-protein signalling at 6 hours in SkMel28 cells implies a non-genomic effect, however, this needs detailed examination of the genes involved in the reported pathways and identification of possible VDREs in their promoters. The *in silico* methodology used here predicted that most of the genes differentially expressed in response to vitamin D were VDR target genes but the limitation of this approach was highlighted when one of the genes not predicted to be a VDR target has been reported in a recent ChIP-sequencing experiment as a VDR target containing a-non classical VDRE (non-DR3 type) (Heikkinen *et al*., 2011). It is therefore, essential to compare the data with available ChIP-sequencing data (such as those held in the ENCODE database) to identify potential non-genomic effects. It might also be that 6 hours was a late time point to identify non-genomic effects as these effects have been reported to be rapid within 40 minutes or 2-3 hours (Deeb *et al*., 2007).

The greatest limitation of this work is that gene expression data were not available from the vitamin D resistant cell line to compare with our results but this could be performed in the future. Future work should also include the identification of whether transcriptional changes identified *in vitro* are mirrored *in vivo* using whole-genome DASL gene expression data from patients from the Leeds Melanoma Cohort, and to what extent these effects are moderated by complex tumour biology.

7.3 Concluding Remarks

In the course of this period of study I have investigated a number of biomarker discovery pipelines. The challenging nature of primary melanoma material has meant that until recently the focus of biomarker discovery was by necessity focused on small studies utilising fresh material or alternate tissue sources such as serum and plasma. The work I have presented on osteopontin has demonstrated that identifying biomarkers in serum and/or plasma present their own challenges and will be difficult using existing tissue-banked material. In contrast interrogation of archival FFPE blocks has provided a wealth of data with respect to differential expression of genes and their copy numbers. These approaches seem most likely to offer novel biomarker findings and my own analysis of the data obtained during this research will be concluded in the near future. Most significantly this research has shown that it is now possible, using advanced genomic techniques, to mine the wealth of data sourced from primary melanoma without reliance on fresh tissue.

Chapter 8
Appendix A

A.1 Cell line authentication reports

A.1.1 Cell lines used at Leeds University

A) MeWo
B) SkMel28
C) SkMel5
D) MelJuso

A.1.2 Cell lines used at Saarland University, Homburg, Germany

A) MeWo
B) SkMel28
C) SkMel5
D) MelJuso

A.1.1-A)

23/07/2012

Cell line STR profile report

CR-UK Centre Genomics Facility
Leeds Institute of Molecular Medicine

Reference data for MeWo

				D3 S1358	TH01 (chr11)	D21 S11	D18 S51	Penta E (chr15)	D5 S818	D13 S317	D7 S820	D16 S539	CSF1PO (chr5)	Penta D (chr21)	Amel	vWA (chr12)	D8 S1179	TPOX (chr2)	FGA (chr4)
Reference profile from :	**CLS**	**n/a**	MSI stable	17	7,9	30,32.2	14,17	5	12,13	8,9	10,12	10,12	12	10	X,Y	15	13,15	8,10	22

Sample data

Sample		D3 S1358	TH01 (chr11)	D21 S11	D18 S51	Penta E (chr15)	D5 S818	D13 S317	D7 S820	D16 S539	CSF1PO (chr5)	Penta D (chr21)	Amel	vWA (chr12)	D8 S1179	TPOX (chr2)	FGA (chr4)
MeWo_JNB_leedsrpt	*GV05076A*	16,?17	7,?9,	28,29	?14,16, ?17	5,?15, 17	11,?12, 13	?8,?9, 11,14	8,11, ?12	11,12	12	9,?10, 12	X,?Y	?15,17, 18,?19	11,13, ?15	8,10	20,22
	GV05076B	16,?17	7,?9	28,29	?14,16, ?17	5,?15, 17	11,13	?8,11, 14	8,11, ?12	11,12	12	9,12	X,?Y	?15,17, 18	11,13, ?15	8,10	20,22

Samples are labelled using the format:
cell line name _ initials of lab head _ date of receipt or other unique identifier

Analysis

Sample match to reference	Sample profile contains multiple additional low level peaks, not all of which have been scored. The major peaks in this sample do not match the reference profile.	
DSMZdb	Using major peaks only, no close match.	DSMZ STR profile database : http://www.dsmz.de/fp/cgi-bin/str.html (registration required). Searches 2278 cell lines from ATCC, DSMZ, JCRB & RIKEN.
CLIMAdb	Not done	Cell line integrated molecular authentication (CLIMA) database http://bioinformatics.istge.it/clima/
CGP		Cancer Genome Project STR data : http://www.sanger.ac.uk/research/projects/cancergenome/archive/ (registration required). Details of CGP data available only to registered users.
Comment	The data are consistent with a contaminated sample in which MM96 is the major component and MeWo is a minor component.	

A.1.1-B)

09/08/2012

Cell line STR profile report

CR-UK Centre Genomics Facility
Leeds Institute of Molecular Medicine

Reference data for SKMEL28

		D3 S1358	TH01 (chr11)	D21 S11	D18 S51	Penta E (chr15)	D5 S818	D13 S317	D7 S820	D16 S539	CSF1PO (chr5)	Penta D (chr21)	Amel	vWA (chr12)	D8 S1179	TPOX (chr2)	FGA (chr4)
Reference profile from :	**ATCC HTB-72**		7				11,13	11,12	9.3,10	9,12	10,12		X,Y	16,19		8,12	

Sample data

Sample		D3 S1358	TH01 (chr11)	D21 S11	D18 S51	Penta E (chr15)	D5 S818	D13 S317	D7 S820	D16 S539	CSF1PO (chr5)	Penta D (chr21)	Amel	vWA (chr12)	D8 S1179	TPOX (chr2)	FGA (chr4)
SKMEL28_JNB_LdsAug1	*GV05084A*	16,18	7	28,29	12,16	8,12	11?,13	11,12	9.3,10	9,12	10,12	9,10	X,Y	16?,19	13	8,12	19
	GV05084B	16,18	7	28,29	12,16	8,12	11?,13	11,12	9.3,10	9,12	10,12	9,10	X,Y	16,19	13	8?,12	19

Samples are labelled using the format:
cell line name _ initials of lab head _ date of receipt or other unique identifier

Analysis

Sample match to reference	Matches 9/9 markers.	
DSMZdb	Closest matches are to SKMEL28, although all SKMEL28 profiles retrieved lack allele 9.3 at D7.	DSMZ STR profile database : http://www.dsmz.de/fp/cgi-bin/str.html (registration required). Searches 2278 cell lines from ATCC, DSMZ, JCRB & RIKEN.
CLIMAdb	Retrieves SKMEL28.	Cell line integrated molecular authentication (CLIMA) database http://bioinformatics.istge.it/clima/
CGP	CGP profile for SKMEL28 lacks allele 9.3 at D7. Matches at 8/8 of the remaining markers in the reference profile and at 5/5 additional markers.	Cancer Genome Project STR data : http://www.sanger.ac.uk/research/projects/cancergenome/archive/ (registration required). Details of CGP data available only to registered users.
Comment		

A.1.1-C)

19/07/2012

Cell line STR profile report

CR-UK Centre Genomics Facility
Leeds Institute of Molecular Medicine

Reference data for SKMEL5

	D3 S1358	TH01 (chr11)	D21 S11	D18 S51	Penta E (chr15)	D5 S818	D13 S317	D7 S820	D16 S539	CSF1PO (chr5)	Penta D (chr21)	Amel	vWA (chr12)	D8 S1179	TPOX (chr2)	FGA (chr4)
Reference profile from : ATCC HTB-70		6,9				11,13	10,12	9,12	10,12	10,13		X	14,18		11	

Sample data

Sample **SKMEL5_JNB_leedsrpt**	D3 S1358	TH01 (chr11)	D21 S11	D18 S51	Penta E (chr15)	D5 S818	D13 S317	D7 S820	D16 S539	CSF1PO (chr5)	Penta D (chr21)	Amel	vWA (chr12)	D8 S1179	TPOX (chr2)	FGA (chr4)
GV05076A	16,17	6,?9	29	15,16	5,12	11,13	10,12	9,12	10,12	10,13	9,11	X	14,18	12,15	11	20.2,22
GV05076B	16,17	6,9	29	15,16	5,12	11,13	10,12	9,12	10,12	10,13	9,11	X	14,18	12,15	11	20.2,22

Samples are labelled using the format:
cell line name _ initials of lab head _ date of receipt or other unique identifier

Analysis

Sample match to reference	Matches 9/9 markers.	
DSMZdb	Perfect match to SKMEL5.	DSMZ STR profile database : http://www.dsmz.de/fp/cgi-bin/str.html (registration required). Searches 2278 cell lines from ATCC, DSMZ, JCRB & RIKEN.
CLIMAdb	Retrieves SKMEL5.	Cell line integrated molecular authentication (CLIMA) database http://bioinformatics.istge.it/clima/
CGP	Perfect match to 9/9 markers in reference profile and to 5/5 additional markers.	Cancer Genome Project STR data : http://www.sanger.ac.uk/research/projects/cancergenome/archive/ (registration required). Details of CGP data available only to registered users.
Comment		

A.1.1-D)

19/07/2012

Cell line STR profile report

CR-UK Centre Genomics Facility
Leeds Institute of Molecular Medicine

Reference data for MelJuso

	D3 S1358	TH01 (chr11)	D21 S11	D18 S51	Penta E (chr15)	D5 S818	D13 S317	D7 S820	D16 S539	CSF1PO (chr5)	Penta D (chr21)	Amel	vWA (chr12)	D8 S1179	TPOX (chr2)	FGA (chr4)
Reference profile from : DSMZ ACC 74		9.3				11,12	8,9	10,12	12,13	10,11		X	16		8,10	

Sample data

Sample **MelJuso_JNB_leedsrpt**	D3 S1358	TH01 (chr11)	D21 S11	D18 S51	Penta E (chr15)	D5 S818	D13 S317	D7 S820	D16 S539	CSF1PO (chr5)	Penta D (chr21)	Amel	vWA (chr12)	D8 S1179	TPOX (chr2)	FGA (chr4)
GV05076A	14,17	9.3	29	15,18	5,13	11,12	8,9	10,12	12,13	10,11	13	X	16	9,14	8,10	22
GV05076B	14,17	9.3	29	15,18	5,13	11,12	8,9	10,12	12,13	10,11	13	X	16	9,14	8,10	22

Samples are labelled using the format:
cell line name _ initials of lab head _ date of receipt or other unique identifier

Analysis

Sample match to reference	Matches 9/9 markers.	
DSMZdb	Perfect match to MelJuso.	DSMZ STR profile database : http://www.dsmz.de/fp/cgi-bin/str.html (registration required). Searches 2278 cell lines from ATCC, DSMZ, JCRB & RIKEN.
CLIMAdb	No record retrieved.	Cell line integrated molecular authentication (CLIMA) database http://bioinformatics.istge.it/clima/
CGP	Perfect match to 9/9 markers in reference profile and to 5/5 additional markers.	Cancer Genome Project STR data : http://www.sanger.ac.uk/research/projects/cancergenome/archive/ (registration required). Details of CGP data available only to registered users.
Comment		

A.1.2-A)

09/08/2012

Cell line STR profile report

CR-UK Centre Genomics Facility
Leeds Institute of Molecular Medicine

Reference data for MeWo

				D3 S1358	TH01 (chr11)	D21 S11	D18 S51	Penta E (chr15)	D5 S818	D13 S317	D7 S820	D16 S539	CSF1PO (chr5)	Penta D (chr21)	Amel	vWA (chr12)	D8 S1179	TPOX (chr2)	FGA (chr4)
Reference profile from :	**CLS**	**n/a**	MSI stable	17	7,9	30,32.2	14,17	5	12,13	8,9	10,12	10,12	12	10	X,Y	15	13,15	8,10	22

Sample data

Sample		D3 S1358	TH01 (chr11)	D21 S11	D18 S51	Penta E (chr15)	D5 S818	D13 S317	D7 S820	D16 S539	CSF1PO (chr5)	Penta D (chr21)	Amel	vWA (chr12)	D8 S1179	TPOX (chr2)	FGA (chr4)
MeWo_JNB_GerAug12	*GV05084A*	17	7,9	30,32.2	14,17	5,15	12,13	8,9	10,12	10,12	12	10	X,Y	15	13,15	8,10	22
	GV05084B	17	7,9	30,32.2	14,17	5,15	12,13	8,9	10,12	10,12	12	10	X,Y	15	13,15	8,10	22

Samples are labelled using the format:
cell line name _ initials of lab head _ date of receipt or other unique identifier

Analysis

Sample match to reference	Matches reference profile at 15/16 markers. CLS reference profile lacks allele 15 at Penta E.	
DSMZdb	Perfect match to MeWo at 9/9 markers. No other close matches.	DSMZ STR profile database : http://www.dsmz.de/fp/cgi-bin/str.html (registration required). Searches 2278 cell lines from ATCC, DSMZ, JCRB & RIKEN.
CLIMAdb	Perfect match to MeWo at 9/9 markers.	Cell line integrated molecular authentication (CLIMA) database http://bioinformatics.istge.it/clima/
CGP	Matches CGP profile at 13/14 markers. CGP MeWo profile lacks allele 9 at D13.	Cancer Genome Project STR data : http://www.sanger.ac.uk/research/projects/cancergenome/archive/ (registration required). Details of CGP data available only to registered users.
Comment		

A.1.2-B)

09/08/2012

Cell line STR profile report

CR-UK Centre Genomics Facility
Leeds Institute of Molecular Medicine

Reference data for SKMEL28

			D3 S1358	TH01 (chr11)	D21 S11	D18 S51	Penta E (chr15)	D5 S818	D13 S317	D7 S820	D16 S539	CSF1PO (chr5)	Penta D (chr21)	Amel	vWA (chr12)	D8 S1179	TPOX (chr2)	FGA (chr4)
Reference profile from :	**ATCC**	**HTB-72**		7				11,13	11,12	9.3,10	9,12	10,12		X,Y	16,19		8,12	

Sample data

Sample		D3 S1358	TH01 (chr11)	D21 S11	D18 S51	Penta E (chr15)	D5 S818	D13 S317	D7 S820	D16 S539	CSF1PO (chr5)	Penta D (chr21)	Amel	vWA (chr12)	D8 S1179	TPOX (chr2)	FGA (chr4)
SKMEL28_JNB_GerAug1	*GV05084A*	16,18	7	28,29	12,16	8?,12	11,13	11,12	9.3,10	9	10,12	9,10	X,Y	16,19	13	8,12	19
	GV05084B	16,18	7	28,29	12,16	8,12	11,13	11,12	9.3,10	9	10,12	9,10	X,Y	16,19	13	8,12	19

Samples are labelled using the format:
cell line name _ initials of lab head _ date of receipt or other unique identifier

Analysis

Sample match to reference	Matches at 8/9 markers. Loss of allele 12 at D16.	
DSMZdb	Closest matches are to SKMEL28, although all SKMEL28 profiles retrieved lack allele 9.3 at D7.	DSMZ STR profile database : http://www.dsmz.de/fp/cgi-bin/str.html (registration required). Searches 2278 cell lines from ATCC, DSMZ, JCRB & RIKEN.
CLIMAdb	Retrieves SKMEL28.	Cell line integrated molecular authentication (CLIMA) database http://bioinformatics.istge.it/clima/
CGP	CGP profile for SKMEL28 lacks allele 9.3 at D7. Matches at 7/8 of the remaining markers in the reference profile (loss of allele 12 at D16 as noted above) and at 5/5 additional markers.	Cancer Genome Project STR data : http://www.sanger.ac.uk/research/projects/cancergenome/archive/ (registration required). Details of CGP data available only to registered users.
Comment	Note that the previous SKMEL28 sample received from this source, SKMEL28_JNB_Jul12, did not match the reference profile but instead matched the profile for MeWo.	

A.1.2-C)

09/08/2012

Cell line STR profile report

CR-UK Centre Genomics Facility
Leeds Institute of Molecular Medicine

Reference data for SKMEL5

	D3 S1358	TH01 (chr11)	D21 S11	D18 S51	Penta E (chr15)	D5 S818	D13 S317	D7 S820	D16 S539	CSF1PO (chr5)	Penta D (chr21)	Amel	vWA (chr12)	D8 S1179	TPOX (chr2)	FGA (chr4)
Reference profile from : ATCC HTB-70		6,9				11,13	10,12	9,12	10,12	10,13		X	14,18		11	

Sample data

Sample **SKMEL5_JNB_GerAug12**	D3 S1358	TH01 (chr11)	D21 S11	D18 S51	Penta E (chr15)	D5 S818	D13 S317	D7 S820	D16 S539	CSF1PO (chr5)	Penta D (chr21)	Amel	vWA (chr12)	D8 S1179	TPOX (chr2)	FGA (chr4)
GV05084A	16,18	7	28,29	12,16	8,12	11,13	11,12	9.3,10	9	10,12	9,10	X,Y	16,19	13	8,12	19
GV05084B	16,18	7	28,29	12,16	8?,12	11,13	11,12	9.3,10	9	10,12	9,10	X,Y	16,19	13	8,12	19

Samples are labelled using the format:
cell line name _ initials of lab head _ date of receipt or other unique identifier

Analysis

Sample match to reference	Does not match reference.	
DSMZdb	Closest matches are to SKMEL28, although all SKMEL28 profiles retrieved lack allele 9.3 at D7.	DSMZ STR profile database : http://www.dsmz.de/fp/cgi-bin/str.html (registration required). Searches 2278 cell lines from ATCC, DSMZ, JCRB & RIKEN.
CLIMAdb	Retrieves SKMEL28.	Cell line integrated molecular authentication (CLIMA) database http://bioinformatics.istge.it/clima/
CGP	Does not match profile.	Cancer Genome Project STR data : http://www.sanger.ac.uk/research/projects/cancergenome/archive/ (registration required). Details of CGP data available only to registered users.
Comment	The previous SKMEL5 sample received from this source, SKMEL5_JNB_Jul12, did not match the SKMEL5 reference profile but did match the SKMEL28 reference profile.	

A.1.2-D)

08/08/2012

Cell line STR profile report

CR-UK Centre Genomics Facility
Leeds Institute of Molecular Medicine

Reference data for MelJuso

			D3 S1358	TH01 (chr11)	D21 S11	D18 S51	Penta E (chr15)	D5 S818	D13 S317	D7 S820	D16 S539	CSF1PO (chr5)	Penta D (chr21)	Amel	vWA (chr12)	D8 S1179	TPOX (chr2)	FGA (chr4)
Reference profile from :	DSMZ	ACC 74		9.3				11,12	8,9	10,12	12,13	10,11		X	16		8,10	

Sample data

Sample MelJuso_JNB_GerAug12	D3 S1358	TH01 (chr11)	D21 S11	D18 S51	Penta E (chr15)	D5 S818	D13 S317	D7 S820	D16 S539	CSF1PO (chr5)	Penta D (chr21)	Amel	vWA (chr12)	D8 S1179	TPOX (chr2)	FGA (chr4)
GV05084A	15	9.3	30	12	16	12	8,12	8,11	9,12	11	11,12	X,Y	15,16	12,15	8	21,24
GV05084B	15	9.3	30	12	16	12	8,12	8,11	9,12	11	11,12	X,Y	15,16	12,15	8	21,24

Samples are labelled using the format:
cell line name _ initials of lab head _ date of receipt or other unique identifier

Analysis

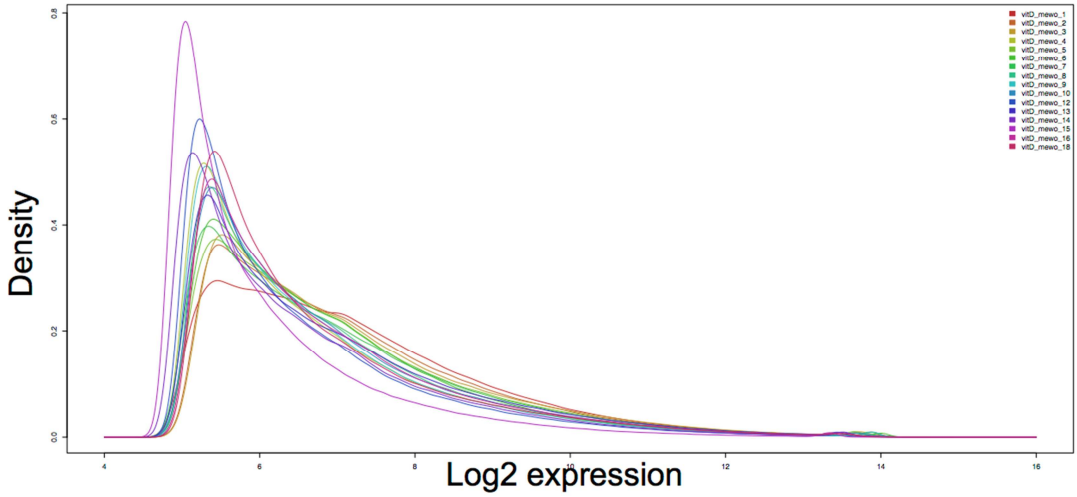
Sample match to reference	Does not match reference profile.	
DSMZdb	No close matches.	DSMZ STR profile database : http://www.dsmz.de/fp/cgi-bin/str.html (registration required). Searches 2278 cell lines from ATCC, DSMZ, JCRB & RIKEN.
CLIMAdb	No record retrieved.	Cell line integrated molecular authentication (CLIMA) database http://bioinformatics.istge.it/clima/
CGP	Does not match profile.	Cancer Genome Project STR data : http://www.sanger.ac.uk/research/projects/cancergenome/archive/ (registration required). Details of CGP data available only to registered users.
Comment	The profile from this sample is the same as for the previous sample MelJuso_JNB_Jul12.	

A.2 Quality control plots before and after microarray gene expression data pre-processing

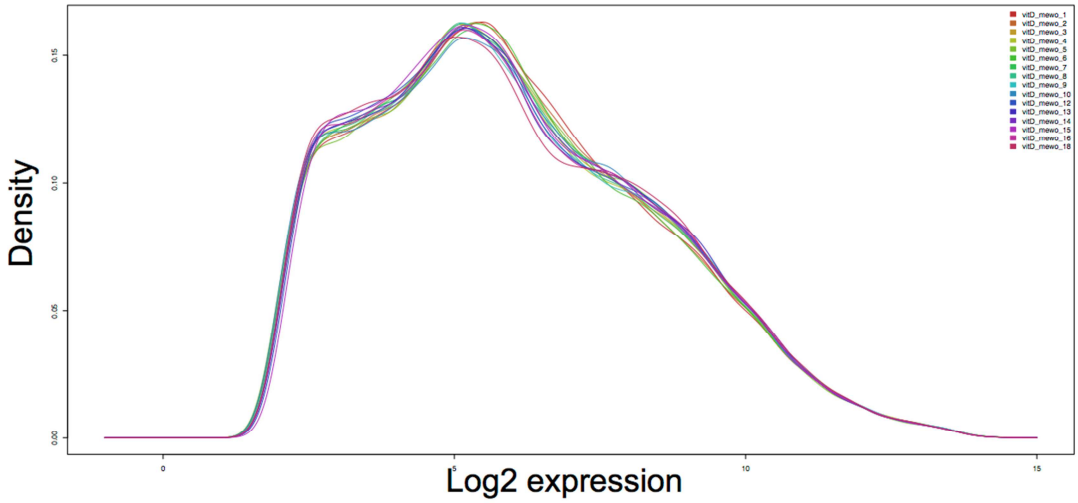
A.2.1 Density distributions of log-intensities of each sample before and after Affymetrix U133 plus 2 microarray gene expression data pre-processing (MeWo)

MeWo

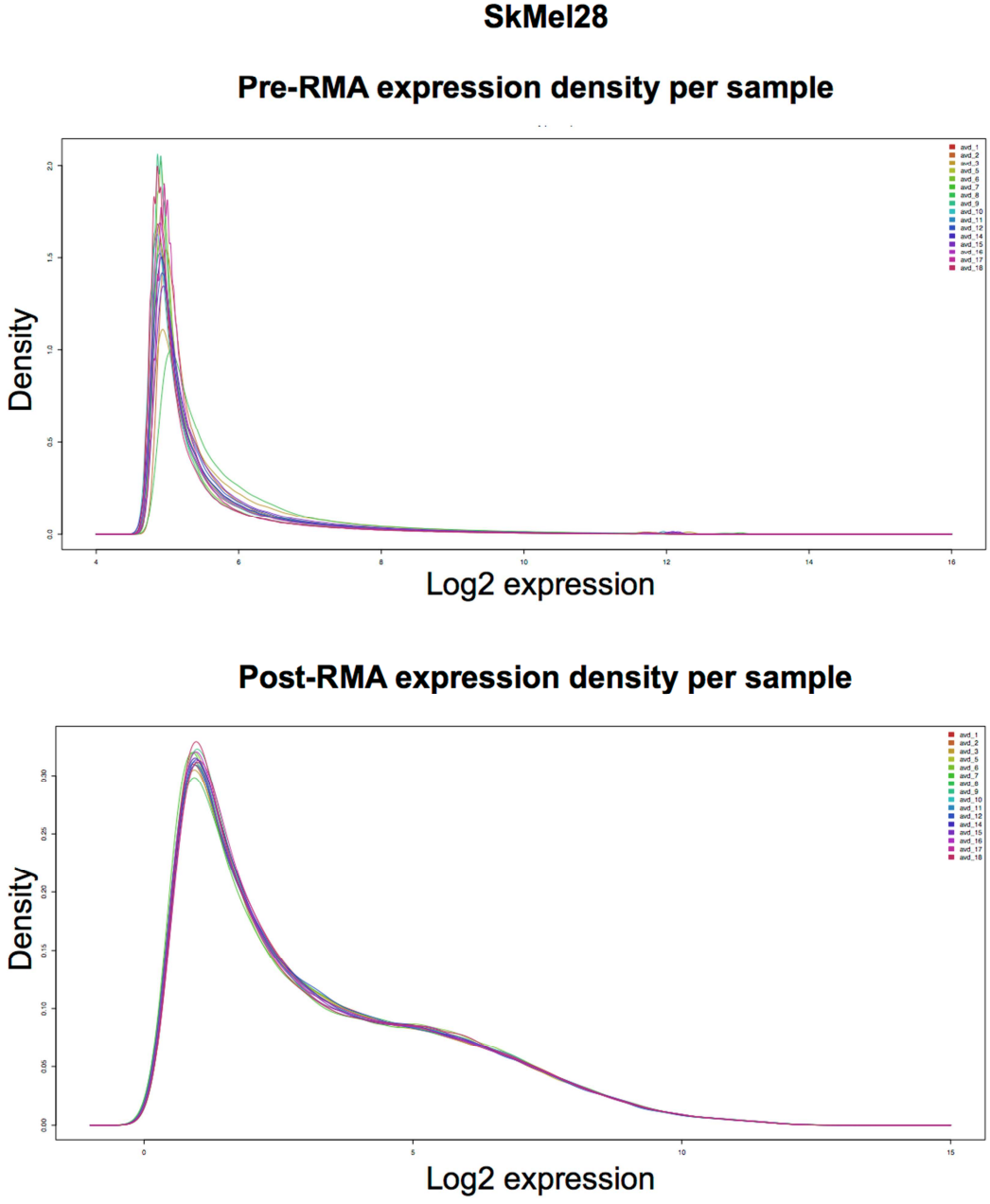
Pre-RMA expression density per sample



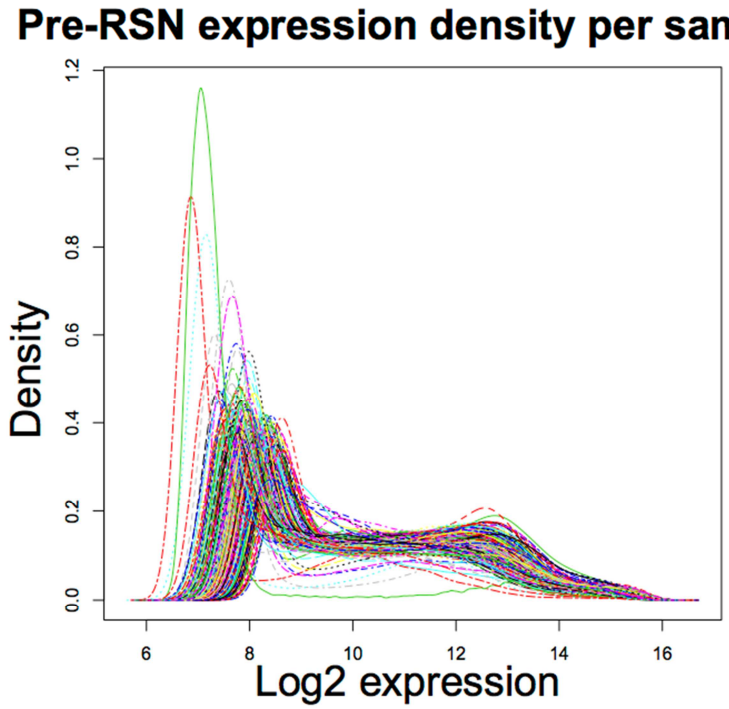
Post-RMA expression density per sample

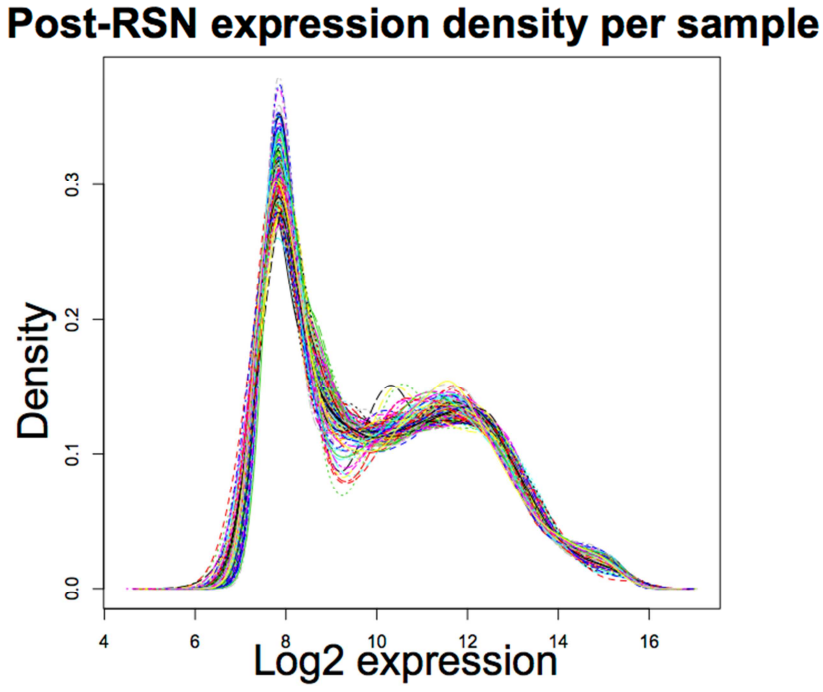


A.2.2 Density distributions of log-intensities of each sample before and after Affymetrix U133 plus 2 microarray gene expression data pre-processing (SkMel28)



A.2.3 Density plots of log-intensity distribution of each sample before and after HT WG-DASL gene expression data pre-processing





A.3 Illumina adapter sequences

The standard Illumina adapter sequences with 5 barcode bases and a T added.

5’ adapter: 5’- ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAACC*T-3’

3’ adapter: 5’-P-GGTTGAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3’

The highlighted bases represent the barcodes. The full list of the 40 different barcodes is shown in the table below (5’ adapter sequence shown).

Barcode	Sequence
1	CAACC-T
2	AACCA-T
3	AAGGA-T
4	AATTA-T
5	ACACA-T
6	GCATG-T
7	TCGAT-T
8	CGATC-T
9	AGCTA-T
10	GGTTG-T
11	TGCAT-T
12	GCCGG-T
13	GATCG-T
14	CCTTC-T
15	TACGT-T
16	TCAGT-T
17	CAGTC-T
18	CGTAC-T
19	TATAT-T
20	AGTCA-T
21	GAAGG-T
22	GACTG-T
23	CATGC-T
24	TCCTT-T
25	GCGCG-T
26	CCGGC-T
27	ACGTA-T
28	TCTCT-T
29	ACTGA-T
30	GGAAG-T
31	TAATT-T
32	CACAC-T
33	GAGAG-T
34	CCAAC-T
35	ACCAA-T
36	GCTAG-T
37	TGACT-T
38	AGAGA-T
39	GGCCG-T
40	CGCGC-T

A.4 Gene expression results in melanoma cell lines in response to vitamin D

A.4.1 Lists of differentially detected probes in response to vitamin D

These lists are in a Microsoft Office Excel spreadsheet on the attached cd under the name “Appendix A-4-1.xlxs”. Differentially detected probes after 6 hours of treatment with vitamin D, after 24 hours, after 48 hours, and the complete lists for each cell line (MeWo and SkMel28) are presented in separate worksheets within the Excel file.

A.4.2 Lists of predicted VDREs in the promoters of differentially expressed genes in response to vitamin D and of all annotated genes in the hg19

These lists are in a Microsoft Office Excel spreadsheet on the attached cd under the name “Appendix A-4-2.xlxs”. Lists of predicted VDREs in the promoters of differentially expressed genes (complete lists used) in response to vitamin D for MeWo and SkMel28 cells and of all annotated genes in the hg19 are presented in separate worksheets within the Excel file.

A.5 Publications

FILIA, A., ELLIOTT, F., WIND, T., FIELD, S., DAVIES, J., KUKALIZCH, K., RANDERSON-MOOR, J., HARLAND, M., BISHOP, D. T., BANKS, R. E. & NEWTON-BISHOP, J. A. 2013. Plasma osteopontin concentrations in patients with cutaneous melanoma. *Oncol Rep,* 30**,** 1575-80.

FILIA, A. Unravelling the mechanisms of vitamin D in the tumour microenvironment. *Cambridge Student Journal of Genetics*. 2013 Apr 16. Edition 1.

A.6 Presentations

FILIA, A., Copy number alterations in melanoma using NGS. *BioGenoMEL meeting, Leiden, Netherlands, Dec 2010*

FILIA, A., DNA extraction & library preparation for NGS. *BioGenoMEL meeting, Cambridge, UK, Nov 2012*

FILIA, A., Using vitamin D *in vitro. BioGenoMEL meeting, Cambridge, UK, Nov 2012*

A.7 Posters

FILIA, A., DROOP, A., THYGESEN, H., LAYE, J., JEWELL, R., SNOWDEN, H., HARLAND, M., BISHOP, D.T. & NEWTON-BISHOP, J.A. The successful generation of genomic data using Illumina HiSeq from small formalin-fixed paraffin-embedded primary melanomas: enabling the identification of predictive biomarkers from tumour samples stored in mature clinical trials. *Accepted for poster presentation, NCRI conference, Liverpool, UK, Nov 2013*

FILIA, A., DROOP, A., HARLAND, M., BISHOP, T., NEWTON-BISHOP, J. Using formalin-fixed paraffin-embedded melanoma tumors for the detection of copy number variation by next generation sequencing. [abstract]. In: Proceedings of the 104th Annual Meeting of the American Association for Cancer Research; 2013 Apr 6-10; Washington, DC. Philadelphia (PA): AACR; Cancer Res 2013;73 (8 Suppl):Abstract nr 3140.

FILIA, A., WIND, T., ELLIOT, F., FIELD, S., KUKALIZCH, K., RANDERSON-MOOR, J., HARLAND, M., BISHOP, D.T., BANKS, R.E. & NEWTON-BISHOP, J.A. Osteopontin as a prognostic biomarker for cutaneous malignant melanoma*, Genes & Cancer meeting, Warwick, UK, Dec 2011*

FILIA, A., WIND, T., ELLIOT, F., FIELD, S., KUKALIZCH, K., RANDERSON-MOOR, J., HARLAND, M., BISHOP, D.T., BANKS, R.E. & NEWTON-BISHOP, J.A. Osteopontin as a prognostic biomarker for cutaneous malignant melanoma*, International Melanoma Congress (SMR), Tampa, Florida, Nov 2011*

Chapter 9 References

ACLAND, K., EVANS, A. V., ABRAHA, H., HEALY, C. M., ROBLIN, P., CALONJE, E., ORCHARD, G., HIGGINS, E., SHERWOOD, R. & RUSSELL-JONES, R. 2002. Serum S100 concentrations are not useful in predicting micrometastatic disease in cutaneous malignant melanoma. *Br J Dermatol,* 146**,** 832-5.

ADAM, Z., ADANY, R., LADANYI, A., TIMAR, J. & BALAZS, M. 2000. Liver metastatic ability of human melanoma cell line is associated with losses of chromosomes 4, 9p21-pter and 10p. *Clin Exp Metastasis,* 18**,** 295-302.

ADAMS, M. D., VEIGL, M. L., WANG, Z., MOLYNEUX, N., SUN, S., GUDA, K., YU, X., MARKOWITZ, S. D. & WILLIS, J. 2012. Global mutational profiling of formalin-fixed human colon cancers from a pathology archive. *Mod Pathol,* 25**,** 1599-608.

AFFYMETRIX. 2009. *GeneChip expression analysis technical manual* [Online]. Santa Clara, CA, USA: Affymetrix. Available: http://media.affymetrix.com/support/downloads/manuals/expression_analysis_technical_manual.pdf [Accessed 10th February 2013].

AGARWALA, S. S., KEILHOLZ, U., GILLES, E., BEDIKIAN, A. Y., WU, J., KAY, R., STEIN, C. A., ITRI, L. M., SUCIU, S. & EGGERMONT, A. M. 2009. LDH correlation with survival in advanced melanoma from two large, randomised trials (Oblimersen GM301 and EORTC 18951). *Eur J Cancer,* 45**,** 1807-14.

AIRD, D., ROSS, M. G., CHEN, W. S., DANIELSSON, M., FENNELL, T., RUSS, C., JAFFE, D. B., NUSBAUM, C. & GNIRKE, A. 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol,* 12**,** R18.

ALONSO, S. R., TRACEY, L., ORTIZ, P., PEREZ-GOMEZ, B., PALACIOS, J., POLLAN, M., LINARES, J., SERRANO, S., SAEZ-CASTILLO, A. I., SANCHEZ, L., PAJARES, R., SANCHEZ-AGUILERA, A., ARTIGA, M. J., PIRIS, M. A. & RODRIGUEZ-PERALTO, J. L. 2007. A high-throughput study in melanoma identifies epithelial-mesenchymal transition as a major determinant of metastasis. *Cancer Res,* 67**,** 3450-60.

AMERICAN TYPE CULTURE COLLECTION STANDARDS DEVELOPMENT ORGANIZATION WORKGROUP, A. S. N. 2010. Cell line misidentification: the beginning of the end. *Nat Rev Cancer,* 10**,** 441-8.

ASHBURNER, M., BALL, C. A., BLAKE, J. A., BOTSTEIN, D., BUTLER, H., CHERRY, J. M., DAVIS, A. P., DOLINSKI, K., DWIGHT, S. S., EPPIG, J. T., HARRIS, M. A., HILL, D. P., ISSEL-TARVER, L., KASARSKIS, A., LEWIS, S., MATESE, J. C., RICHARDSON, J. E., RINGWALD, M., RUBIN, G. M. & SHERLOCK, G. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet,* 25**,** 25-9.

AYACHE, S., PANELLI, M. C., BYRNE, K. M., SLEZAK, S., LEITMAN, S. F., MARINCOLA, F. M. & STRONCEK, D. F. 2006. Comparison of proteomic profiles of serum, plasma, and modified media supplements used for cell culture and expansion. *J Transl Med,* 4**,** 40.

BABRAHAM BIOINFORMATICS. 2010. *FastQC* [Online]. Available: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

BALABAN, G., HERLYN, M., GUERRY, D. T., BARTOLO, R., KOPROWSKI, H., CLARK, W. H. & NOWELL, P. C. 1984. Cytogenetics of human malignant melanoma and premalignant lesions. *Cancer Genet Cytogenet,* 11**,** 429-39.

BALCH, C. M., BUZAID, A. C., SOONG, S. J., ATKINS, M. B., CASCINELLI, N., COIT, D. G., FLEMING, I. D., GERSHENWALD, J. E., HOUGHTON, A., JR., KIRKWOOD, J. M., MCMASTERS, K. M., MIHM, M. F., MORTON, D. L., REINTGEN, D. S., ROSS, M. I., SOBER, A., THOMPSON, J. A. &

THOMPSON, J. F. 2001a. Final version of the American Joint Committee on Cancer staging system for cutaneous melanoma. *J Clin Oncol,* 19**,** 3635-48.

BALCH, C. M., GERSHENWALD, J. E., SOONG, S. J., THOMPSON, J. F., ATKINS, M. B., BYRD, D. R., BUZAID, A. C., COCHRAN, A. J., COIT, D. G., DING, S., EGGERMONT, A. M., FLAHERTY, K. T., GIMOTTY, P. A., KIRKWOOD, J. M., MCMASTERS, K. M., MIHM, M. C., JR., MORTON, D. L., ROSS, M. I., SOBER, A. J. & SONDAK, V. K. 2009. Final version of 2009 AJCC melanoma staging and classification. *J Clin Oncol,* 27**,** 6199-206.

BALCH, C. M., GERSHENWALD, J. E., SOONG, S. J., THOMPSON, J. F., DING, S., BYRD, D. R., CASCINELLI, N., COCHRAN, A. J., COIT, D. G., EGGERMONT, A. M., JOHNSON, T., KIRKWOOD, J. M., LEONG, S. P., MCMASTERS, K. M., MIHM, M. C., JR., MORTON, D. L., ROSS, M. I. & SONDAK, V. K. 2010. Multivariate analysis of prognostic factors among 2,313 patients with stage III melanoma: comparison of nodal micrometastases versus macrometastases. *J Clin Oncol,* 28**,** 2452-9.

BALCH, C. M., SOONG, S. J., GERSHENWALD, J. E., THOMPSON, J. F., REINTGEN, D. S., CASCINELLI, N., URIST, M., MCMASTERS, K. M., ROSS, M. I., KIRKWOOD, J. M., ATKINS, M. B., THOMPSON, J. A., COIT, D. G., BYRD, D., DESMOND, R., ZHANG, Y., LIU, P. Y., LYMAN, G. H. & MORABITO, A. 2001b. Prognostic factors analysis of 17,600 melanoma patients: validation of the American Joint Committee on Cancer melanoma staging system. *J Clin Oncol,* 19**,** 3622-34.

BARAK, V., KAISERMAN, I., FRENKEL, S., HENDLER, K., KALICKMAN, I. & PE'ER, J. 2011. The dynamics of serum tumor markers in predicting metastatic uveal melanoma (part 1). *Anticancer Res,* 31**,** 345-9.

BARRETT, J. H., ILES, M. M., HARLAND, M., TAYLOR, J. C., AITKEN, J. F., ANDRESEN, P. A., AKSLEN, L. A., ARMSTRONG, B. K., AVRIL, M. F., AZIZI, E., BAKKER, B., BERGMAN, W., BIANCHI-SCARRA, G., BRESSAC-DE PAILLERETS, B., CALISTA, D., CANNON-ALBRIGHT, L. A., CORDA, E., CUST, A. E., DEBNIAK, T., DUFFY, D., DUNNING, A. M., EASTON, D. F., FRIEDMAN, E., GALAN, P., GHIORZO, P., GILES, G. G., HANSSON, J., HOCEVAR, M., HOIOM, V., HOPPER, J. L., INGVAR, C., JANSSEN, B., JENKINS, M. A., JONSSON, G., KEFFORD, R. F., LANDI, G., LANDI, M. T., LANG, J., LUBINSKI, J., MACKIE, R., MALVEHY, J., MARTIN, N. G., MOLVEN, A., MONTGOMERY, G. W., VAN NIEUWPOORT, F. A., NOVAKOVIC, S., OLSSON, H., PASTORINO, L., PUIG, S., PUIG-BUTILLE, J. A., RANDERSON-MOOR, J., SNOWDEN, H., TUOMINEN, R., VAN BELLE, P., VAN DER STOEP, N., WHITEMAN, D. C., ZELENIKA, D., HAN, J., FANG, S., LEE, J. E., WEI, Q., LATHROP, G. M., GILLANDERS, E. M., BROWN, K. M., GOLDSTEIN, A. M., KANETSKY, P. A., MANN, G. J., MACGREGOR, S., ELDER, D. E., AMOS, C. I., HAYWARD, N. K., GRUIS, N. A., DEMENAIS, F., BISHOP, J. A., BISHOP, D. T. & GENO, M. E. L. C. 2011. Genome-wide association study identifies three new melanoma susceptibility loci. *Nat Genet,* 43**,** 1108-13.

BARTKOVA, J., LUKAS, J., GULDBERG, P., ALSNER, J., KIRKIN, A. F., ZEUTHEN, J. & BARTEK, J. 1996. The p16-cyclin D/Cdk4-pRb pathway as a functional unit frequently altered in melanoma pathogenesis. *Cancer Res,* 56**,** 5475-83.

BASTIAN, B. C. 2003. Understanding the progression of melanocytic neoplasia using genomic analysis: from fields to cancer. *Oncogene,* 22**,** 3081-6.

BASTIAN, B. C., KASHANI-SABET, M., HAMM, H., GODFREY, T., MOORE, D. H., 2ND, BROCKER, E. B., LEBOIT, P. E. & PINKEL, D. 2000. Gene amplifications characterize acral melanoma and permit the detection of occult tumor cells in the surrounding skin. *Cancer Res,* 60**,** 1968-73.

BASTIAN, B. C., OLSHEN, A. B., LEBOIT, P. E. & PINKEL, D. 2003. Classifying melanocytic tumors based on DNA copy number changes. *Am J Pathol,* 163**,** 1765-70.

BELLAHCENE, A., CASTRONOVO, V., OGBUREKE, K. U., FISHER, L. W. & FEDARKO, N. S. 2008. Small integrin-binding ligand N-linked glycoproteins (SIBLINGs): multifunctional proteins in cancer. *Nat Rev Cancer,* 8**,** 212-26.

BELVEDERE, O., BERRI, S., CHALKLEY, R., CONWAY, C., BARBONE, F., PISA, F., MACLENNAN, K., DALY, C., ALSOP, M., MORGAN, J., MENIS, J., TCHERVENIAKOV, P., PAPAGIANNOPOULOS, K., RABBITTS, P. & WOOD, H. M. 2012. A computational index derived from whole-genome copy number analysis is a novel tool for prognosis in early stage lung squamous cell carcinoma. *Genomics,* 99**,** 18-24.

BENJAMINI, Y. & HOCHBERG, Y. 1995. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological,* 57**,** 289-300.

BENJAMINI, Y. & SPEED, T. P. 2012. Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res,* 40**,** e72.

BERGMAN, R., BRUCKER-TUDERMAN, S., HERCOGOVA, J. & BASTIAN, B. C. 2006. Melanocytic tumours. Nodular melanoma. *World Health Organisation Classification of Tumours. Pathology and Genetics of Skin Tumours.* Lyon, France: IARC press.

BEROUKHIM, R., MERMEL, C. H., PORTER, D., WEI, G., RAYCHAUDHURI, S., DONOVAN, J., BARRETINA, J., BOEHM, J. S., DOBSON, J., URASHIMA, M., MC HENRY, K. T., PINCHBACK, R. M., LIGON, A. H., CHO, Y. J., HAERY, L., GREULICH, H., REICH, M., WINCKLER, W., LAWRENCE, M. S., WEIR, B. A., TANAKA, K. E., CHIANG, D. Y., BASS, A. J., LOO, A., HOFFMAN, C., PRENSNER, J., LIEFELD, T., GAO, Q., YECIES, D., SIGNORETTI, S., MAHER, E., KAYE, F. J., SASAKI, H., TEPPER, J. E., FLETCHER, J. A., TABERNERO, J., BASELGA, J., TSAO, M. S., DEMICHELIS, F., RUBIN, M. A., JANNE, P. A., DALY, M. J., NUCERA, C., LEVINE, R. L., EBERT, B. L., GABRIEL, S., RUSTGI, A. K., ANTONESCU, C. R., LADANYI, M., LETAI, A., GARRAWAY, L. A., LODA, M., BEER, D. G., TRUE, L. D., OKAMOTO, A., POMEROY, S. L., SINGER, S., GOLUB, T. R., LANDER, E. S., GETZ, G., SELLERS, W. R. & MEYERSON, M. 2010. The landscape of somatic copy-number alteration across human cancers. *Nature,* 463**,** 899-905.

BIBIKOVA, M., TALANTOV, D., CHUDIN, E., YEAKLEY, J. M., CHEN, J., DOUCET, D., WICKHAM, E., ATKINS, D., BARKER, D., CHEE, M., WANG, Y. & FAN, J. B. 2004. Quantitative gene expression profiling in formalin-fixed, paraffin-embedded tissues using universal bead arrays. *Am J Pathol,* 165**,** 1799-807.

BISHOP, D. T., DEMENAIS, F., ILES, M. M., HARLAND, M., TAYLOR, J. C., CORDA, E., RANDERSON-MOOR, J., AITKEN, J. F., AVRIL, M. F., AZIZI, E., BAKKER, B., BIANCHI-SCARRA, G., BRESSAC-DE PAILLERETS, B., CALISTA, D., CANNON-ALBRIGHT, L. A., CHIN, A. W. T., DEBNIAK, T., GALORE-HASKEL, G., GHIORZO, P., GUT, I., HANSSON, J., HOCEVAR, M., HOIOM, V., HOPPER, J. L., INGVAR, C., KANETSKY, P. A., KEFFORD, R. F., LANDI, M. T., LANG, J., LUBINSKI, J., MACKIE, R., MALVEHY, J., MANN, G. J., MARTIN, N. G., MONTGOMERY, G. W., VAN NIEUWPOORT, F. A., NOVAKOVIC, S., OLSSON, H., PUIG, S., WEISS, M., VAN WORKUM, W., ZELENIKA, D., BROWN, K. M., GOLDSTEIN, A. M., GILLANDERS, E. M., BOLAND, A., GALAN, P., ELDER, D. E., GRUIS, N. A., HAYWARD, N. K., LATHROP, G. M., BARRETT, J. H. & BISHOP, J. A. 2009. Genome-wide association study identifies three loci associated with melanoma risk. *Nat Genet,* 41**,** 920-5.

BLASBERG, J. D., PASS, H. I., GOPARAJU, C. M., FLORES, R. M., LEE, S. & DONINGTON, J. S. 2010. Reduction of elevated plasma osteopontin levels with resection of non-small-cell lung cancer. *J Clin Oncol,* 28**,** 936-41.

BOLSTAD, B. M., COLLIN, F., BRETTSCHNEIDER, J., SIMPSON, K., COPE, L., IRIZARRY, R. A. & SPEED, T. P. 2005. Quality Assessment of Affymetrix GeneChip Data. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor.* New York: Springer.

BOLSTAD, B. M., IRIZARRY, R. A., ASTRAND, M. & SPEED, T. P. 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics,* 19**,** 185-93.

BOONE, B., JACOBS, K., FERDINANDE, L., TAILDEMAN, J., LAMBERT, J., PEETERS, M., BRACKE, M., PAUWELS, P. & BROCHEZ, L. 2011. EGFR in melanoma: clinical significance and potential therapeutic target. *J Cutan Pathol,* 38**,** 492-502.

BOSSERHOFF, A. K. 2006. Novel biomarkers in malignant melanoma. *Clin Chim Acta,* 367**,** 28-35.

BRESLOW, A. 1970. Thickness, cross-sectional areas and depth of invasion in the prognosis of cutaneous melanoma. *Ann Surg,* 172**,** 902-8.

BROZYNA, A. A., JOZWICKI, W., JANJETOVIC, Z. & SLOMINSKI, A. T. 2011. Expression of vitamin D receptor decreases during progression of pigmented skin lesions. *Hum Pathol,* 42**,** 618-31.

BRYNE, J. C., VALEN, E., TANG, M. H., MARSTRAND, T., WINTHER, O., DA PIEDADE, I., KROGH, A., LENHARD, B. & SANDELIN, A. 2008. JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Res,* 36**,** D102-6.

BUSAM, K. J., ZHAO, H., COIT, D. G., KUCUKGOL, D., JUNGBLUTH, A. A., NOBREGA, J. & VIALE, A. 2005. Distinction of desmoplastic melanoma from non-desmoplastic melanoma by gene expression profiling. *J Invest Dermatol,* 124**,** 412-8.

CAMPBELL, F. C., XU, H., EL-TANANI, M., CROWE, P. & BINGHAM, V. 2010. The yin and yang of vitamin D receptor (VDR) signaling in neoplastic progression: operational networks and tissue-specific growth control. *Biochem Pharmacol,* 79**,** 1-9.

CANCER RESEARCH UK STATISTICAL INFORMATION TEAM. 2010a. *Skin cancer incidence statistics [online]* [Online]. Available: http://info.cancerresearchuk.org/cancerstats/types/skin/incidence [Accessed 29th August 2013].

CANCER RESEARCH UK STATISTICAL INFORMATION TEAM. 2010b. *Skin cancer survival statistics [online]* [Online]. Available: http://www.cancerresearchuk.org/cancer-info/cancerstats/types/skin/survival/ [Accessed 2nd September 2013].

CARRINGTON, M. N., THARP-HILTBOLD, B., KNOTH, J. & WARD, F. E. 1988. 1,25-Dihydroxyvitamin D3 decreases expression of HLA class II molecules in a melanoma cell line. *J Immunol,* 140**,** 4013-8.

CHANG, Y. M., BARRETT, J. H., BISHOP, D. T., ARMSTRONG, B. K., BATAILLE, V., BERGMAN, W., BERWICK, M., BRACCI, P. M., ELWOOD, J. M., ERNSTOFF, M. S., GALLAGHER, R. P., GREEN, A. C., GRUIS, N. A., HOLLY, E. A., INGVAR, C., KANETSKY, P. A., KARAGAS, M. R., LEE, T. K., LE MARCHAND, L., MACKIE, R. M., OLSSON, H., OSTERLIND, A., REBBECK, T. R., SASIENI, P., SISKIND, V., SWERDLOW, A. J., TITUS-ERNSTOFF, L., ZENS, M. S. & NEWTON-BISHOP, J. A. 2009. Sun exposure and melanoma risk at different latitudes: a pooled analysis of 5700 cases and 7216 controls. *Int J Epidemiol,* 38**,** 814-30.

CHAPMAN, P. B., HAUSCHILD, A., ROBERT, C., HAANEN, J. B., ASCIERTO, P., LARKIN, J., DUMMER, R., GARBE, C., TESTORI, A., MAIO, M., HOGG, D., LORIGAN, P., LEBBE, C., JOUARY, T., SCHADENDORF, D., RIBAS, A.,

O'DAY, S. J., SOSMAN, J. A., KIRKWOOD, J. M., EGGERMONT, A. M., DRENO, B., NOLOP, K., LI, J., NELSON, B., HOU, J., LEE, R. J., FLAHERTY, K. T., MCARTHUR, G. A. & GROUP, B.-S. 2011. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N Engl J Med,* 364**,** 2507-16.

CHATZINASIOU, F., LILL, C. M., KYPREOU, K., STEFANAKI, I., NICOLAOU, V., SPYROU, G., EVANGELOU, E., ROEHR, J. T., KODELA, E., KATSAMBAS, A., TSAO, H., IOANNIDIS, J. P., BERTRAM, L. & STRATIGOS, A. J. 2011. Comprehensive field synopsis and systematic meta-analyses of genetic association studies in cutaneous melanoma. *J Natl Cancer Inst,* 103**,** 1227-35.

CHEN, P., HU, P., XIE, D., QIN, Y., WANG, F. & WANG, H. 2010. Meta-analysis of vitamin D, calcium and the prevention of breast cancer. *Breast Cancer Res Treat,* 121**,** 469-77.

CHEN, Y. C., LIU, T., YU, C. H., CHIANG, T. Y. & HWANG, C. C. 2013. Effects of GC bias in next-generation-sequencing data on de novo genome assembly. *PLoS One,* 8**,** e62856.

CHENG, X., ZHAO, X., KHURANA, S., BRUGGEMAN, L. A. & KAO, H. Y. 2013. Microarray analyses of glucocorticoid and vitamin D3 target genes in differentiating cultured human podocytes. *PLoS One,* 8**,** e60213.

CHERNOFF, K. A., BORDONE, L., HORST, B., SIMON, K., TWADELL, W., LEE, K., COHEN, J. A., WANG, S., SILVERS, D. N., BRUNNER, G. & CELEBI, J. T. 2009. GAB2 amplifications refine molecular classification of melanoma. *Clin Cancer Res,* 15**,** 4288-91.

CHIN, L. 2003. The genetics of malignant melanoma: lessons from mouse and man. *Nat Rev Cancer,* 3**,** 559-70.

CHIN, L., MERLINO, G. & DEPINHO, R. A. 1998. Malignant melanoma: modern black plague and genetic black box. *Genes Dev,* 12**,** 3467-81.

CHO, H., KANG, E. S., KIM, Y. T. & KIM, J. H. 2009. Diagnostic and prognostic impact of osteopontin expression in endometrial cancer. *Cancer Invest,* 27**,** 313-23.

CHURCH, D. M., SCHNEIDER, V. A., GRAVES, T., AUGER, K., CUNNINGHAM, F., BOUK, N., CHEN, H. C., AGARWALA, R., MCLAREN, W. M., RITCHIE, G. R., ALBRACHT, D., KREMITZKI, M., ROCK, S., KOTKIEWICZ, H., KREMITZKI, C., WOLLAM, A., TRANI, L., FULTON, L., FULTON, R., MATTHEWS, L., WHITEHEAD, S., CHOW, W., TORRANCE, J., DUNN, M., HARDEN, G., THREADGOLD, G., WOOD, J., COLLINS, J., HEATH, P., GRIFFITHS, G., PELAN, S., GRAFHAM, D., EICHLER, E. E., WEINSTOCK, G., MARDIS, E. R., WILSON, R. K., HOWE, K., FLICEK, P. & HUBBARD, T. 2011. Modernizing reference genome assemblies. *PLoS Biol,* 9**,** e1001091.

CLARK, W. H., JR., ELDER, D. E., GUERRY, D. T., EPSTEIN, M. N., GREENE, M. H. & VAN HORN, M. 1984. A study of tumor progression: the precursor lesions of superficial spreading and nodular melanoma. *Hum Pathol,* 15**,** 1147-65.

CLARK, W. H., JR., FROM, L., BERNARDINO, E. A. & MIHM, M. C. 1969. The histogenesis and biologic behavior of primary human malignant melanomas of the skin. *Cancer Res,* 29**,** 705-27.

COLOMBINO, M., CAPONE, M., LISSIA, A., COSSU, A., RUBINO, C., DE GIORGI, V., MASSI, D., FONSATTI, E., STAIBANO, S., NAPPI, O., PAGANI, E., CASULA, M., MANCA, A., SINI, M., FRANCO, R., BOTTI, G., CARACO, C., MOZZILLO, N., ASCIERTO, P. A. & PALMIERI, G. 2012. BRAF/NRAS mutation frequencies among primary tumors and metastases in patients with melanoma. *J Clin Oncol,* 30**,** 2522-9.

CONSORTIUM, E. P., BERNSTEIN, B. E., BIRNEY, E., DUNHAM, I., GREEN, E. D., GUNTER, C. & SNYDER, M. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature,* 489**,** 57-74.

CONWAY, C., BESWICK, S., ELLIOTT, F., CHANG, Y. M., RANDERSON-MOOR, J., HARLAND, M., AFFLECK, P., MARSDEN, J., SANDERS, D. S., BOON, A., KNOWLES, M. A., BISHOP, D. T. & NEWTON-BISHOP, J. A. 2010. Deletion at chromosome arm 9p in relation to BRAF/NRAS mutations and prognostic significance for primary melanoma. *Genes Chromosomes Cancer,* 49**,** 425-38.

CONWAY, C., CHALKLEY, R., HIGH, A., MACLENNAN, K., BERRI, S., CHENGOT, P., ALSOP, M., EGAN, P., MORGAN, J., TAYLOR, G. R., CHESTER, J., SEN, M., RABBITTS, P. & WOOD, H. M. 2012. Next-generation sequencing for simultaneous determination of human papillomavirus load, subtype, and associated genomic copy number changes in tumors. *J Mol Diagn,* 14**,** 104-11.

CONWAY, C., MITRA, A., JEWELL, R., RANDERSON-MOOR, J., LOBO, S., NSENGIMANA, J., EDWARD, S., SANDERS, D. S., COOK, M., POWELL, B., BOON, A., ELLIOTT, F., DE KORT, F., KNOWLES, M. A., BISHOP, D. T. & NEWTON-BISHOP, J. 2009. Gene expression profiling of paraffin-embedded primary melanoma using the DASL assay identifies increased osteopontin expression as predictive of reduced relapse-free survival. *Clin Cancer Res,* 15**,** 6939-46.

COSTA, J. L., EIJK, P. P., VAN DE WIEL, M. A., TEN BERGE, D., SCHMITT, F., NARVAEZ, C. J., WELSH, J. & YLSTRA, B. 2009. Anti-proliferative action of vitamin D in MCF7 is still active after siRNA-VDR knock-down. *BMC Genomics,* 10**,** 499.

COWAN, J. M., HALABAN, R. & FRANCKE, U. 1988. Cytogenetic analysis of melanocytes from premalignant nevi and melanomas. *J Natl Cancer Inst,* 80**,** 1159-64.

CURTIN, J. A., BUSAM, K., PINKEL, D. & BASTIAN, B. C. 2006. Somatic activation of KIT in distinct subtypes of melanoma. *J Clin Oncol,* 24**,** 4340-6.

CURTIN, J. A., FRIDLYAND, J., KAGESHITA, T., PATEL, H. N., BUSAM, K. J., KUTZNER, H., CHO, K. H., AIBA, S., BROCKER, E. B., LEBOIT, P. E., PINKEL, D. & BASTIAN, B. C. 2005. Distinct sets of genetic alterations in melanoma. *N Engl J Med,* 353**,** 2135-47.

DANIELSSON, C., FEHSEL, K., POLLY, P. & CARLBERG, C. 1998. Differential apoptotic response of human melanoma cells to 1 alpha,25-dihydroxyvitamin D3 and its analogues. *Cell Death Differ,* 5**,** 946-52.

DANIELSSON, C., TORMA, H., VAHLQUIST, A. & CARLBERG, C. 1999. Positive and negative interaction of 1,25-dihydroxyvitamin D3 and the retinoid CD437 in the induction of human melanoma cell apoptosis. *Int J Cancer,* 81**,** 467-70.

DAS, R., PHILIP, S., MAHABELESHWAR, G. H., BULBULE, A. & KUNDU, G. C. 2005. Osteopontin: it's role in regulation of cell motility and nuclear factor kappa B-mediated urokinase type plasminogen activator expression. *IUBMB Life,* 57**,** 441-7.

DAVAR, D., TARHINI, A. A. & KIRKWOOD, J. M. 2012. Adjuvant therapy for melanoma. *Cancer J,* 18**,** 192-202.

DAVIES, H., BIGNELL, G. R., COX, C., STEPHENS, P., EDKINS, S., CLEGG, S., TEAGUE, J., WOFFENDIN, H., GARNETT, M. J., BOTTOMLEY, W., DAVIS, N., DICKS, E., EWING, R., FLOYD, Y., GRAY, K., HALL, S., HAWES, R., HUGHES, J., KOSMIDOU, V., MENZIES, A., MOULD, C., PARKER, A., STEVENS, C., WATT, S., HOOPER, S., WILSON, R., JAYATILAKE, H., GUSTERSON, B. A., COOPER, C., SHIPLEY, J., HARGRAVE, D., PRITCHARD-JONES, K., MAITLAND, N., CHENEVIX-TRENCH, G., RIGGINS, G. J., BIGNER, D. D., PALMIERI, G., COSSU, A.,

FLANAGAN, A., NICHOLSON, A., HO, J. W., LEUNG, S. Y., YUEN, S. T., WEBER, B. L., SEIGLER, H. F., DARROW, T. L., PATERSON, H., MARAIS, R., MARSHALL, C. J., WOOSTER, R., STRATTON, M. R. & FUTREAL, P. A. 2002. Mutations of the BRAF gene in human cancer. *Nature,* 417**,** 949-54.

DEEB, K. K., TRUMP, D. L. & JOHNSON, C. S. 2007. Vitamin D signalling pathways in cancer: potential for anticancer therapeutics. *Nat Rev Cancer,* 7**,** 684-700.

DEGENHARDT, Y., HUANG, J., GRESHOCK, J., HORIATES, G., NATHANSON, K., YANG, X., HERLYN, M. & WEBER, B. 2010. Distinct MHC gene expression patterns during progression of melanoma. *Genes Chromosomes Cancer,* 49**,** 144-54.

DEICHMANN, M., BENNER, A., BOCK, M., JACKEL, A., UHL, K., WALDMANN, V. & NAHER, H. 1999. S100-Beta, melanoma-inhibiting activity, and lactate dehydrogenase discriminate progressive from nonprogressive American Joint Committee on Cancer stage IV melanoma. *J Clin Oncol,* 17**,** 1891-6.

DESILVA, B., SMITH, W., WEINER, R., KELLEY, M., SMOLEC, J., LEE, B., KHAN, M., TACEY, R., HILL, H. & CELNIKER, A. 2003. Recommendations for the bioanalytical method validation of ligand-binding assays to support pharmacokinetic assessments of macromolecules. *Pharm Res,* 20**,** 1885-900.

DEWAR, D. J., NEWELL, B., GREEN, M. A., TOPPING, A. P., POWELL, B. W. & COOK, M. G. 2004. The microanatomic location of metastatic melanoma in sentinel lymph nodes predicts nonsentinel lymph node involvement. *J Clin Oncol,* 22**,** 3345-9.

DING, L., WENDL, M. C., KOBOLDT, D. C. & MARDIS, E. R. 2010. Analysis of next-generation genomic data in cancer: accomplishments and challenges. *Hum Mol Genet,* 19**,** R188-96.

DOHM, J. C., LOTTAZ, C., BORODINA, T. & HIMMELBAUER, H. 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res,* 36**,** e105.

DU, P., KIBBE, W. A. & LIN, S. M. 2008. lumi: a pipeline for processing Illumina microarray. *Bioinformatics,* 24**,** 1547-8.

DUAN, J., ZHANG, J. G., DENG, H. W. & WANG, Y. P. 2013. Comparative studies of copy number variation detection methods for next-generation sequencing technologies. *PLoS One,* 8**,** e59128.

ECKHART, L., BACH, J., BAN, J. & TSCHACHLER, E. 2000. Melanin binds reversibly to thermostable DNA polymerase and inhibits its activity. *Biochem Biophys Res Commun,* 271**,** 726-30.

EGGERMONT, A. M. & ROBERT, C. 2011. New drugs in melanoma: it's a whole new world. *Eur J Cancer,* 47**,** 2150-7.

EGGERMONT, A. M. & ROBERT, C. 2012. Melanoma in 2011: a new paradigm tumor for drug development. *Nat Rev Clin Oncol,* 9**,** 74-6.

EGGERMONT, A. M., TESTORI, A., MARSDEN, J., HERSEY, P., QUIRT, I., PETRELLA, T., GOGAS, H., MACKIE, R. M. & HAUSCHILD, A. 2009. Utility of adjuvant systemic therapy in melanoma. *Ann Oncol,* 20 Suppl 6**,** vi30-4.

EILERS, P. H. & GOEMAN, J. J. 2004. Enhancing scatterplots with smoothed densities. *Bioinformatics,* 20**,** 623-8.

EISMAN, J. A., BARKLA, D. H. & TUTTON, P. J. 1987. Suppression of in vivo growth of human cancer solid tumor xenografts by 1,25-dihydroxyvitamin D3. *Cancer Res,* 47**,** 21-5.

EKINS, S., NIKOLSKY, Y., BUGRIM, A., KIRILLOV, E. & NIKOLSKAYA, T. 2007. Pathway mapping tools for analysis of high content data. *Methods Mol Biol,* 356**,** 319-50.

EL-TANANI, M. K., CAMPBELL, F. C., KURISETTY, V., JIN, D., MCCANN, M. & RUDLAND, P. S. 2006. The regulation and role of osteopontin in malignant transformation and cancer. *Cytokine Growth Factor Rev,* 17**,** 463-74.

ELDER, D. E. & MURPHY, G. F. 2010. Malignant tumors (melanoma and related lesions). *Atlas of tumor pathology, fourth series: Melanocytic tumors of the skin.* Washington: Armed Forces Institute of Pathology.

ESSA, S., DENZER, N., MAHLKNECHT, U., KLEIN, R., COLLNOT, E. M., TILGEN, W. & REICHRATH, J. 2010. VDR microRNA expression and epigenetic silencing of vitamin D signaling in melanoma cells. *J Steroid Biochem Mol Biol,* 121**,** 110-3.

EVANS, S. R., HOUGHTON, A. M., SCHUMAKER, L., BRENNER, R. V., BURAS, R. R., DAVOODI, F., NAUTA, R. J. & SHABAHANG, M. 1996. Vitamin D receptor and growth inhibition by 1,25-dihydroxyvitamin D3 in human malignant melanoma cell lines. *J Surg Res,* 61**,** 127-33.

EWING, B. & GREEN, P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res,* 8**,** 186-94.

FAOUR, W. H., HE, Q., MANCINI, A., JOVANOVIC, D., ANTONIOU, J. & DI BATTISTA, J. A. 2006. Prostaglandin E2 stimulates p53 transactivational activity through specific serine 15 phosphorylation in human synovial fibroblasts. Role in suppression of c/EBP/NF-kappaB-mediated MEKK1-induced MMP-1 expression. *J Biol Chem,* 281**,** 19849-60.

FIELD, S. & NEWTON-BISHOP, J. A. 2011. Melanoma and vitamin D. *Mol Oncol,* 5**,** 197-214.

FILIA, A., ELLIOTT, F., WIND, T., FIELD, S., DAVIES, J., KUKALIZCH, K., RANDERSON-MOOR, J., HARLAND, M., BISHOP, D. T., BANKS, R. E. & NEWTON-BISHOP, J. A. 2013. Plasma osteopontin concentrations in patients with cutaneous melanoma. *Oncol Rep,* 30**,** 1575-80.

FINDEISEN, P., ZAPATKA, M., PECCERELLA, T., MATZK, H., NEUMAIER, M., SCHADENDORF, D. & UGUREL, S. 2009. Serum amyloid A as a prognostic marker in melanoma identified by proteomic profiling. *J Clin Oncol,* 27**,** 2199-208.

FISHER, L. W., TORCHIA, D. A., FOHR, B., YOUNG, M. F. & FEDARKO, N. S. 2001. Flexible structures of SIBLING proteins, bone sialoprotein, and osteopontin. *Biochem Biophys Res Commun,* 280**,** 460-5.

FISHMAN, P., MADI, L., BAR-YEHUDA, S., BARER, F., DEL VALLE, L. & KHALILI, K. 2002. Evidence for involvement of Wnt signaling pathway in IB-MECA mediated suppression of melanoma cells. *Oncogene,* 21**,** 4060-4.

FLAHERTY, K. T., PUZANOV, I., KIM, K. B., RIBAS, A., MCARTHUR, G. A., SOSMAN, J. A., O'DWYER, P. J., LEE, R. J., GRIPPO, J. F., NOLOP, K. & CHAPMAN, P. B. 2010. Inhibition of mutated, activated BRAF in metastatic melanoma. *N Engl J Med,* 363**,** 809-19.

FLEET, J. C., DESMET, M., JOHNSON, R. & LI, Y. 2012. Vitamin D and cancer: a review of molecular mechanisms. *Biochem J,* 441**,** 61-76.

FOOD AND DRUG ADMINISTRATION 2001. Guidance for industry: Bioanalytical Method Validation. US Department of Health and Human Services, FDA, Centre for Drug Evaluation and Research, Rockville, MD, 2001.

FRAMPTON, R. J., OMOND, S. A. & EISMAN, J. A. 1983. Inhibition of human cancer cell growth by 1,25-dihydroxyvitamin D3 metabolites. *Cancer Res,* 43**,** 4443-7.

FRANKEL, A. 2012. Formalin fixation in the '-omics' era: a primer for the surgeon-scientist. *ANZ J Surg,* 82**,** 395-402.

FRANZEN, A. & HEINEGARD, D. 1985. Isolation and characterization of two sialoproteins present only in bone calcified matrix. *Biochem J,* 232**,** 715-24.

FURNEY, S. J., TURAJLIC, S., STAMP, G., NOHADANI, M., CARLISLE, A., THOMAS, J. M., HAYES, A., STRAUSS, D., GORE, M., VAN DEN OORD, J., LARKIN, J. & MARAIS, R. 2013. Genome sequencing of mucosal

melanomas reveals that they are driven by distinct mechanisms from cutaneous melanoma. *J Pathol,* 230**,** 261-9.

GANDINI, S., SERA, F., CATTARUZZA, M. S., PASQUINI, P., ABENI, D., BOYLE, P. & MELCHI, C. F. 2005a. Meta-analysis of risk factors for cutaneous melanoma: I. Common and atypical naevi. *Eur J Cancer,* 41**,** 28-44.

GANDINI, S., SERA, F., CATTARUZZA, M. S., PASQUINI, P., PICCONI, O., BOYLE, P. & MELCHI, C. F. 2005b. Meta-analysis of risk factors for cutaneous melanoma: II. Sun exposure. *Eur J Cancer,* 41**,** 45-60.

GANDINI, S., SERA, F., CATTARUZZA, M. S., PASQUINI, P., ZANETTI, R., MASINI, C., BOYLE, P. & MELCHI, C. F. 2005c. Meta-analysis of risk factors for cutaneous melanoma: III. Family history, actinic damage and phenotypic factors. *Eur J Cancer,* 41**,** 2040-59.

GARLAND, C. F. & GARLAND, F. C. 1980. Do sunlight and vitamin D reduce the likelihood of colon cancer? *Int J Epidemiol,* 9**,** 227-31.

GARRAWAY, L. A., WIDLUND, H. R., RUBIN, M. A., GETZ, G., BERGER, A. J., RAMASWAMY, S., BEROUKHIM, R., MILNER, D. A., GRANTER, S. R., DU, J., LEE, C., WAGNER, S. N., LI, C., GOLUB, T. R., RIMM, D. L., MEYERSON, M. L., FISHER, D. E. & SELLERS, W. R. 2005. Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature,* 436**,** 117-22.

GAST, A., SCHERER, D., CHEN, B., BLOETHNER, S., MELCHERT, S., SUCKER, A., HEMMINKI, K., SCHADENDORF, D. & KUMAR, R. 2010. Somatic alterations in the melanoma genome: a high-resolution array-based comparative genomic hybridization study. *Genes Chromosomes Cancer,* 49**,** 733-45.

GENTLEMAN, R. C., CAREY, V. J., BATES, D. M., BOLSTAD, B., DETTLING, M., DUDOIT, S., ELLIS, B., GAUTIER, L., GE, Y., GENTRY, J., HORNIK, K., HOTHORN, T., HUBER, W., IACUS, S., IRIZARRY, R., LEISCH, F., LI, C., MAECHLER, M., ROSSINI, A. J., SAWITZKI, G., SMITH, C., SMYTH, G., TIERNEY, L., YANG, J. Y. & ZHANG, J. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol,* 5**,** R80.

GERAMI, P., JEWELL, S. S., POURYAZDANPARAST, P., WAYNE, J. D., HAGHIGHAT, Z., BUSAM, K. J., RADEMAKER, A. & MORRISON, L. 2011. Copy number gains in 11q13 and 8q24 [corrected] are highly linked to prognosis in cutaneous malignant melanoma. *J Mol Diagn,* 13**,** 352-8.

GIORGI, F. M., BOLGER, A. M., LOHSE, M. & USADEL, B. 2010. Algorithm-driven artifacts in median Polish summarization of microarray data. *BMC Bioinformatics,* 11**,** 553.

GOGAS, H., EGGERMONT, A. M., HAUSCHILD, A., HERSEY, P., MOHR, P., SCHADENDORF, D., SPATZ, A. & DUMMER, R. 2009. Biomarkers in melanoma. *Ann Oncol,* 20 Suppl 6**,** vi8-13.

GOULD ROTHBERG, B. E., BRACKEN, M. B. & RIMM, D. L. 2009. Tissue biomarkers for prognosis in cutaneous melanoma: a systematic review and meta-analysis. *J Natl Cancer Inst,* 101**,** 452-74.

GRABACKA, M., PLONKA, P. M., URBANSKA, K. & REISS, K. 2006. Peroxisome proliferator-activated receptor alpha activation decreases metastatic potential of melanoma cells in vitro via down-regulation of Akt. *Clin Cancer Res,* 12**,** 3028-36.

GRAY-SCHOPFER, V., WELLBROCK, C. & MARAIS, R. 2007. Melanoma biology and new targeted therapy. *Nature,* 445**,** 851-7.

GRAY-SCHOPFER, V. C., CHEONG, S. C., CHONG, H., CHOW, J., MOSS, T., ABDEL-MALEK, Z. A., MARAIS, R., WYNFORD-THOMAS, D. & BENNETT, D. C. 2006. Cellular senescence in naevi and immortalisation in melanoma: a role for p16? *Br J Cancer,* 95**,** 496-505.

GRESHOCK, J., NATHANSON, K., MEDINA, A., WARD, M. R., HERLYN, M., WEBER, B. L. & ZAKS, T. Z. 2009. Distinct patterns of DNA copy number alterations associate with BRAF mutations in melanomas and melanoma-derived cell lines. *Genes Chromosomes Cancer,* 48**,** 419-28.

GRUBER, B. M. & ANUSZEWSKA, E. L. 2002. Influence of vitamin D3 metabolites on cell proliferation and cytotoxicity of adriamycin in human normal and neoplastic cells. *Toxicol In Vitro,* 16**,** 663-7.

GUO, H., CARLSON, J. A. & SLOMINSKI, A. 2012. Role of TRPM in melanocytes and melanoma. *Exp Dermatol,* 21**,** 650-4.

GUSNANTO, A., WOOD, H. M., PAWITAN, Y., RABBITTS, P. & BERRI, S. 2012. Correcting for cancer genome size and tumour cell content enables better estimation of copy number alterations from next-generation sequence data. *Bioinformatics,* 28**,** 40-7.

HANAHAN, D. & WEINBERG, R. A. 2011. Hallmarks of cancer: the next generation. *Cell,* 144**,** 646-74.

HANEKE, E. & BASTIAN, B. C. 2006. Melanocytic tumours. Superficial spreading melanoma. *World Health Organisation Classification of Tumours. Pathology and Genetics of Skin Tumours.* Lyon, France: IARC press.

HAQQ, C., NOSRATI, M., SUDILOVSKY, D., CROTHERS, J., KHODABAKHSH, D., PULLIAM, B. L., FEDERMAN, S., MILLER, J. R., 3RD, ALLEN, R. E., SINGER, M. I., LEONG, S. P., LJUNG, B. M., SAGEBIEL, R. W. & KASHANI-SABET, M. 2005. The gene expression signatures of melanoma progression. *Proc Natl Acad Sci U S A,* 102**,** 6092-7.

HARANT, H., ANDREW, P. J., REDDY, G. S., FOGLAR, E. & LINDLEY, I. J. 1997. 1alpha,25-dihydroxyvitamin D3 and a variety of its natural metabolites transcriptionally repress nuclear-factor-kappaB-mediated interleukin-8 gene expression. *Eur J Biochem,* 250**,** 63-71.

HARITOGLOU, I., WOLF, A., MAIER, T., HARITOGLOU, C., HEIN, R. & SCHALLER, U. C. 2009. Osteopontin and 'melanoma inhibitory activity': comparison of two serological tumor markers in metastatic uveal melanoma patients. *Ophthalmologica,* 223**,** 239-43.

HAUSSLER, M. R., JURUTKA, P. W., MIZWICKI, M. & NORMAN, A. W. 2011. Vitamin D receptor (VDR)-mediated actions of 1alpha,25(OH)(2)vitamin D(3): genomic and non-genomic mechanisms. *Best Pract Res Clin Endocrinol Metab,* 25**,** 543-59.

HAWKINS, R. D., HON, G. C. & REN, B. 2010. Next-generation genomics: an integrative approach. *Nat Rev Genet,* 11**,** 476-86.

HEENAN, P., SPATZ, A., CERIO, R. & BASTIAN, B. C. 2006. Melanocytic tumours. Lentigo maligna. *World Health Organisation Classification of Tumours. Pathology and Genetics of Skin Tumours.* Lyon, France: IARC press.

HEIKKINEN, S., VAISANEN, S., PEHKONEN, P., SEUTER, S., BENES, V. & CARLBERG, C. 2011. Nuclear hormone 1alpha,25-dihydroxyvitamin D3 elicits a genome-wide shift in the locations of VDR chromatin occupancy. *Nucleic Acids Res,* 39**,** 9181-93.

HILL, V. K., GARTNER, J. J., SAMUELS, Y. & GOLDSTEIN, A. M. 2013. The genetics of melanoma: recent advances. *Annu Rev Genomics Hum Genet,* 14**,** 257-79.

HIRSCH, D., KEMMERLING, R., DAVIS, S., CAMPS, J., MELTZER, P. S., RIED, T. & GAISER, T. 2013. Chromothripsis and focal copy number alterations determine poor outcome in malignant melanoma. *Cancer Res,* 73**,** 1454-60.

HODIS, E., WATSON, I. R., KRYUKOV, G. V., AROLD, S. T., IMIELINSKI, M., THEURILLAT, J. P., NICKERSON, E., AUCLAIR, D., LI, L., PLACE, C., DICARA, D., RAMOS, A. H., LAWRENCE, M. S., CIBULSKIS, K., SIVACHENKO, A., VOET, D., SAKSENA, G., STRANSKY, N., ONOFRIO, R. C., WINCKLER, W., ARDLIE, K., WAGLE, N., WARGO, J., CHONG, K., MORTON, D. L., STEMKE-HALE, K., CHEN, G., NOBLE, M., MEYERSON,

M., LADBURY, J. E., DAVIES, M. A., GERSHENWALD, J. E., WAGNER, S. N., HOON, D. S., SCHADENDORF, D., LANDER, E. S., GABRIEL, S. B., GETZ, G., GARRAWAY, L. A. & CHIN, L. 2012. A landscape of driver mutations in melanoma. *Cell,* 150**,** 251-63.

HOFMANN, M. A., GUSSMANN, F., FRITSCHE, A., BIESOLD, S., SCHICKE, B., KUCHLER, I., VOIT, C. & TREFZER, U. 2009. Diagnostic value of melanoma inhibitory activity serum marker in the follow-up of patients with stage I or II cutaneous melanoma. *Melanoma Res,* 19**,** 17-23.

HOLICK, M. F. 2003. Vitamin D: A millenium perspective. *J Cell Biochem,* 88**,** 296-307.

HOLICK, M. F. 2004. Sunlight and vitamin D for bone health and prevention of autoimmune diseases, cancers, and cardiovascular disease. *Am J Clin Nutr,* 80**,** 1678S-88S.

HOMSI, J., KASHANI-SABET, M., MESSINA, J. L. & DAUD, A. 2005. Cutaneous melanoma: prognostic factors. *Cancer Control,* 12**,** 223-9.

HORN, S., FIGL, A., RACHAKONDA, P. S., FISCHER, C., SUCKER, A., GAST, A., KADEL, S., MOLL, I., NAGORE, E., HEMMINKI, K., SCHADENDORF, D. & KUMAR, R. 2013. TERT promoter mutations in familial and sporadic melanoma. *Science,* 339**,** 959-61.

HORST, B., GRUVBERGER-SAAL, S. K., HOPKINS, B. D., BORDONE, L., YANG, Y., CHERNOFF, K. A., UZOMA, I., SCHWIPPER, V., LIEBAU, J., NOWAK, N. J., BRUNNER, G., OWENS, D., RIMM, D. L., PARSONS, R. & CELEBI, J. T. 2009. Gab2-mediated signaling promotes melanoma metastasis. *Am J Pathol,* 174**,** 1524-33.

HUBBELL, E., LIU, W. M. & MEI, R. 2002. Robust estimators for expression analysis. *Bioinformatics,* 18**,** 1585-92.

ILES, M. M., LAW, M. H., STACEY, S. N., HAN, J., FANG, S., PFEIFFER, R., HARLAND, M., MACGREGOR, S., TAYLOR, J. C., ABEN, K. K., AKSLEN, L. A., AVRIL, M. F., AZIZI, E., BAKKER, B., BENEDIKTSDOTTIR, K. R., BERGMAN, W., SCARRA, G. B., BROWN, K. M., CALISTA, D., CHAUDRU, V., FARGNOLI, M. C., CUST, A. E., DEMENAIS, F., DE WAAL, A. C., DEBNIAK, T., ELDER, D. E., FRIEDMAN, E., GALAN, P., GHIORZO, P., GILLANDERS, E. M., GOLDSTEIN, A. M., GRUIS, N. A., HANSSON, J., HELSING, P., HOCEVAR, M., HOIOM, V., HOPPER, J. L., INGVAR, C., JANSSEN, M., JENKINS, M. A., KANETSKY, P. A., KIEMENEY, L. A., LANG, J., LATHROP, G. M., LEACHMAN, S., LEE, J. E., LUBINSKI, J., MACKIE, R. M., MANN, G. J., MARTIN, N. G., MAYORDOMO, J. I., MOLVEN, A., MULDER, S., NAGORE, E., NOVAKOVIC, S., OKAMOTO, I., OLAFSSON, J. H., OLSSON, H., PEHAMBERGER, H., PERIS, K., GRASA, M. P., PLANELLES, D., PUIG, S., PUIG-BUTILLE, J. A., RANDERSON-MOOR, J., REQUENA, C., RIVOLTINI, L., RODOLFO, M., SANTINAMI, M., SIGURGEIRSSON, B., SNOWDEN, H., SONG, F., SULEM, P., THORISDOTTIR, K., TUOMINEN, R., VAN BELLE, P., VAN DER STOEP, N., VAN ROSSUM, M. M., WEI, Q., WENDT, J., ZELENIKA, D., ZHANG, M., LANDI, M. T., THORLEIFSSON, G., BISHOP, D. T., AMOS, C. I., HAYWARD, N. K., STEFANSSON, K., BISHOP, J. A., BARRETT, J. H., GENO, M. E. L. C., Q, M. & INVESTIGATORS, A. 2013. A variant in FTO shows association with melanoma risk not due to BMI. *Nat Genet,* 45**,** 428-32, 432e1.

ILLUMINA. 2012. *Whole-Genome Gene Expression DASL HT Assay Guide* [Online]. San Diego, California, U.S.A: Illumina. Available: http://supportres.illumina.com/documents/myillumina/97db80bc-667e-4e34-a44d-9f8761fff229/wgdasl_ht_assay_guide_15018210_d.pdf [Accessed 7th September 2013].

IRIZARRY, R. A., BOLSTAD, B. M., COLLIN, F., COPE, L. M., HOBBS, B. & SPEED, T. P. 2003a. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res,* 31**,** e15.

IRIZARRY, R. A., HOBBS, B., COLLIN, F., BEAZER-BARCLAY, Y. D., ANTONELLIS, K. J., SCHERF, U. & SPEED, T. P. 2003b. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics,* 4**,** 249-64.

ISA, S., KAWAGUCHI, T., TERAMUKAI, S., MINATO, K., OHSAKI, Y., SHIBATA, K., YONEI, T., HAYASHIBARA, K., FUKUSHIMA, M., KAWAHARA, M., FURUSE, K. & MACK, P. C. 2009. Serum osteopontin levels are highly prognostic for survival in advanced non-small cell lung cancer: results from JMTO LC 0004. *J Thorac Oncol,* 4**,** 1104-10.

ISHIBASHI, M., ARAI, M., TANAKA, S., ONDA, K. & HIRANO, T. 2012. Antiproliferative and apoptosis-inducing effects of lipophilic vitamins on human melanoma A375 cells in vitro. *Biol Pharm Bull,* 35**,** 10-7.

JAEGER, J., KOCZAN, D., THIESEN, H. J., IBRAHIM, S. M., GROSS, G., SPANG, R. & KUNZ, M. 2007. Gene expression signatures for tumor progression, tumor subtype, and tumor thickness in laser-microdissected melanoma tissues. *Clin Cancer Res,* 13**,** 806-15.

JANJETOVIC, Z., BROZYNA, A. A., TUCKEY, R. C., KIM, T. K., NGUYEN, M. N., JOZWICKI, W., PFEFFER, S. R., PFEFFER, L. M. & SLOMINSKI, A. T. 2011. High basal NF-kappaB activity in nonpigmented melanoma cells is associated with an enhanced sensitivity to vitamin D3 derivatives. *Br J Cancer,* 105**,** 1874-84.

JEANMOUGIN, M., DE REYNIES, A., MARISA, L., PACCARD, C., NUEL, G. & GUEDJ, M. 2010. Should we abandon the t-test in the analysis of gene expression microarray data: a comparison of variance modeling strategies. *PLoS One,* 5**,** e12336.

JENAB, M., BUENO-DE-MESQUITA, H. B., FERRARI, P., VAN DUIJNHOVEN, F. J., NORAT, T., PISCHON, T., JANSEN, E. H., SLIMANI, N., BYRNES, G., RINALDI, S., TJONNELAND, A., OLSEN, A., OVERVAD, K., BOUTRON-RUAULT, M. C., CLAVEL-CHAPELON, F., MOROIS, S., KAAKS, R., LINSEISEN, J., BOEING, H., BERGMANN, M. M., TRICHOPOULOU, A., MISIRLI, G., TRICHOPOULOS, D., BERRINO, F., VINEIS, P., PANICO, S., PALLI, D., TUMINO, R., ROS, M. M., VAN GILS, C. H., PEETERS, P. H., BRUSTAD, M., LUND, E., TORMO, M. J., ARDANAZ, E., RODRIGUEZ, L., SANCHEZ, M. J., DORRONSORO, M., GONZALEZ, C. A., HALLMANS, G., PALMQVIST, R., RODDAM, A., KEY, T. J., KHAW, K. T., AUTIER, P., HAINAUT, P. & RIBOLI, E. 2010. Association between pre-diagnostic circulating vitamin D concentration and risk of colorectal cancer in European populations:a nested case-control study. *BMJ,* 340**,** b5500.

JEWELL, R., CONWAY, C., MITRA, A., RANDERSON-MOOR, J., LOBO, S., NSENGIMANA, J., HARLAND, M., MARPLES, M., EDWARD, S., COOK, M., POWELL, B., BOON, A., DE KORT, F., PARKER, K. A., CREE, I. A., BARRETT, J. H., KNOWLES, M. A., BISHOP, D. T. & NEWTON-BISHOP, J. 2010. Patterns of expression of DNA repair genes and relapse from melanoma. *Clin Cancer Res,* 16**,** 5211-21.

JONSSON, G., BUSCH, C., KNAPPSKOG, S., GEISLER, J., MILETIC, H., RINGNER, M., LILLEHAUG, J. R., BORG, A. & LONNING, P. E. 2010. Gene expression profiling-based identification of molecular subtypes in stage IV melanomas with different clinical outcome. *Clin Cancer Res,* 16**,** 3356-67.

JONSSON, G., DAHL, C., STAAF, J., SANDBERG, T., BENDAHL, P. O., RINGNER, M., GULDBERG, P. & BORG, A. 2007. Genomic profiling of malignant melanoma using tiling-resolution arrayCGH. *Oncogene,* 26**,** 4738-48.

KABBARAH, O. & CHIN, L. 2005. Revealing the genomic heterogeneity of melanoma. *Cancer Cell,* 8**,** 439-41.

KADKOL, S. S., LIN, A. Y., BARAK, V., KALICKMAN, I., LEACH, L., VALYI-NAGY, K., MAJUMDAR, D., SETTY, S., MANIOTIS, A. J., FOLBERG, R. & PE'ER, J. 2006. Osteopontin expression and serum levels in metastatic uveal melanoma: a pilot study. *Invest Ophthalmol Vis Sci,* 47**,** 802-6.

KAMB, A., SHATTUCK-EIDENS, D., EELES, R., LIU, Q., GRUIS, N. A., DING, W., HUSSEY, C., TRAN, T., MIKI, Y., WEAVER-FELDHAUS, J. & ET AL. 1994. Analysis of the p16 gene (CDKN2) as a candidate for the chromosome 9p melanoma susceptibility locus. *Nat Genet,* 8**,** 23-6.

KANEHISA, M., GOTO, S., SATO, Y., FURUMICHI, M. & TANABE, M. 2012. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res,* 40**,** D109-14.

KASHANI-SABET, M., VENNA, S., NOSRATI, M., RANGEL, J., SUCKER, A., EGBERTS, F., BAEHNER, F. L., SIMKO, J., LEONG, S. P., HAQQ, C., HAUSCHILD, A., SCHADENDORF, D., MILLER, J. R., 3RD & SAGEBIEL, R. W. 2009. A multimarker prognostic assay for primary cutaneous melanoma. *Clin Cancer Res,* 15**,** 6987-92.

KAUFFMANN, A., ROSSELLI, F., LAZAR, V., WINNEPENNINCKX, V., MANSUET-LUPO, A., DESSEN, P., VAN DEN OORD, J. J., SPATZ, A. & SARASIN, A. 2008. High expression of DNA repair pathways is associated with metastasis in melanoma patients. *Oncogene,* 27**,** 565-73.

KETCHART, W., SMITH, K. M., KRUPKA, T., WITTMANN, B. M., HU, Y., RAYMAN, P. A., DOUGHMAN, Y. Q., ALBERT, J. M., BAI, X., FINKE, J. H., XU, Y., EXNER, A. A. & MONTANO, M. M. 2013. Inhibition of metastasis by HEXIM1 through effects on cell invasion and angiogenesis. *Oncogene,* 32**,** 3829-39.

KIM, R. D., CURTIN, J. A. & BASTIAN, B. C. 2008. Lack of somatic alterations of MC1R in primary melanoma. *Pigment Cell Melanoma Res,* 21**,** 579-82.

KIRCHER, M., HEYN, P. & KELSO, J. 2011. Addressing challenges in the production and analysis of illumina sequencing data. *BMC Genomics,* 12**,** 382.

KLUGER, H. M., HOYT, K., BACCHIOCCHI, A., MAYER, T., KIRSCH, J., KLUGER, Y., SZNOL, M., ARIYAN, S., MOLINARO, A. & HALABAN, R. 2011. Plasma markers for identifying patients with metastatic melanoma. *Clin Cancer Res,* 17**,** 2417-25.

KORNER, H., EPANCHINTSEV, A., BERKING, C., SCHULER-THURNER, B., SPEICHER, M. R., MENSSEN, A. & HERMEKING, H. 2007. Digital karyotyping reveals frequent inactivation of the dystrophin/DMD gene in malignant melanoma. *Cell Cycle,* 6**,** 189-98.

KOVALENKO, P. L., ZHANG, Z., CUI, M., CLINTON, S. K. & FLEET, J. C. 2010. 1,25 dihydroxyvitamin D-mediated orchestration of anticancer, transcript-level effects in the immortalized, non-transformed prostate epithelial cell line, RWPE1. *BMC Genomics,* 11**,** 26.

KRAHN, G., KASKEL, P., SANDER, S., WAIZENHOFER, P. J., WORTMANN, S., LEITER, U. & PETER, R. U. 2001. S100 beta is a more reliable tumor marker in peripheral blood for patients with newly occurred melanoma metastases compared with MIA, albumin and lactate-dehydrogenase. *Anticancer Res,* 21**,** 1311-6.

KRAUTHAMMER, M., KONG, Y., HA, B. H., EVANS, P., BACCHIOCCHI, A., MCCUSKER, J. P., CHENG, E., DAVIS, M. J., GOH, G., CHOI, M., ARIYAN, S., NARAYAN, D., DUTTON-REGESTER, K., CAPATANA, A., HOLMAN, E. C., BOSENBERG, M., SZNOL, M., KLUGER, H. M., BRASH, D. E., STERN, D. F., MATERIN, M. A., LO, R. S., MANE, S., MA, S., KIDD, K. K., HAYWARD, N. K., LIFTON, R. P., SCHLESSINGER, J., BOGGON, T.

J. & HALABAN, R. 2012. Exome sequencing identifies recurrent somatic RAC1 mutations in melanoma. *Nat Genet,* 44**,** 1006-14.

KRIEBITZSCH, C., VERLINDEN, L., EELEN, G., TAN, B. K., VAN CAMP, M., BOUILLON, R. & VERSTUYF, A. 2009. The impact of 1,25(OH)2D3 and its structural analogs on gene expression in cancer cells--a microarray approach. *Anticancer Res,* 29**,** 3471-83.

KULASINGAM, V. & DIAMANDIS, E. P. 2008. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. *Nat Clin Pract Oncol,* 5**,** 588-99.

KUMAR, S. M., ZHANG, G., BASTIAN, B. C., ARCASOY, M. O., KARANDE, P., PUSHPARAJAN, A., ACS, G. & XU, X. 2012. Erythropoietin receptor contributes to melanoma cell survival in vivo. *Oncogene,* 31**,** 1649-60.

LAW, M. H., MACGREGOR, S. & HAYWARD, N. K. 2012. Melanoma genetics: recent findings take us beyond well-traveled pathways. *J Invest Dermatol,* 132**,** 1763-74.

LEE, J. D., UNGER, E. R., GITTENGER, C., LEE, D. R., HEBERT, R. & MAIZE, J. C. 2001. Interphase cytogenetic analysis of 1q12 satellite III DNA in melanocytic lesions: increased aneuploidy with malignant histology. *Am J Dermatopathol,* 23**,** 176-80.

LEE, J. H., PARK, S., CHEON, S., LEE, J. H., KIM, S., HUR, D. Y., KIM, T. S., YOON, S. R., YANG, Y., BANG, S. I., PARK, H., LEE, H. T. & CHO, D. 2011. 1,25-Dihydroxyvitamin D(3) enhances NK susceptibility of human melanoma cells via Hsp60-mediated FAS expression. *Eur J Immunol,* 41**,** 2937-46.

LEONG, S. P., CADY, B., JABLONS, D. M., GARCIA-AGUILAR, J., REINTGEN, D., JAKUB, J., PENDAS, S., DUHAIME, L., CASSELL, R., GARDNER, M., GIULIANO, R., ARCHIE, V., CALVIN, D., MENSHA, L., SHIVERS, S., COX, C., WERNER, J. A., KITAGAWA, Y. & KITAJIMA, M. 2006. Clinical patterns of metastasis. *Cancer Metastasis Rev,* 25**,** 221-32.

LEVY, C., KHALED, M. & FISHER, D. E. 2006. MITF: master regulator of melanocyte development and melanoma oncogene. *Trends Mol Med,* 12**,** 406-14.

LEW, Q. J., CHIA, Y. L., CHU, K. L., LAM, Y. T., GURUMURTHY, M., XU, S., LAM, K. P., CHEONG, N. & CHAO, S. H. 2012. Identification of HEXIM1 as a positive regulator of p53. *J Biol Chem,* 287**,** 36443-54.

LI, H. & DURBIN, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics,* 25**,** 1754-60.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & GENOME PROJECT DATA PROCESSING, S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.

LI, W., CHEN, J., JANJETOVIC, Z., KIM, T. K., SWEATMAN, T., LU, Y., ZJAWIONY, J., TUCKEY, R. C., MILLER, D. & SLOMINSKI, A. 2010. Chemical synthesis of 20S-hydroxyvitamin D3, which shows antiproliferative activity. *Steroids,* 75**,** 926-35.

LI, Y., LI, L., WANG, J. T., KAN, X. & LU, J. G. 2011. Elevated content of osteopontin in plasma and tumor tissues of patients with laryngeal and hypopharyngeal carcinoma associated with metastasis and prognosis. *Med Oncol*.

LIEBEN, L. & CARMELIET, G. 2013. Vitamin D signaling in osteocytes: effects on bone and mineral homeostasis. *Bone,* 54**,** 237-43.

LIMON, J., DAL CIN, P., SAIT, S. N., KARAKOUSIS, C. & SANDBERG, A. A. 1988. Chromosome changes in metastatic human melanoma. *Cancer Genet Cytogenet,* 30**,** 201-11.

LIN, W. M., BAKER, A. C., BEROUKHIM, R., WINCKLER, W., FENG, W., MARMION, J. M., LAINE, E., GREULICH, H., TSENG, H., GATES, C.,

HODI, F. S., DRANOFF, G., SELLERS, W. R., THOMAS, R. K., MEYERSON, M., GOLUB, T. R., DUMMER, R., HERLYN, M., GETZ, G. & GARRAWAY, L. A. 2008. Modeling genomic diversity and tumor dependency in malignant melanoma. *Cancer Res,* 68**,** 664-73.

LIPSHUTZ, R. J., FODOR, S. P., GINGERAS, T. R. & LOCKHART, D. J. 1999. High density synthetic oligonucleotide arrays. *Nat Genet,* 21**,** 20-4.

LITTLE, S. E., VUONONVIRTA, R., REIS-FILHO, J. S., NATRAJAN, R., IRAVANI, M., FENWICK, K., MACKAY, A., ASHWORTH, A., PRITCHARD-JONES, K. & JONES, C. 2006. Array CGH using whole genome amplification of fresh-frozen and formalin-fixed, paraffin-embedded tumor DNA. *Genomics,* 87**,** 298-306.

LIU, L., LI, Y., LI, S., HU, N., HE, Y., PONG, R., LIN, D., LU, L. & LAW, M. 2012. Comparison of next-generation sequencing systems. *J Biomed Biotechnol,* 2012**,** 251364.

LOCASALE, J. W., GRASSIAN, A. R., MELMAN, T., LYSSIOTIS, C. A., MATTAINI, K. R., BASS, A. J., HEFFRON, G., METALLO, C. M., MURANEN, T., SHARFI, H., SASAKI, A. T., ANASTASIOU, D., MULLARKY, E., VOKES, N. I., SASAKI, M., BEROUKHIM, R., STEPHANOPOULOS, G., LIGON, A. H., MEYERSON, M., RICHARDSON, A. L., CHIN, L., WAGNER, G., ASARA, J. M., BRUGGE, J. S., CANTLEY, L. C. & VANDER HEIDEN, M. G. 2011. Phosphoglycerate dehydrogenase diverts glycolytic flux and contributes to oncogenesis. *Nat Genet,* 43**,** 869-74.

LOCKHART, D. J., DONG, H., BYRNE, M. C., FOLLETTIE, M. T., GALLO, M. V., CHEE, M. S., MITTMANN, M., WANG, C., KOBAYASHI, M., HORTON, H. & BROWN, E. L. 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol,* 14**,** 1675-80.

LOPEZ-OTIN, C. & MATRISIAN, L. M. 2007. Emerging roles of proteases in tumour suppression. *Nat Rev Cancer,* 7**,** 800-8.

LORIGAN, P., EISEN, T. & HAUSCHILD, A. 2008. Systemic therapy for metastatic malignant melanoma--from deeply disappointing to bright future? *Exp Dermatol,* 17**,** 383-94.

MACK, P. C., REDMAN, M. W., CHANSKY, K., WILLIAMSON, S. K., FARNETH, N. C., LARA, P. N., JR., FRANKLIN, W. A., LE, Q. T., CROWLEY, J. J. & GANDARA, D. R. 2008. Lower osteopontin plasma levels are associated with superior outcomes in advanced non-small-cell lung cancer patients receiving platinum-based chemotherapy: SWOG Study S0003. *J Clin Oncol,* 26**,** 4771-6.

MAIER, T., LAUBENDER, R. P., STURM, R. A., KLINGENSTEIN, A., KORTING, H. C., RUZICKA, T. & BERKING, C. 2011. Osteopontin expression in plasma of melanoma patients and in melanocytic tumours. *J Eur Acad Dermatol Venereol*.

MAIER, T., LAUBENDER, R. P., STURM, R. A., KLINGENSTEIN, A., KORTING, H. C., RUZICKA, T. & BERKING, C. 2012. Osteopontin expression in plasma of melanoma patients and in melanocytic tumours. *J Eur Acad Dermatol Venereol,* 26**,** 1084-91.

MALDONADO, J. L., FRIDLYAND, J., PATEL, H., JAIN, A. N., BUSAM, K., KAGESHITA, T., ONO, T., ALBERTSON, D. G., PINKEL, D. & BASTIAN, B. C. 2003. Determinants of BRAF mutations in primary melanomas. *J Natl Cancer Inst,* 95**,** 1878-90.

MANDRUZZATO, S., CALLEGARO, A., TURCATEL, G., FRANCESCATO, S., MONTESCO, M. C., CHIARION-SILENI, V., MOCELLIN, S., ROSSI, C. R., BICCIATO, S., WANG, E., MARINCOLA, F. M. & ZANOVELLO, P. 2006. A gene expression signature associated with survival in metastatic melanoma. *J Transl Med,* 4**,** 50.

MANI, I., SHARMA, V., TAMBOLI, I. & RAMAN, G. 2001. Interaction of melanin with proteins--the importance of an acidic intramelanosomal pH. *Pigment Cell Res,* 14**,** 170-9.

MARDIS, E. R. 2008. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet,* 9**,** 387-402.

MARSDEN, J. R., NEWTON-BISHOP, J. A., BURROWS, L., COOK, M., CORRIE, P. G., COX, N. H., GORE, M. E., LORIGAN, P., MACKIE, R., NATHAN, P., PEACH, H., POWELL, B., WALKER, C. & BRITISH ASSOCIATION OF DERMATOLOGISTS CLINICAL STANDARDS, U. 2010. Revised U.K. guidelines for the management of cutaneous melanoma 2010. *Br J Dermatol,* 163**,** 238-56.

MARZINKE, M. A. & CLAGETT-DAME, M. 2012. The all-trans retinoic acid (atRA)-regulated gene Calmin (Clmn) regulates cell cycle exit and neurite outgrowth in murine neuroblastoma (Neuro2a) cells. *Exp Cell Res,* 318**,** 85-93.

MASER, R. S., CHOUDHURY, B., CAMPBELL, P. J., FENG, B., WONG, K. K., PROTOPOPOV, A., O'NEIL, J., GUTIERREZ, A., IVANOVA, E., PERNA, I., LIN, E., MANI, V., JIANG, S., MCNAMARA, K., ZAGHLUL, S., EDKINS, S., STEVENS, C., BRENNAN, C., MARTIN, E. S., WIEDEMEYER, R., KABBARAH, O., NOGUEIRA, C., HISTEN, G., ASTER, J., MANSOUR, M., DUKE, V., FORONI, L., FIELDING, A. K., GOLDSTONE, A. H., ROWE, J. M., WANG, Y. A., LOOK, A. T., STRATTON, M. R., CHIN, L., FUTREAL, P. A. & DEPINHO, R. A. 2007. Chromosomally unstable mouse tumours have genomic alterations similar to diverse human cancers. *Nature,* 447**,** 966-71.

MCSHANE, L. M., ALTMAN, D. G., SAUERBREI, W., TAUBE, S. E., GION, M., CLARK, G. M. & STATISTICS SUBCOMMITTEE OF THE, N. C. I. E. W. G. O. C. D. 2005. REporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer,* 93**,** 387-91.

MENON, R., DENG, M., BOEHM, D., BRAUN, M., FEND, F., BOEHM, D., BISKUP, S. & PERNER, S. 2012. Exome Enrichment and SOLiD Sequencing of Formalin Fixed Paraffin Embedded (FFPE) Prostate Cancer Tissue. *Int J Mol Sci,* 13**,** 8933-42.

MERTENS, F., JOHANSSON, B., HOGLUND, M. & MITELMAN, F. 1997. Chromosomal imbalance maps of malignant solid tumors: a cytogenetic survey of 3185 neoplasms. *Cancer Res,* 57**,** 2765-80.

METZKER, M. L. 2010. Sequencing technologies - the next generation. *Nat Rev Genet,* 11**,** 31-46.

MEYERSON, M., GABRIEL, S. & GETZ, G. 2010. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet,* 11**,** 685-96.

MEYLE, K. D. & GULDBERG, P. 2009. Genetic risk factors for melanoma. *Hum Genet,* 126**,** 499-510.

MEZAWA, H., SUGIURA, T., WATANABE, M., NORIZOE, C., TAKAHASHI, D., SHIMOJIMA, A., TAMEZ, S., TSUTSUMI, Y., YANAGA, K. & URASHIMA, M. 2010. Serum vitamin D levels and survival of patients with colorectal cancer: post-hoc analysis of a prospective cohort study. *BMC Cancer,* 10**,** 347.

MICHALOGLOU, C., VREDEVELD, L. C., SOENGAS, M. S., DENOYELLE, C., KUILMAN, T., VAN DER HORST, C. M., MAJOOR, D. M., SHAY, J. W., MOOI, W. J. & PEEPER, D. S. 2005. BRAFE600-associated senescence-like cell cycle arrest of human naevi. *Nature,* 436**,** 720-4.

MILANI, C., KATAYAMA, M. L., DE LYRA, E. C., WELSH, J., CAMPOS, L. T., BRENTANI, M. M., MACIEL MDO, S., ROELA, R. A., DEL VALLE, P. R., GOES, J. C., NONOGAKI, S., TAMURA, R. E. & FOLGUEIRA, M. A. 2013. Transcriptional effects of 1,25 dihydroxyvitamin D(3) physiological and

supra-physiological concentrations in breast cancer organotypic culture. *BMC Cancer,* 13**,** 119.
MILLER, A. J. & MIHM, M. C., JR. 2006. Melanoma. *N Engl J Med,* 355**,** 51-65.
MINOCHE, A. E., DOHM, J. C. & HIMMELBAUER, H. 2011. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biol,* 12**,** R112.
MOCELLIN, S., PASQUALI, S., ROSSI, C. R. & NITTI, D. 2010. Interferon alpha adjuvant therapy in patients with high-risk melanoma: a systematic review and meta-analysis. *J Natl Cancer Inst,* 102**,** 493-501.
MOCELLIN, S., ZAVAGNO, G. & NITTI, D. 2008. The prognostic value of serum S100B in patients with cutaneous melanoma: a meta-analysis. *Int J Cancer,* 123**,** 2370-6.
MOHRI, T., NAKAJIMA, M., TAKAGI, S., KOMAGATA, S. & YOKOI, T. 2009. MicroRNA regulates human vitamin D receptor. *Int J Cancer,* 125**,** 1328-33.
MOLIFE, R. & HANCOCK, B. W. 2002. Adjuvant therapy of malignant melanoma. *Crit Rev Oncol Hematol,* 44**,** 81-102.
MUELLER, O., HAHNENBERGER, K., DITTMANN, M., YEE, H., DUBROW, R., NAGLE, R. & ILSLEY, D. 2000. A microfluidic system for high-speed reproducible DNA sizing and quantitation. *Electrophoresis,* 21**,** 128-34.
MULLIN, G. E. & DOBS, A. 2007. Vitamin d and its role in cancer and immunity: a prescription for sunlight. *Nutr Clin Pract,* 22**,** 305-22.
NAEF, F., HACKER, C. R., PATIL, N. & MAGNASCO, M. 2002. Empirical characterization of the expression ratio noise structure in high-density oligonucleotide arrays. *Genome Biol,* 3**,** RESEARCH0018.
NAMIKI, T., YANAGAWA, S., IZUMO, T., ISHIKAWA, M., TACHIBANA, M., KAWAKAMI, Y., YOKOZEKI, H., NISHIOKA, K. & KANEKO, Y. 2005. Genomic alterations in primary cutaneous melanomas detected by metaphase comparative genomic hybridization with laser capture or manual microdissection: 6p gains may predict poor outcome. *Cancer Genet Cytogenet,* 157**,** 1-11.
NEWTON-BISHOP, J. & RANDERSON-MOOR, J. 2013. Vitamin D and melanoma. *Handbook of vitamin D in human health.* The Netherlands: Wageningen Academic Publishers
NEWTON-BISHOP, J. A., BESWICK, S., RANDERSON-MOOR, J., CHANG, Y. M., AFFLECK, P., ELLIOTT, F., CHAN, M., LEAKE, S., KARPAVICIUS, B., HAYNES, S., KUKALIZCH, K., WHITAKER, L., JACKSON, S., GERRY, E., NOLAN, C., BERTRAM, C., MARSDEN, J., ELDER, D. E., BARRETT, J. H. & BISHOP, D. T. 2009. Serum 25-hydroxyvitamin D3 levels are associated with breslow thickness at presentation and survival from melanoma. *J Clin Oncol,* 27**,** 5439-44.
NEWTON-BISHOP, J. A., CHANG, Y. M., ELLIOTT, F., CHAN, M., LEAKE, S., KARPAVICIUS, B., HAYNES, S., FITZGIBBON, E., KUKALIZCH, K., RANDERSON-MOOR, J., ELDER, D. E., BISHOP, D. T. & BARRETT, J. H. 2011. 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,* 47**,** 732-41.
NEWTON-BISHOP, J. A., CHANG, Y. M., ILES, M. M., TAYLOR, J. C., BAKKER, B., CHAN, M., LEAKE, S., KARPAVICIUS, B., HAYNES, S., FITZGIBBON, E., ELLIOTT, F., KANETSKY, P. A., HARLAND, M., BARRETT, J. H. & BISHOP, D. T. 2010. Melanocytic nevi, nevus genes, and melanoma risk in a large case-control study in the United Kingdom. *Cancer Epidemiol Biomarkers Prev,* 19**,** 2043-54.
NEWTON BISHOP, J. A. & BISHOP, D. T. 2005. The genetics of susceptibility to cutaneous melanoma. *Drugs Today (Barc),* 41**,** 193-203.
NORTH, J. P., VETTO, J. T., MURALI, R., WHITE, K. P., WHITE, C. R., JR. & BASTIAN, B. C. 2011. Assessment of copy number status of chromosomes

6 and 11 by FISH provides independent prognostic information in primary melanoma. *Am J Surg Pathol,* 35**,** 1146-50.

NURNBERG, B., GRABER, S., GARTNER, B., GEISEL, J., PFOHLER, C., SCHADENDORF, D., TILGEN, W. & REICHRATH, J. 2009. Reduced serum 25-hydroxyvitamin D levels in stage IV melanoma patients. *Anticancer Res,* 29**,** 3669-74.

OHSIE, S. J., SARANTOPOULOS, G. P., COCHRAN, A. J. & BINDER, S. W. 2008. Immunohistochemical characteristics of melanoma. *J Cutan Pathol,* 35**,** 433-44.

OLSEN, C. M., CARROLL, H. J. & WHITEMAN, D. C. 2010. Estimating the attributable fraction for melanoma: a meta-analysis of pigmentary characteristics and freckling. *Int J Cancer,* 127**,** 2430-45.

OSBORNE, J. E. & HUTCHINSON, P. E. 2002. Vitamin D and systemic cancer: is this relevant to malignant melanoma? *Br J Dermatol,* 147**,** 197-213.

PALMER, S. R., ERICKSON, L. A., ICHETOVKIN, I., KNAUER, D. J. & MARKOVIC, S. N. 2011. Circulating serologic and molecular biomarkers in malignant melanoma. *Mayo Clin Proc,* 86**,** 981-90.

PARKIN, D. M., MESHER, D. & SASIENI, P. 2011. 13. Cancers attributable to solar (ultraviolet) radiation exposure in the UK in 2010. *Br J Cancer,* 105 Suppl 2**,** S66-9.

PARMITER, A. H., BALABAN, G., CLARK, W. H., JR. & NOWELL, P. C. 1988. Possible involvement of the chromosome region 10q24----q26 in early stages of melanocytic neoplasia. *Cancer Genet Cytogenet,* 30**,** 313-7.

PASCHEN, A., SUCKER, A., HILL, B., MOLL, I., ZAPATKA, M., NGUYEN, X. D., SIM, G. C., GUTMANN, I., HASSEL, J., BECKER, J. C., STEINLE, A., SCHADENDORF, D. & UGUREL, S. 2009. Differential clinical significance of individual NKG2D ligands in melanoma: soluble ULBP2 as an indicator of poor prognosis superior to S100B. *Clin Cancer Res,* 15**,** 5208-15.

PAYETTE, M. J., KATZ, M., 3RD & GRANT-KELS, J. M. 2009. Melanoma prognostic factors found in the dermatopathology report. *Clin Dermatol,* 27**,** 53-74.

PEPE, M. S., ETZIONI, R., FENG, Z., POTTER, J. D., THOMPSON, M. L., THORNQUIST, M., WINGET, M. & YASUI, Y. 2001. Phases of biomarker development for early detection of cancer. *J Natl Cancer Inst,* 93**,** 1054-61.

PICKRELL, J. K., GAFFNEY, D. J., GILAD, Y. & PRITCHARD, J. K. 2011. False positive peaks in ChIP-seq and other sequencing-based functional assays caused by unannotated high copy number regions. *Bioinformatics,* 27**,** 2144-6.

PIRKER, C., HOLZMANN, K., SPIEGL-KREINECKER, S., ELBLING, L., THALLINGER, C., PEHAMBERGER, H., MICKSCHE, M. & BERGER, W. 2003. Chromosomal imbalances in primary and metastatic melanomas: over-representation of essential telomerase genes. *Melanoma Res,* 13**,** 483-92.

PLATZ, A., EGYHAZI, S., RINGBORG, U. & HANSSON, J. 2008. Human cutaneous melanoma; a review of NRAS and BRAF mutation frequencies in relation to histogenetic subclass and body site. *Mol Oncol,* 1**,** 395-405.

PLUMER, A., DUAN, H., SUBRAMANIAM, S., LUCAS, F. L., MIESFELDT, S., NG, A. K. & LIAW, L. 2008. Development of fragment-specific osteopontin antibodies and ELISA for quantification in human metastatic breast cancer. *BMC Cancer,* 8**,** 38.

POLLOCK, P. M., HARPER, U. L., HANSEN, K. S., YUDT, L. M., STARK, M., ROBBINS, C. M., MOSES, T. Y., HOSTETTER, G., WAGNER, U., KAKAREKA, J., SALEM, G., POHIDA, T., HEENAN, P., DURAY, P., KALLIONIEMI, O., HAYWARD, N. K., TRENT, J. M. & MELTZER, P. S. 2003. High frequency of BRAF mutations in nevi. *Nat Genet,* 33**,** 19-20.

POYNTER, J. N., ELDER, J. T., FULLEN, D. R., NAIR, R. P., SOENGAS, M. S., JOHNSON, T. M., REDMAN, B., THOMAS, N. E. & GRUBER, S. B. 2006. BRAF and NRAS mutations in melanoma and melanocytic nevi. *Melanoma Res,* 16**,** 267-73.

PRUITT, K. D., TATUSOVA, T., BROWN, G. R. & MAGLOTT, D. R. 2012. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res,* 40**,** D130-5.

R DEVELOPMENT CORE TEAM 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

RAKOSY, Z., ECSEDI, S., TOTH, R., VIZKELETI, L., HERNANDEZ-VARGAS, H., LAZAR, V., EMRI, G., SZATMARI, I., HERCEG, Z., ADANY, R. & BALAZS, M. 2013. Integrative genomics identifies gene signature associated with melanoma ulceration. *PLoS One,* 8**,** e54958.

RAKOSY, Z., VIZKELETI, L., ECSEDI, S., BEGANY, A., EMRI, G., ADANY, R. & BALAZS, M. 2008. Characterization of 9p21 copy number alterations in human melanoma by fluorescence in situ hybridization. *Cancer Genet Cytogenet,* 182**,** 116-21.

RAKOSY, Z., VIZKELETI, L., ECSEDI, S., VOKO, Z., BEGANY, A., BAROK, M., KREKK, Z., GALLAI, M., SZENTIRMAY, Z., ADANY, R. & BALAZS, M. 2007. EGFR gene copy number alterations in primary cutaneous malignant melanomas are associated with poor prognosis. *Int J Cancer,* 121**,** 1729-37.

RAMAGOPALAN, S. V., HEGER, A., BERLANGA, A. J., MAUGERI, N. J., LINCOLN, M. R., BURRELL, A., HANDUNNETTHI, L., HANDEL, A. E., DISANTO, G., ORTON, S. M., WATSON, C. T., MORAHAN, J. M., GIOVANNONI, G., PONTING, C. P., EBERS, G. C. & KNIGHT, J. C. 2010. A ChIP-seq defined genome-wide map of vitamin D receptor binding: associations with disease and evolution. *Genome Res,* 20**,** 1352-60.

RAMIREZ-MONTAGUT, T., BLACHERE, N. E., SVIDERSKAYA, E. V., BENNETT, D. C., RETTIG, W. J., GARIN-CHESA, P. & HOUGHTON, A. N. 2004. FAPalpha, a surface peptidase expressed during wound healing, is a tumor suppressor. *Oncogene,* 23**,** 5435-46.

RAMOS-VARA, J. A. 2005. Technical aspects of immunohistochemistry. *Vet Pathol,* 42**,** 405-26.

RANDERSON-MOOR, J. A., TAYLOR, J. C., ELLIOTT, F., CHANG, Y. M., BESWICK, S., KUKALIZCH, K., AFFLECK, P., LEAKE, S., HAYNES, S., KARPAVICIUS, B., MARSDEN, J., GERRY, E., BALE, L., BERTRAM, C., FIELD, H., BARTH, J. H., SILVA IDOS, S., SWERDLOW, A., KANETSKY, P. A., BARRETT, J. H., BISHOP, D. T. & BISHOP, J. A. 2009. 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,* 45**,** 3271-81.

RANGEL, J., NOSRATI, M., TORABIAN, S., SHAIKH, L., LEONG, S. P., HAQQ, C., MILLER, J. R., 3RD, SAGEBIEL, R. W. & KASHANI-SABET, M. 2008. Osteopontin as a molecular prognostic marker for melanoma. *Cancer,* 112**,** 144-50.

RAVO, M., MUTARELLI, M., FERRARO, L., GROBER, O. M., PARIS, O., TARALLO, R., VIGILANTE, A., CIMINO, D., DE BORTOLI, M., NOLA, E., CICATIELLO, L. & WEISZ, A. 2008. Quantitative expression profiling of highly degraded RNA from formalin-fixed, paraffin-embedded breast tumor biopsies by oligonucleotide microarrays. *Lab Invest,* 88**,** 430-40.

REICHRATH, J., RECH, M., MOEINI, M., MEESE, E., TILGEN, W. & SEIFERT, M. 2007. In vitro comparison of the vitamin D endocrine system in 1,25(OH)2D3-responsive and -resistant melanoma cells. *Cancer Biol Ther,* 6**,** 48-55.

REINIGER, I. W., WOLF, A., WELGE-LUSSEN, U., MUELLER, A. J., KAMPIK, A. & SCHALLER, U. C. 2007. Osteopontin as a serologic marker for metastatic uveal melanoma: results of a pilot study. *Am J Ophthalmol,* 143**,** 705-7.

RITTLING, S. R. & CHAMBERS, A. F. 2004. Role of osteopontin in tumour progression. *Br J Cancer,* 90**,** 1877-81.

RIZZO, J. M. & BUCK, M. J. 2012. Key principles and clinical applications of "next-generation" DNA sequencing. *Cancer Prev Res (Phila),* 5**,** 887-900.

ROBERT, C., THOMAS, L., BONDARENKO, I., O'DAY, S., M, D. J., GARBE, C., LEBBE, C., BAURAIN, J. F., TESTORI, A., GROB, J. J., DAVIDSON, N., RICHARDS, J., MAIO, M., HAUSCHILD, A., MILLER, W. H., JR., GASCON, P., LOTEM, M., HARMANKAYA, K., IBRAHIM, R., FRANCIS, S., CHEN, T. T., HUMPHREY, R., HOOS, A. & WOLCHOK, J. D. 2011. Ipilimumab plus dacarbazine for previously untreated metastatic melanoma. *N Engl J Med,* 364**,** 2517-26.

RODRIGUES, L. R., TEIXEIRA, J. A., SCHMITT, F. L., PAULSSON, M. & LINDMARK-MANSSON, H. 2007. The role of osteopontin in tumor progression and metastasis in breast cancer. *Cancer Epidemiol Biomarkers Prev,* 16**,** 1087-97.

RONAGHI, M., UHLEN, M. & NYREN, P. 1998. A sequencing method based on real-time pyrophosphate. *Science,* 281**,** 363, 365.

ROSE, A. E., WANG, G., HANNIFORD, D., MONNI, S., TU, T., SHAPIRO, R. L., BERMAN, R. S., PAVLICK, A. C., PAGANO, M., DARVISHIAN, F., MAZUMDAR, M., HERNANDO, E. & OSMAN, I. 2011. Clinical relevance of SKP2 alterations in metastatic melanoma. *Pigment Cell Melanoma Res,* 24**,** 197-206.

ROSS, J. S. & CRONIN, M. 2011. Whole cancer genome sequencing by next-generation methods. *Am J Clin Pathol,* 136**,** 527-39.

ROSS, M. I. 2006. Early-stage melanoma: staging criteria and prognostic modeling. *Clin Cancer Res,* 12**,** 2312s-2319s.

ROYCHOWDHURY, S., IYER, M. K., ROBINSON, D. R., LONIGRO, R. J., WU, Y. M., CAO, X., KALYANA-SUNDARAM, S., SAM, L., BALBIN, O. A., QUIST, M. J., BARRETTE, T., EVERETT, J., SIDDIQUI, J., KUNJU, L. P., NAVONE, N., ARAUJO, J. C., TRONCOSO, P., LOGOTHETIS, C. J., INNIS, J. W., SMITH, D. C., LAO, C. D., KIM, S. Y., ROBERTS, J. S., GRUBER, S. B., PIENTA, K. J., TALPAZ, M. & CHINNAIYAN, A. M. 2011. Personalized oncology through integrative high-throughput sequencing: a pilot study. *Sci Transl Med,* 3**,** 111ra121.

RYU, B., KIM, D. S., DELUCA, A. M. & ALANI, R. M. 2007. Comprehensive expression profiling of tumor cell lines identifies molecular signatures of melanoma progression. *PLoS One,* 2**,** e594.

SARNAIK, A. A., YU, B., YU, D., MORELLI, D., HALL, M., BOGLE, D., YAN, L., TARGAN, S., SOLOMON, J., NICHOL, G., YELLIN, M. & WEBER, J. S. 2011. Extended dose ipilimumab with a peptide vaccine: immune correlates associated with clinical benefit in patients with resected high-risk stage IIIc/IV melanoma. *Clin Cancer Res,* 17**,** 896-906.

SAUTER, E. R., YEO, U. C., VON STEMM, A., ZHU, W., LITWIN, S., TICHANSKY, D. S., PISTRITTO, G., NESBIT, M., PINKEL, D., HERLYN, M. & BASTIAN, B. C. 2002. Cyclin D1 is a candidate oncogene in cutaneous melanoma. *Cancer Res,* 62**,** 3200-6.

SAWYERS, C. L. 2008. The cancer biomarker problem. *Nature,* 452**,** 548-52.

SCHROEDER, A., MUELLER, O., STOCKER, S., SALOWSKY, R., LEIBER, M., GASSMANN, M., LIGHTFOOT, S., MENZEL, W., GRANZOW, M. & RAGG, T. 2006. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol,* 7**,** 3.

SCHWEIGER, M. R., KERICK, M., TIMMERMANN, B., ALBRECHT, M. W., BORODINA, T., PARKHOMCHUK, D., ZATLOUKAL, K. & LEHRACH, H.

2009. Genome-wide massively parallel sequencing of formaldehyde fixed-paraffin embedded (FFPE) tumor tissues for copy-number- and mutation-analysis. *PLoS One,* 4**,** e5548.

SCOTT, K. L., KABBARAH, O., LIANG, M. C., IVANOVA, E., ANAGNOSTOU, V., WU, J., DHAKAL, S., WU, M., CHEN, S., FEINBERG, T., HUANG, J., SACI, A., WIDLUND, H. R., FISHER, D. E., XIAO, Y., RIMM, D. L., PROTOPOPOV, A., WONG, K. K. & CHIN, L. 2009. GOLPH3 modulates mTOR signalling and rapamycin sensitivity in cancer. *Nature,* 459**,** 1085-90.

SEGURA, M. F., HANNIFORD, D., MENENDEZ, S., REAVIE, L., ZOU, X., ALVAREZ-DIAZ, S., ZAKRZEWSKI, J., BLOCHIN, E., ROSE, A., BOGUNOVIC, D., POLSKY, D., WEI, J., LEE, P., BELITSKAYA-LEVY, I., BHARDWAJ, N., OSMAN, I. & HERNANDO, E. 2009. Aberrant miR-182 expression promotes melanoma metastasis by repressing FOXO3 and microphthalmia-associated transcription factor. *Proc Natl Acad Sci U S A,* 106**,** 1814-9.

SEIFERT, M., RECH, M., MEINEKE, V., TILGEN, W. & REICHRATH, J. 2004. Differential biological effects of 1,25-dihydroxyVitamin D3 on melanoma cell lines in vitro. *J Steroid Biochem Mol Biol,* 89-90**,** 375-9.

SENGER, D. R., ASCH, B. B., SMITH, B. D., PERRUZZI, C. A. & DVORAK, H. F. 1983. A secreted phosphoprotein marker for neoplastic transformation of both epithelial and fibroblastic cells. *Nature,* 302**,** 714-5.

SENGER, D. R., PERRUZZI, C. A. & PAPADOPOULOS, A. 1989. Elevated expression of secreted phosphoprotein I (osteopontin, 2ar) as a consequence of neoplastic transformation. *Anticancer Res,* 9**,** 1291-9.

SENNELS, H. P., JACOBSEN, S., JENSEN, T., HANSEN, M. S., OSTERGAARD, M., NIELSEN, H. J. & SORENSEN, S. 2007. Biological variation and reference intervals for circulating osteopontin, osteoprotegerin, total soluble receptor activator of nuclear factor kappa B ligand and high-sensitivity C-reactive protein. *Scand J Clin Lab Invest,* 67**,** 821-35.

SERTZNIG, P., DUNLOP, T., SEIFERT, M., TILGEN, W. & REICHRATH, J. 2009a. Cross-talk between vitamin D receptor (VDR)- and peroxisome proliferator-activated receptor (PPAR)-signaling in melanoma cells. *Anticancer Res,* 29**,** 3647-58.

SERTZNIG, P., SEIFERT, M., TILGEN, W. & REICHRATH, J. 2009b. Activation of vitamin D receptor (VDR)- and peroxisome proliferator-activated receptor (PPAR)-signaling pathways through 1,25(OH)(2)D(3) in melanoma cell lines and other skin-derived cell lines. *Dermatoendocrinol,* 1**,** 232-8.

SHANNAN, B., SEIFERT, M., LESKOV, K., BOOTHMAN, D., PFOHLER, C., TILGEN, W. & REICHRATH, J. 2006. Clusterin (CLU) and melanoma growth: CLU is expressed in malignant melanoma and 1,25-dihydroxyvitamin D3 modulates expression of CLU in melanoma cell lines in vitro. *Anticancer Res,* 26**,** 2707-16.

SHLIEN, A. & MALKIN, D. 2009. Copy number variations and cancer. *Genome Med,* 1**,** 62.

SHMELKOV, E., TANG, Z., AIFANTIS, I. & STATNIKOV, A. 2011. Assessing quality and completeness of human transcriptional regulatory pathways on a genome-wide scale. *Biol Direct,* 6**,** 15.

SIM, S. H., MESSENGER, M. P., GREGORY, W. M., WIND, T. C., VASUDEV, N. S., CARTLEDGE, J., THOMPSON, D., SELBY, P. J. & BANKS, R. E. 2012. Prognostic utility of pre-operative circulating osteopontin, carbonic anhydrase IX and CRP in renal cell carcinoma. *Br J Cancer,* 107**,** 1131-7.

SIMBOLO, M., GOTTARDI, M., CORBO, V., FASSAN, M., MAFFICINI, A., MALPELI, G., LAWLOR, R. T. & SCARPA, A. 2013. DNA qualification workflow for next generation sequencing of histopathological samples. *PLoS One,* 8**,** e62692.

SIVERTSSON, A., PLATZ, A., HANSSON, J. & LUNDEBERG, J. 2002. Pyrosequencing as an alternative to single-strand conformation polymorphism analysis for detection of N-ras mutations in human melanoma metastases. *Clin Chem,* 48**,** 2164-70.

SLINGLUFF, C. L., JR., VOLLMER, R. T., REINTGEN, D. S. & SEIGLER, H. F. 1988. Lethal "thin" malignant melanoma. Identifying patients at risk. *Ann Surg,* 208**,** 150-61.

SLOMINSKI, A. T., KIM, T. K., JANJETOVIC, Z., TUCKEY, R. C., BIENIEK, R., YUE, J., LI, W., CHEN, J., NGUYEN, M. N., TANG, E. K., MILLER, D., CHEN, T. C. & HOLICK, M. 2011. 20-Hydroxyvitamin D2 is a noncalcemic analog of vitamin D with potent antiproliferative and prodifferentiation activities in normal and malignant cells. *Am J Physiol Cell Physiol,* 300**,** C526-41.

SMITH, A. P., HOEK, K. & BECKER, D. 2005. Whole-genome expression profiling of the melanoma progression pathway reveals marked molecular differences between nevi/melanoma in situ and advanced-stage melanomas. *Cancer Biol Ther,* 4**,** 1018-29.

SMYTH, G. K. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol,* 3**,** Article3.

SNITCOVSKY, I., LEITAO, G. M., PASINI, F. S., BRUNIALTI, K. C., MANGONE, F. R., MAISTRO, S., DE CASTRO, G., JR., VILLAR, R. C. & FEDERICO, M. H. 2009. Plasma osteopontin levels in patients with head and neck cancer undergoing chemoradiotherapy. *Arch Otolaryngol Head Neck Surg,* 135**,** 807-11.

SOIKKELI, J., PODLASZ, P., YIN, M., NUMMELA, P., JAHKOLA, T., VIROLAINEN, S., KROGERUS, L., HEIKKILA, P., VON SMITTEN, K., SAKSELA, O. & HOLTTA, E. 2010. Metastatic outgrowth encompasses COL-I, FN1, and POSTN up-regulation and assembly to fibrillar networks regulating cell adhesion, migration, and growth. *Am J Pathol,* 177**,** 387-403.

SOUFIR, N., AVRIL, M. F., CHOMPRET, A., DEMENAIS, F., BOMBLED, J., SPATZ, A., STOPPA-LYONNET, D., BENARD, J. & BRESSAC-DE PAILLERETS, B. 1998. Prevalence of p16 and CDK4 germline mutations in 48 melanoma-prone families in France. The French Familial Melanoma Study Group. *Hum Mol Genet,* 7**,** 209-16.

SPITTLE, C., WARD, M. R., NATHANSON, K. L., GIMOTTY, P. A., RAPPAPORT, E., BROSE, M. S., MEDINA, A., LETRERO, R., HERLYN, M. & EDWARDS, R. H. 2007. Application of a BRAF pyrosequencing assay for mutation detection and copy number analysis in malignant melanoma. *J Mol Diagn,* 9**,** 464-71.

SRINIVASAN, M., SEDMAK, D. & JEWELL, S. 2002. Effect of fixatives and tissue processing on the content and integrity of nucleic acids. *Am J Pathol,* 161**,** 1961-71.

STAGG, J. & SMYTH, M. J. 2010. Extracellular adenosine triphosphate and adenosine in cancer. *Oncogene,* 29**,** 5346-58.

STAHLECKER, J., GAUGER, A., BOSSERHOFF, A., BUTTNER, R., RING, J. & HEIN, R. 2000. MIA as a reliable tumor marker in the serum of patients with malignant melanoma. *Anticancer Res,* 20**,** 5041-4.

STARK, M. & HAYWARD, N. 2007. Genome-wide loss of heterozygosity and copy number analysis in melanoma using high-density single-nucleotide polymorphism arrays. *Cancer Res,* 67**,** 2632-42.

STATACORP. 2007. Stata Statistical Software: Release 10. College Station, TX: StataCorp LP.

STATACORP. 2011. Stata Statistical Software: Release 12. College Station, TX: StataCorp LP.

SULLIVAN, J., BLAIR, L., ALNAJAR, A., AZIZ, T., NG, C. Y., CHIPITSYNA, G., GONG, Q., WITKIEWICZ, A., WEBER, G. F., DENHARDT, D. T., YEO, C. J. & ARAFAT, H. A. 2009. Expression of a prometastatic splice variant of osteopontin, OPNC, in human pancreatic ductal adenocarcinoma. *Surgery,* 146**,** 232-40.

SUN, J., XU, H. M., ZHOU, H. J., DONG, Q. Z., ZHAO, Y., FU, L. Y., HEI, Z. Y., YE, Q. H., REN, N., JIA, H. L. & QIN, L. X. 2010. The prognostic significance of preoperative plasma levels of osteopontin in patients with TNM stage-I of hepatocellular carcinoma. *J Cancer Res Clin Oncol,* 136**,** 1-7.

SVIDERSKAYA, E. V., GRAY-SCHOPFER, V. C., HILL, S. P., SMIT, N. P., EVANS-WHIPP, T. J., BOND, J., HILL, L., BATAILLE, V., PETERS, G., KIPLING, D., WYNFORD-THOMAS, D. & BENNETT, D. C. 2003. p16/cyclin-dependent kinase inhibitor 2A deficiency in human melanocyte senescence, apoptosis, and immortalization: possible implications for melanoma progression. *J Natl Cancer Inst,* 95**,** 723-32.

SWOBODA, A., RASIN-STREDEN, D., SCHANAB, O., OKAMOTO, I., PEHAMBERGER, H., PETZELBAUER, P. & MIKULA, M. 2011. Identification of genetic disparity between primary and metastatic melanoma in human patients. *Genes Chromosomes Cancer,* 50**,** 680-8.

TAKATA, M., SUZUKI, T., ANSAI, S., KIMURA, T., SHIRASAKI, F., HATTA, N. & SAIDA, T. 2005. Genome profiling of melanocytic tumors using multiplex ligation-dependent probe amplification (MLPA): Its usefulness as an adjunctive diagnostic tool for melanocytic tumors. *J Dermatol Sci,* 40**,** 51-7.

TALANTOV, D., MAZUMDER, A., YU, J. X., BRIGGS, T., JIANG, Y., BACKUS, J., ATKINS, D. & WANG, Y. 2005. Novel genes associated with malignant melanoma but not benign melanocytic lesions. *Clin Cancer Res,* 11**,** 7234-42.

TEO, S. M., PAWITAN, Y., KU, C. S., CHIA, K. S. & SALIM, A. 2012. Statistical challenges associated with detecting copy number variations with next-generation sequencing. *Bioinformatics,* 28**,** 2711-8.

THOMPSON, F. H., EMERSON, J., OLSON, S., WEINSTEIN, R., LEAVITT, S. A., LEONG, S. P., EMERSON, S., TRENT, J. M., NELSON, M. A., SALMON, S. E. & ET AL. 1995. Cytogenetics of 158 patients with regional or disseminated melanoma. Subset analysis of near-diploid and simple karyotypes. *Cancer Genet Cytogenet,* 83**,** 93-104.

THOMPSON, J. F., SOONG, S. J., BALCH, C. M., GERSHENWALD, J. E., DING, S., COIT, D. G., FLAHERTY, K. T., GIMOTTY, P. A., JOHNSON, T., JOHNSON, M. M., LEONG, S. P., ROSS, M. I., BYRD, D. R., CASCINELLI, N., COCHRAN, A. J., EGGERMONT, A. M., MCMASTERS, K. M., MIHM, M. C., JR., MORTON, D. L. & SONDAK, V. K. 2011. Prognostic significance of mitotic rate in localized primary cutaneous melanoma: an analysis of patients in the multi-institutional American Joint Committee on Cancer melanoma staging database. *J Clin Oncol,* 29**,** 2199-205.

TOKURA, Y., BASTIAN, B. C. & DUNCAN, L. 2006. Melanocytic tumours. Acral-lentiginous melanoma. *World Health Organisation Classification of Tumours. Pathology and Genetics of Skin Tumours.* Lyon, France: IARC press.

TSAO, H., CHIN, L., GARRAWAY, L. A. & FISHER, D. E. 2012. Melanoma: from mutations to medicine. *Genes Dev,* 26**,** 1131-55.

TURAJLIC, S., FURNEY, S. J., LAMBROS, M. B., MITSOPOULOS, C., KOZAREWA, I., GEYER, F. C., MACKAY, A., HAKAS, J., ZVELEBIL, M., LORD, C. J., ASHWORTH, A., THOMAS, M., STAMP, G., LARKIN, J., REIS-FILHO, J. S. & MARAIS, R. 2012. Whole genome sequencing of matched primary and metastatic acral melanomas. *Genome Res,* 22**,** 196-207.

U.S. NATIONAL INSTITUTES OF HEALTH. 2013. Available: http://clinicaltrials.gov/ [Accessed 20 August 2013].

UDART, M., UTIKAL, J., KRAHN, G. M. & PETER, R. U. 2001. Chromosome 7 aneusomy. A marker for metastatic melanoma? Expression of the epidermal growth factor receptor gene and chromosome 7 aneusomy in nevi, primary malignant melanomas and metastases. *Neoplasia,* 3**,** 245-54.

UGUREL, S., HOUBEN, R., SCHRAMA, D., VOIGT, H., ZAPATKA, M., SCHADENDORF, D., BROCKER, E. B. & BECKER, J. C. 2007. Microphthalmia-associated transcription factor gene amplification in metastatic melanoma is a prognostic marker for patient survival, but not a predictive marker for chemosensitivity and chemotherapy response. *Clin Cancer Res,* 13**,** 6344-50.

UTIKAL, J., SCHADENDORF, D. & UGUREL, S. 2007. Serologic and immunohistochemical prognostic biomarkers of cutaneous malignancies. *Arch Dermatol Res,* 298**,** 469-77.

VAN AKKOOI, A. C., NOWECKI, Z. I., VOIT, C., SCHAFER-HESTERBERG, G., MICHEJ, W., DE WILT, J. H., RUTKOWSKI, P., VERHOEF, C. & EGGERMONT, A. M. 2008. Sentinel node tumor burden according to the Rotterdam criteria is the most important prognostic factor for survival in melanoma patients: a multicenter study in 388 patients with positive sentinel nodes. *Ann Surg,* 248**,** 949-55.

VAN RAAMSDONK, C. D., BEZROOKOVE, V., GREEN, G., BAUER, J., GAUGLER, L., O'BRIEN, J. M., SIMPSON, E. M., BARSH, G. S. & BASTIAN, B. C. 2009. Frequent somatic mutations of GNAQ in uveal melanoma and blue naevi. *Nature,* 457**,** 599-602.

VOGEL, C. & MARCOTTE, E. M. 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet,* 13**,** 227-32.

WASSERMAN, W. W. & SANDELIN, A. 2004. Applied bioinformatics for the identification of regulatory elements. *Nat Rev Genet,* 5**,** 276-87.

WELLBROCK, C. & MARAIS, R. 2005. Elevated expression of MITF counteracts B-RAF-stimulated melanocyte and melanoma cell proliferation. *J Cell Biol,* 170**,** 703-8.

WERNER, T. 2008. Bioinformatics applications for pathway analysis of microarray data. *Curr Opin Biotechnol,* 19**,** 50-4.

WHEATLEY-PRICE, P., YANG, B., PATSIOS, D., PATEL, D., MA, C., XU, W., LEIGHL, N., FELD, R., CHO, B. C., O'SULLIVAN, B., ROBERTS, H., TSAO, M. S., TAMMEMAGI, M., ANRAKU, M., CHEN, Z., DE PERROT, M. & LIU, G. 2010. Soluble mesothelin-related Peptide and osteopontin as markers of response in malignant mesothelioma. *J Clin Oncol,* 28**,** 3316-22.

WHITEMAN, D. C., PAVAN, W. J. & BASTIAN, B. C. 2011. The melanomas: a synthesis of epidemiological, clinical, histopathological, genetic, and biological aspects, supporting distinct subtypes, causal pathways, and cells of origin. *Pigment Cell Melanoma Res,* 24**,** 879-97.

WILTSHIRE, R. N., DENNIS, T. R., SONDAK, V. K., MELTZER, P. S. & TRENT, J. M. 2001. Application of molecular cytogenetic techniques in a case study of human cutaneous metastatic melanoma. *Cancer Genet Cytogenet,* 131**,** 97-103.

WIND, T. C., MESSENGER, M. P., THOMPSON, D., SELBY, P. J. & BANKS, R. E. 2011. Measuring carbonic anhydrase IX as a hypoxia biomarker: differences in concentrations in serum and plasma using a commercial enzyme-linked immunosorbent assay due to influences of metal ions. *Ann Clin Biochem,* 48**,** 112-20.

WINNEPENNINCKX, V., LAZAR, V., MICHIELS, S., DESSEN, P., STAS, M., ALONSO, S. R., AVRIL, M. F., ORTIZ ROMERO, P. L., ROBERT, T., BALACESCU, O., EGGERMONT, A. M., LENOIR, G., SARASIN, A., TURSZ, T., VAN DEN OORD, J. J. & SPATZ, A. 2006. Gene expression

profiling of primary cutaneous melanoma and clinical outcome. *J Natl Cancer Inst,* 98**,** 472-82.

WOOD, H. M., BELVEDERE, O., CONWAY, C., DALY, C., CHALKLEY, R., BICKERDIKE, M., MCKINLEY, C., EGAN, P., ROSS, L., HAYWARD, B., MORGAN, J., DAVIDSON, L., MACLENNAN, K., ONG, T. K., PAPAGIANNOPOULOS, K., COOK, I., ADAMS, D. J., TAYLOR, G. R. & RABBITTS, P. 2010. Using next-generation sequencing for high resolution multiplex analysis of copy number variation from nanogram quantities of DNA from formalin-fixed paraffin-embedded specimens. *Nucleic Acids Res,* 38**,** e151.

WU, H., GOEL, V. & HALUSKA, F. G. 2003. PTEN signaling pathways in melanoma. *Oncogene,* 22**,** 3113-22.

YANCOVITZ, M., LITTERMAN, A., YOON, J., NG, E., SHAPIRO, R. L., BERMAN, R. S., PAVLICK, A. C., DARVISHIAN, F., CHRISTOS, P., MAZUMDAR, M., OSMAN, I. & POLSKY, D. 2012. Intra- and inter-tumor heterogeneity of BRAF(V600E))mutations in primary and metastatic melanoma. *PLoS One,* 7**,** e29336.

ZHANG, J., CHIODINI, R., BADR, A. & ZHANG, G. 2011. The impact of next-generation sequencing on genomics. *J Genet Genomics,* 38**,** 95-109.

ZHOU, Y., DAI, D. L., MARTINKA, M., SU, M., ZHANG, Y., CAMPOS, E. I., DOROCICZ, I., TANG, L., HUNTSMAN, D., NELSON, C., HO, V. & LI, G. 2005. Osteopontin expression correlates with melanoma invasion. *J Invest Dermatol,* 124**,** 1044-52.