THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Dissecting the genetic architecture of familial melanoma

Lauren Gabriel Aoude

Bachelor of Engineering (Mechanical)

Bachelor of Arts (European Studies)

A thesis submitted for the degree of Doctor of Philosophy at

The University of Queensland in 2014

School of Medicine / QIMR Berghofer Medical Research Institute

Abstract

Cutaneous malignant melanoma (CMM) is a disease that has had a significant impact on the lives of Australians, affecting 1 in 17 people. Around 10% of those diagnosed have a family history of melanoma and in the majority of these cases there is underlying predisposition that makes them more susceptible to melanoma development. Known high-risk loci (*CDKN2A*, *CDK4*) account for approximately half of these families. There remains, therefore, a large gap in knowledge regarding the causes of susceptibility to CMM.

Uveal malignant melanoma (UMM) is a rare type of melanoma that forms in the uveal tract of the eye. Unlike cutaneous melanoma, ultraviolet radiation does not contribute to the formation of UMM. To date, only one gene has been identified that contributes to risk of uveal melanoma: *BAP1*.

The main objective of this thesis was to discover the underlying germline mutations that contribute to increased risk in high-density CMM and UMM families lacking mutations in any of the known melanoma risk loci. The key techniques that were used to achieve this were whole-genome and exome sequencing of multiple cases from these families. By targeting novel variants that segregated with disease in families, truncating mutations leading to CMM development were found in *POT1, ACD* and *TERF2IP* (Chapters 4 and 5). This technique was also used to discover truncating germline *BAP1* mutations in a UMM family from Denmark (Chapter 7). A second method of gene discovery that was used was the candidate gene approach. This first involved a literature search to discover likely candidates for melanoma susceptibility. The protein-coding regions of the chosen genes were then Sanger sequenced to identify novel variants in probands from melanoma families. This method was used to interrogate the role of *PALB2, SOX10* and *MITF* in CMM predisposition (Chapter 3).

The contribution of known melanoma predisposition genes *CDKN2A, CDK4* and *BAP1* to a population-based CMM sample from Queensland was also assessed. This was achieved through a combination of targeted pull-down of the selected genes using the Ion Torrent Personal Genome Machine, and also single nucleotide polymorphism analysis using Sequenom mass spectrometry. Chapters 2 and 6 of this thesis detail the contribution of these genes to CMM and UMM predisposition in population-based samples from Australia.

Overall this project has made a major contribution to the field of familial melanoma genetics. We described germline mutations in members of the shelterin complex (*POT1, ACD* and *TERF2IP*) in CMM, which accounts for 9% of families lacking mutations in previously known risk loci

CDKN2A, and *CDK4*. In combination with a previous finding of mutations in *TERT*, the work that is presented in this thesis has contributed to the discovery of a new pathway related to melanoma predisposition and thus represents a significant advance in the field.

Additionally, we have estimated the contribution of known risk genes *CDKN2A* and *BAP1* to susceptibility in a population-based sample of Queensland melanoma cases (N=1,109). We showed that *CDKN2A* accounts for around 1.31% of Queensland melanoma families and *BAP1* accounts for around 0.63%. Additionally, we showed *CDK4* mutations are extremely rare, not being found in 1,550 melanoma families from Queensland.

In summary, next-generation sequencing techniques have been an invaluable tool in melanoma predisposition gene discovery. We have identified several new genes that contribute to melanoma susceptibility when mutated in the familial setting. By discovering and characterising these genes, better surveillance and treatment strategies may be available in the future for individuals with familial melanoma.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the General Award Rules of The University of Queensland, immediately made available for research and study in accordance with the *Copyright Act 1968*.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

Publications during candidature

Lauren G. Aoude, Michael Gartside, Peter Johansson, Jane M. Palmer, Judith Symmons, Nicholas G. Martin, Grant W. Montgomery, Nicholas K. Hayward. Prevalence of germline *BAP1, CDKN2A* and *CDK4* mutations in an Australian population-based sample of cutaneous melanoma cases. *Twin Research and Human Genetics* (manuscript submitted 02/10/2014).

Lauren G. Aoude, Karin A. W. Wadt and Nicholas K. Hayward. Genetics of Familial Melanoma: 20 years after CDKN2A. *Pigment Cell & Melanoma Research* (manuscript submitted 30/09/2014).

Karin A. W. Wadt, **Lauren G. Aoude**, Lisa Golmard, Thomas v. O. Hansen, Xavier Sastre-Garau, Nicholas K. Hayward, Anne-Marie Gerdes. Germline *RAD51B* truncating mutation in a family with cutaneous melanoma. *BMC Cancer* (manuscript submitted 11/06/2014).

Lauren G. Aoude*, Antonia L. Pritchard*, Carla Daniela Robles-Espinoza*, Karin Wadt*, Mark Harland*, Jiyeon Choi, Michael Gartside, Víctor Quesada, Peter Johansson, Jane M. Palmer, Andrew J. Ramsay, Xijun Zhang, Kristine Jones, Judith Symmons, Elizabeth A. Holland, Helen Schmid, Vanessa Bonazzi, Susan Woods, Ken Dutton-Regester, Mitchell S. Stark, Helen Snowden, Remco van Doorn, Grant W. Montgomery, Nicholas G. Martin, Thomas M. Keane, Carlos López-Otín, Anne-Marie Gerdes, Håkan Olsson, Christian Ingvar, Åke Borg, Nelleke A. Gruis, Jeffrey M. Trent, Göran Jönsson, D. Timothy Bishop, Graham J. Mann, Julia A. Newton-Bishop, Kevin M. Brown, David J. Adams, Nicholas K. Hayward. Nonsense mutations in the shelterin complex genes *ACD* and *TERF2IP* in familial melanoma. *Journal of the National Cancer Institute* (in press).

Karin A. W. Wadt*, **Lauren G. Aoude***, Peter Johansson, Annalisa Gatt, Antonia Pritchard, Oana Crainic, Morten Tolstrup Andersen, Jens F Kiilgaard, Steffen Heegaard, Lone Sunde, Birgitte Federspiel, John Thompson, Stan McCarthy, Steven Kao, Anne-Marie Gerdes, Hensin Tsao, Göran Jönsson, Klaus Busam, Ruta Gupta, Jeffrey M. Trent, Kevin M. Brown, Richard Scolyer, Nicholas K. Hayward. A recurrent germline *BAP1* mutation in melanoma families from two continents and extension of the *BAP1* tumour predisposition spectrum to include basal cell carcinoma. *Clinical Genetics*. 2014 Sep; doi:10.1111/cge.12501

Lauren G. Aoude, Mai Xu, Zhen Zhen Zhao, Michael Kovacs, Jane M. Palmer, Peter Johansson, Judith Symmons, Jeffrey M. Trent, Nicholas G. Martin, Grant W. Montgomery, Kevin M. Brown and Nicholas K. Hayward. Assessment of *PALB2* as a candidate melanoma susceptibility gene. *PLOS One*. 2014 June; 9(6):e100683.

Carla Daniela Robles-Espinoza*, Mark Harland*, Andrew J. Ramsay*, **Lauren G. Aoude***, Víctor Quesada, Zhihao Ding, Antonia L. Pritchard, Jessamy C. Tiffen, Mia Petljak, Jane M. Palmer, Judith Symmons, Peter Johansson, Mitchell S. Stark, Michael G. Gartside, Helen Snowden, Grant W. Montgomery, Nicholas G. Martin, Jimmy Z. Liu, Jiyeon Choi, Matthew Makowski, Kevin M. Brown, Thomas M. Keane, Carlos López-Otín, Nelleke A. Gruis, Nicholas K. Hayward, D. Timothy Bishop, Julia A. Newton-Bishop, David J. Adams. *POT1* loss-of-function variants predispose to familial melanoma. *Nature Genetics*. 2014 May; 46(5):478-81.

Ken Dutton-Regester, Hojabr Kakavand, **Lauren G. Aoude**, Mitchell S. Stark, Michael G. Gartside, Peter Johansson, Linda O'Connor, Cathy Lanagan, Varsha Tembe, Gulietta M. Pupo, Lauren E. Haydu, Christopher W. Schmidt, Graham J. Mann, John F. Thompson, Richard A. Scolyer, Nicholas K. Hayward. Melanomas of unknown primary have a mutation profile consistent with cutaneous sun-exposed melanoma. *Pigment Cell & Melanoma Research*. 2013 Nov; 26(6):852-60.

Julia C. Cronin, Dawn E. Watkins-Chow, Art Incao, Joanne H. Hasskamp, Nicola Schönewolf, **Lauren G. Aoude**, Nicholas K. Hayward, Boris C. Bastian, Reinhard Dummer, Stacie K. Loftus, William J. Pavan. SOX10 ablation arrests cell cycle, induces senescence, and suppresses melanomagenesis. *Cancer Research*. 2013 Sep; 73(18):5709-18.

Lauren G. Aoude, Karin Wadt, Anders Bojesen, Dorthe Crüger, Åke Borg, Jeffrey M. Trent, Kevin M. Brown, Anne-Marie Gerdes, Göran Jönsson, Nicholas K. Hayward. A *BAP1* mutation in a Danish family predisposes to uveal melanoma and other cancers. *PLOS One*. 2013 Aug; 8(8):e72144.

Lauren G. Aoude, Claire M. Vajdic, Anne Kricker, Bruce Armstrong, Nicholas K. Hayward. Prevalence of germline *BAP1* mutation in a population-based sample of uveal melanoma cases. *Pigment Cell & Melanoma Research*. 2013 Mar; 26(2):278-9.

Ken Dutton-Regester, **Lauren Aoude**, Derek J. Nancarrow, Mitchell S Stark, Linda O'Connor, Cathy Lanagan, Gulietta M. Pupo, Varsha Tembe, Candace D. Carter, Michael O'Rourke, Richard A. Scolyer, Graham J. Mann, Chris Schmidt, Adrian Herington, Nicholas K. Hayward. Identification of *TFG* (TRK fused gene) as a putative metastatic melanoma tumour suppressor gene. *Genes, Chromosomes and Cancer*. 2012 May; 51(5):452-61.

Ken Dutton-Regester, Darryl Irwin, Priscilla Hunt, **Lauren G. Aoude**, Varsha Tembe, Gulietta M. Pupo, Cathy Lanagan, Candice D. Carter, Linda O'Connor, Michael O'Rourke, Richard A. Scolyer, Graham J. Mann, Christopher W. Schmidt, Adrian Herington, Nicholas K. Hayward. A high throughput panel for identifying clinically-relevant mutation profiles in melanoma. *Molecular Cancer Therapeutics*. 2012 Apr; 11(4):888-97.

Mitchell S. Stark,* Ken Dutton-Regester,* Susan L. Woods,* Michael Gartside,* Vanessa F. Bonazzi, **Lauren G. Aoude**, Donna Muzny, Donald Chow, Jonathan J. Ellis, Jeffrey Reid, Victoria Zismann, Sonika Tyagi, Irene Newsham, YuanQing Wu, Jane M. Palmer, Thomas Pollak, David Youngkin, Bradford R. Brooks, Chris Sereduk, Catherine Lanagan, Christopher W. Schmidt, Bostjan Kobe, Hongwei Yin, Kevin M. Brown, Richard Gibbs, Jeffrey Trent, Nicholas K. Hayward. Frequent somatic mutations in *MAP3K5* and *MAP3K5* in metastatic melanoma identified by exome sequencing. *Nature Genetics*. 2011 Dec; 44(2):165-9.

Satoru Yokoyama*, Susan L. Woods*, Glen M. Boyle*, **Lauren G. Aoude*,** Stuart MacGregor*, Victoria Zismann*, Michael Gartside, Anne E. Cust, Rizwan Haq, Mark Harland, John Taylor, David L. Duffy, Kelly Holohan, Ken Dutton-Regester, Jane M. Palmer, Vanessa Bonazzi, Mitchell S. Stark, Judith Symmons, Matthew H. Law, Christopher Schmidt, Cathy Lanagan, Linda O'Connor, Elizabeth A. Holland, Helen Schmid, Judith A. Maskiell, Jodie Jetann, Megan Ferguson, Mark A. Jenkins, Richard F. Kefford, Graham G. Giles, Bruce K. Armstrong, Joanne F. Aitken, John L. Hopper, David C. Whiteman, Paul D. Pharoah, Douglas F. Easton, Alison M. Dunning, Julia A. Newton-Bishop, Grant W. Montgomery, Nicholas G. Martin, Graham J. Mann, D. Timothy Bishop, Hensin Tsao, Jeffrey M. Trent, David E. Fisher, Nicholas K. Hayward, Kevin M. Brown. A novel recurrent mutation in *MITF* predisposes to familial and sporadic melanoma. *Nature*. 2011 Nov; 480(7375):99-103.

***Joint first author**

Publications included in this thesis

Lauren G. Aoude, Michael Gartside, et al. Prevalence of germline *BAP1, CDKN2A* and *CDK4* mutations in a population-based sample of cutaneous melanoma cases. *Twin Research and Human Genetics* (manuscript submitted 02/10/2014) – Incorporated as Chapter 2

Contributor	Statement of contribution
Lauren Aoude (candidate)	Designed experiments (50%) Performed experiments: Ion Torrent, iPLEX, Sanger sequencing (90%) Wrote the manuscript (100%)
Michael Gartside	Performed experiments: Ion Torrent sequencing (10%)
Peter Johansson	Bioinformatics (100%)
Jane Palmer, Judith Symmons	Collected patient data (50% each)
Nicholas Martin, Grant Montgomery	Sample contribution (33% each)
Nicholas Hayward	Designed experiments (50%) Sample contribution (33 %) Edited the manuscript (100%)

Lauren G. Aoude, Mai Xu, et al. Assessment of *PALB2* as a candidate melanoma susceptibility gene. *PLOS One*. 2014 Jun; 9(6):e100683 – Incorporated as Chapter 3

Contributor	Statement of contribution
Lauren Aoude (candidate)	Designed experiments (33%) Performed experiments: NGS, co-segregation analysis, case-control analysis, qPCR (60%) Wrote the manuscript (100%)
Mai Xu	Performed experiments: Sanger sequencing (25%)
Zhen Zhen Zhao	Performed experiments: Case-control analysis (10%)
Michael Kovacs	Performed experiments: Sanger sequencing (5%)
Jane Palmer, Judith Symmons	Collected patient data (50% each)
Peter Johansson	Copy number analysis (50%)
Jeffrey Trent, Nicholas Martin, Grant Montgomery	Sample contribution (25% each)
Kevin Brown	Designed experiments (33%)
Nicholas Hayward	Designed experiments (33%) Sample contribution (25%) Edited the manuscript (100%)

Carla Daniela Robles-Espinoza*, Mark Harland*, Andrew J. Ramsay*, **Lauren G. Aoude***, et al. *POT1* loss-of-function variants predispose to familial melanoma. *Nature Genetics*. 2014 May; 46(5):478-81 – Incorporated as Chapter 4

Contributor	Statement of contribution
Lauren Aoude (candidate)	Designed experiments (10%) Performed experiments: NGS, Sanger sequencing, co-segregation analysis, RT-PCR (40%) Data analysis (30%) Wrote methods section and figures and edited the manuscript (15%)
Carla Daniela Robles-Espinoza	Designed experiments (10%) Bioinformatics (25%) Data analysis (40%) Wrote the manuscript (30%)
Andrew Ramsay	Protein modelling (50%) Performed experiments: EMSAs (15%)
Mark Harland	Performed experiments (10%) Collected patient data (33%) Edited the manuscript (5%)
Víctor Quesada	Protein modelling (50%)
Jiyeon Choi, Jimmy Liu, Matthew Makowski, Alison Dunning, Mia Petljak, Karen Pooley	Performed experiments (5% each)
Mitchell Stark, Michael Gartside, Jessamy Tiffen, Antonia Pritchard, Carlos López-Otín, Kevin Brown	Data analysis (5% each)
Zhihao Ding, Peter Johansson, Thomas Keane	Bioinformatics (25% each)
Jane Palmer, Judith Symmons	Collected patient data (33% each)
Nicholas Martin, Grant Montgomery, Nelleke Gruis, Helen Snowden,	Sample contribution (20% each)
Timothy Bishop, Julia Newton-Bishop	Designed experiments (20% each) Wrote and edited the manuscript (5% each)
Nicholas Hayward	Designed experiments (20%) Sample contribution (20%) Wrote and edited the manuscript (10%)
David Adams	Designed experiments (20%) Wrote and edited the manuscript (30%)

Lauren G. Aoude*, Antonia L. Pritchard*, et al. Nonsense mutations in the shelterin complex genes *ACD* and *TERF2IP* in familial melanoma. *Journal of the National Cancer Institute* (in press) – Incorporated as Chapter 5

Contributor	Statement of contribution
Lauren Aoude (candidate)	Designed experiments (20%) Performed experiments: NGS, Sanger sequencing, co-segregation analysis (65%) Data Analysis (20%) Wrote the manuscript, created tables and figures (40%)
Antonia Pritchard	Designed experiments (20%) Data Analysis (20%) Performed experiments: functional work (25%) Wrote the manuscript (30%)
Carla Daniela Robles-Espinoza	Bioinformatics (33%) Data Analysis (20%) Edited the manuscript (40%)
Karin Wadt	Collected patient data (20%) Edited the manuscript (5%)
Mark Harland	Collected patient data (20%) Performed experiments: co-segregation analysis (10%) Edited the manuscript (5%)
Michael Gartside, Vanessa Bonazzi, Susan Woods, Ken Dutton-Regester, Mitchell Stark, Remco van Doorn, Xijun Zhang, Kristine Jones	Acquisition of data (12% each)
Víctor Quesada, Andrew Ramsay, Carlos López-Otín	Protein modelling (33% each)
Peter Johansson, Thomas Keane	Bioinformatics (33% each)
Jane Palmer, Judith Symmons, Elizabeth Holland, Helen Schmid, Göran Jönsson	Collected patient data (12% each)
Julia Newton-Bishop, Timothy Bishop, Graham Mann, Nicholas Martin, Grant Montgomery, Helen Snowden, Nelleke Gruis, Anne-Marie Gerdes, Håkan Olsson, Christian Ingvar, Åke Borg, Jeffrey Trent	Sample contribution (8% each)
Kevin Brown, Jiyeon Choi	Data Analysis (20% each) Edited the manuscript (5% each)
David Adams	Designed experiments (30%) Edited the manuscript (40%)
Nicholas Hayward	Designed experiments (30%) Sample contribution (8%) Wrote the manuscript (30%)

Lauren G. Aoude, Claire M. Vajdic, et al. Prevalence of germline *BAP1* mutation in a population-based sample of uveal melanoma cases. *Pigment Cell & Melanoma Research.* 2013 Mar; 26(2):278-9 – Incorporated as Chapter 6

Contributor	Statement of contribution
Lauren Aoude (candidate)	Designed the experiments (50%) Performed Sanger sequencing experiments (100%) Wrote the manuscript (90%)
Claire Vajdic	Collected patient data (100%)
Anne Kricker, Bruce Armstrong	Sample contribution (50% each)
Nicholas Hayward	Designed the experiments (50%) Wrote and edited the manuscript (10%)

Lauren G. Aoude, Karin Wadt, et al. A *BAP1* mutation in a Danish family predisposes to uveal melanoma and other cancers. *PLOS One*. 2013 Aug; 8(8):e72144 – Incorporated as Chapter 7

Contributor	Statement of contribution
Lauren Aoude (candidate)	Designed the experiments (15%) Performed experiments: NGS, co-segregation analysis, case-control analysis (90%) Wrote and edited the manuscript (90%)
Karin Wadt	Collected clinical data (100%)
Anders Bojesen, Dorthe Crüger, Åke Borg, Anne-Marie Gerdes	Sample contribution (25% each)
Jeffrey Trent, Kevin Brown	Designed the experiments (15% each)
Göran Jönsson	Designed the experiments (15%) Performed Experiments: RT-PCR (10%)
Nicholas Hayward	Designed the experiments (40%) Wrote and edited the manuscript (10%)

Contributions by others to the thesis

This has been a collaborative effort with many international groups pooling their samples and next-generation sequencing data to produce high-impact research. The Queensland Familial Melanoma Project sample collection that formed the framework for much of this thesis was led by Nicholas Hayward, Nicholas Martin and Grant Montgomery. This has been an invaluable sample resource providing DNA on approximately 1,690 probands from Queensland melanoma families. The members of GenoMEL also provided many samples, in particular: Kevin Brown, Göran Jönsson and Karin Wadt, who have supplied clinical data and DNA samples from their respective research groups. Peter Johansson provided the bioinformatics support and generated the pipeline through which all raw sequence data was processed.

Statement of parts of the thesis submitted to qualify for the award of another degree

None.

Acknowledgements

I could not have accomplished a project of this size without the support of the people around me. I would firstly like to thank my principal supervisor Nicholas Hayward. This has been a challenging but very rewarding experience. I have developed so much as a researcher and I owe it to his guidance and support over the past three years. I owe a huge thank you to everyone in the Oncogenomics research group. I am very lucky to be part of such an open and supportive group. Regardless of what is going on, I enjoy coming to work every day and that is something I do not take for granted. A special thanks to Jane Palmer for her tireless efforts chasing patient data. I would also like to thank Stuart MacGregor for his support as my secondary supervisor throughout the course of my PhD. A big thank you to Antonia Pritchard for not only being a huge help in the lab, but also being such a great friend and keeping me sane through it all.

I am extremely grateful for the financial support that allowed me to pursue this path. I would like to thank the Australia and New Zealand Banking Group trustees and the Katherine Hannay Estate for providing me with PhD scholarship support throughout my candidature. I am also thankful to QIMR Berghofer for providing me with a PhD top-up scholarship and supporting my conference travel. I would also like to thank the Cancer Council Queensland for providing me with financial support for conference travel abroad; these opportunities to travel and learn from leaders in the field are invaluable as a student.

Lastly, I would like to say a huge thank you to my family and friends for all their love and support over these past three years. Mum, Dad and Anthony you have always supported me in my studies and encouraged me to pursue my dreams. To my wonderful husband Johnno, thank you for understanding when I've spent weekends at work when I could have been home with you instead. You have supported me through the whole process and somehow we planned a wedding in the middle of it all! I love you dearly and I couldn't have done it without you by my side.

Keywords

Cutaneous melanoma, uveal melanoma, germline mutation, next-generation sequencing

Australian and New Zealand standard research classifications (ANZSRC)

ANZSRC code: 111203, Cancer Genetics, 100%

Fields of research (FoR) classification

FoR code: 0604, Biomedical Sciences: Genetics, 100%

Table of contents

List of figures

List of tables

List of abbreviations used in this thesis

ALM	acral lentiginous melanoma
cDNA	complementary deoxyribonucleic acid
CLL	chronic lymphocytic leukaemia
CMM	cutaneous malignant melanoma
CNV	copy number variation
COSMIC	Catalogue of Somatic Mutations in Cancer
DKC	dyskeratosis congenita
DNA	deoxyribonucleic acid
dsDNA	double-stranded DNA
ESP6500	NHLBI Exome Sequencing Project
GWAS	genome-wide association study
LCL	lymphoblastoid cell line
LMM	lentigo maligna melanoma
LOVD	Leiden Open (source) Variant Database
MAF	minor allele frequency
NGS	next-generation sequencing
NM	nodular melanoma
OB	oligonucleotide/oligosaccharide-binding
OT2	One Touch 2
PCR	polymerase chain reaction
PGM	Personal Genome Machine
QIMR	Queensland Institute of Medical Research
qPCR	quantitative polymerase chain reaction
SNP	single nucleotide polymorphism
ssDNA	single-stranded DNA
SSM	superficial spreading melanoma
RT-PCR	reverse-transcription polymerase chain reaction
UMM	uveal malignant melanoma
UVR	ultraviolet radiation

Chapter 1

Introduction

1.1 Significance of melanoma

Melanoma is a cancer that forms from melanocytes, occurring when genetic and epigenetic events lead to the development of malignancy. There are several forms of skin cancer but melanoma is the most aggressive, accounting for 75% of skin cancer deaths in Australia despite making up only 2.3% of all skin cancers [1]. Caught in the early stages (I and II) it can be effectively treated with local excision; however, by the time it progresses to stage IV, patient prognosis is poor, with a five-year survival rate of less than 10% [2]. This disease affects the lives of people globally, but nowhere more so than in Australia, with Queensland recording the highest incidence of invasive cutaneous melanoma in the world [3]. The Australian Institute of Health and Welfare reported over 12,510 new cases in 2012, making it the fourth most common cancer nationally. The risk of being diagnosed with melanoma by the age of 85, in Australia, is 1 in 14 for males and 1 in 24 for females [4]. It is also the most common cancer type found in adolescents and young adults, with 2,251 cases diagnosed in the 15–39 year old age group in the period 2003–2007. This figure accounts for about one quarter of all cancers in this demographic [5]. It is clear that this disease has a profound effect in Australia and across the world, and effective screening and treatment strategies are needed to combat the burden of this disease. Identifying mutation events that lead to melanoma development is one step on this path.

There are many environmental and genetic factors that play a part in melanomagenesis. While exposure to ultraviolet radiation (UVR) has a significant role in melanoma development, there can also be an underlying predisposition that contributes to a person's risk. Studies have shown that around 10% of melanoma cases have a family history of the disease [6-11]. Currently, known germline variants account for approximately half of the high-density familial cases, leaving a large proportion of families with as yet unidentified genetic risk factors. The discovery of these could ultimately lead to better treatment strategies and prevention plans for at-risk individuals and their families.

1.2 Melanoma subtypes

Cutaneous malignant melanoma (CMM) is a form of skin cancer that develops when unregulated growth of melanocytes leads to the formation of tumours. There are four subtypes by which primary tumours are classified. This stratifies tumours according to their anatomic location and their growth features. The subtypes are: superficial spreading melanoma (SSM), lentigo maligna melanoma (LMM), nodular melanoma (NM) and acral lentiginous melanoma (ALM) [12, 13].

SSM is the most common CMM subtype, making up approximately 70% of all cases. These melanomas can be found on any part of the body and often arise in a pre-existing naevus, and then spread along the epidermis before becoming invasive (Figure 1). Phenotypically, they present with irregular borders and varied colours [14].

LMM is a slow-growing neoplasm, usually occurring on the chronically sun-exposed areas of the head and neck. They develop from a freckle-like lesion (Hutchinson's melanotic freckle) into invasive melanoma [15]. The histological features of this type of CMM are atypical melanocytes that proliferate at the dermal-epidermal junction. LMM accounts for about 1% of CMM cases and usually develops later in life, with an age of diagnosis often in the 60s or 70s [16].

NM is an invasive melanoma that is generally black in appearance (though it can present with other colours on occasion). This is the most aggressive form of melanoma and accounts for around 15% of cases, usually occurring in older individuals [14].

Lastly, ALM is melanoma originating in the epidermis of the palms, soles of the feet or in the nail bed. It is the most common form of melanoma in Asian and African-American populations but is rarer in Caucasian populations, making up about 5% of CMM cases in the latter group [14].

There are two types of non-cutaneous melanoma: mucosal and uveal. Unlike CMM, these melanoma types are not related to UVR exposure. Mucosal melanoma is quite rare, accounting for only 2% of all melanoma cases. This melanoma develops in the mucosal membranes of the body, with the most common areas being the nasal passage and oral cavity. Uveal malignant melanoma (UMM) occurs in the uveal tract of the eye. This is also a very rare melanoma subtype, occurring in around 5.6 per million people in the USA [17]. Refer to Chapter 1 section 1.7 for a more in-depth introduction to UMM.

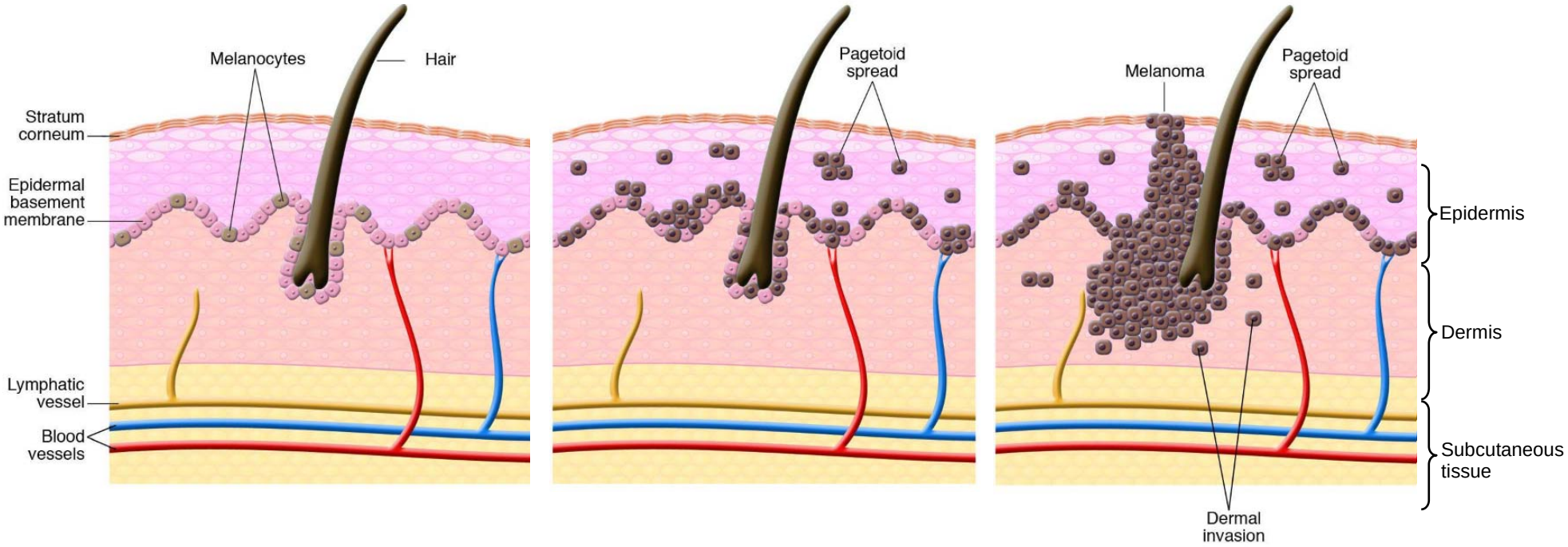


Figure 1: Histological progression from normal skin to melanoma. Evolution from normal skin (left panel) to radial growth phase (middle panel) to vertical growth phase (right panel). Pagetoid spread is the migration of melanocytes into the epidermis. Adapted from Chudnovsky et al., 2005.

1.3 Pathways to familial cutaneous malignant melanoma development

Approximately 10% of CMM cases occur in individuals that have a relative with melanoma and approximately 1% of cases have a strong history of melanoma (two or more first degree relatives with the disease). Familial CMM is known to have both genetic and environmental risk factors.

Firstly, familial melanoma may cluster by chance due to latitude-dependent ambient UVR. A case study of capital cities along the Eastern coast of Australia shows this effect well [18]. The trend in melanoma incidence increases as latitude decreases, even though the population demographic of each location is roughly the same. Melbourne has a melanoma incidence of 26 cases per 10^5 individuals, Sydney has a rate of 27.3 cases per 10^5 individuals and Brisbane has a rate of 44.6 cases per 10^5 individuals. Melanoma rates increase as the latitude becomes smaller, with the cities having locations of 37.5^{o}S, 33.6^{o}S and 27.3^{o}S respectively. This trend can be seen across the world, not just in the Australian population [18]. Figure 2 shows a graphical representation of the UVR gradient in Australia.

The second pathway through which familial melanoma can occur is the presence of high-penetrance mutations in critical genes such as *CDKN2A* and *CDK4* [19]. Mutation of these genes results in a Mendelian-like pattern of melanoma development which co-segregates with disease in a manner consistent with autosomal dominant inheritance. These high-penetrance genes are often associated with high-density CMM families that present with early-onset melanoma, multiple primary melanoma and multiple cancer types.

The third way that familial melanoma can develop is due to the presence of multiple medium or low penetrance genes that, when combined with high UVR exposure, increase a person's susceptibility relative to the general population. The effect of this has been well described in prostate cancer, where the association between multi-locus genetic risk profiles and increased risk of cancer development compared to the general population has been shown [20]. Examples of medium risk genes for CMM are *BAP1* and *MITF* (Chapter 1.4.2) [21, 22] while examples of lower risk genes tend to be those responsible for pigmentation and naevus development (Chapter 1.4.3).

1.4 Cutaneous melanoma predisposition genes

A proportion of CMM in families and in the general population can be attributed to known predisposition genes. These genes can be divided into three categories: high-risk, medium-risk and low-risk. High-risk alleles are rare but confer an extremely high chance of CMM development in carriers and cause Mendelian-like disease in families (Figure 3). Historically, these have been identified through candidate gene sequencing and linkage studies followed by positional cloning. Low-risk alleles are found much more frequently in the general population but the effects they confer are less profound. Low-risk alleles in CMM have been found through genome-wide association studies (GWAS). Medium-risk alleles occur at relatively low frequency and confer a more intermediate effect.

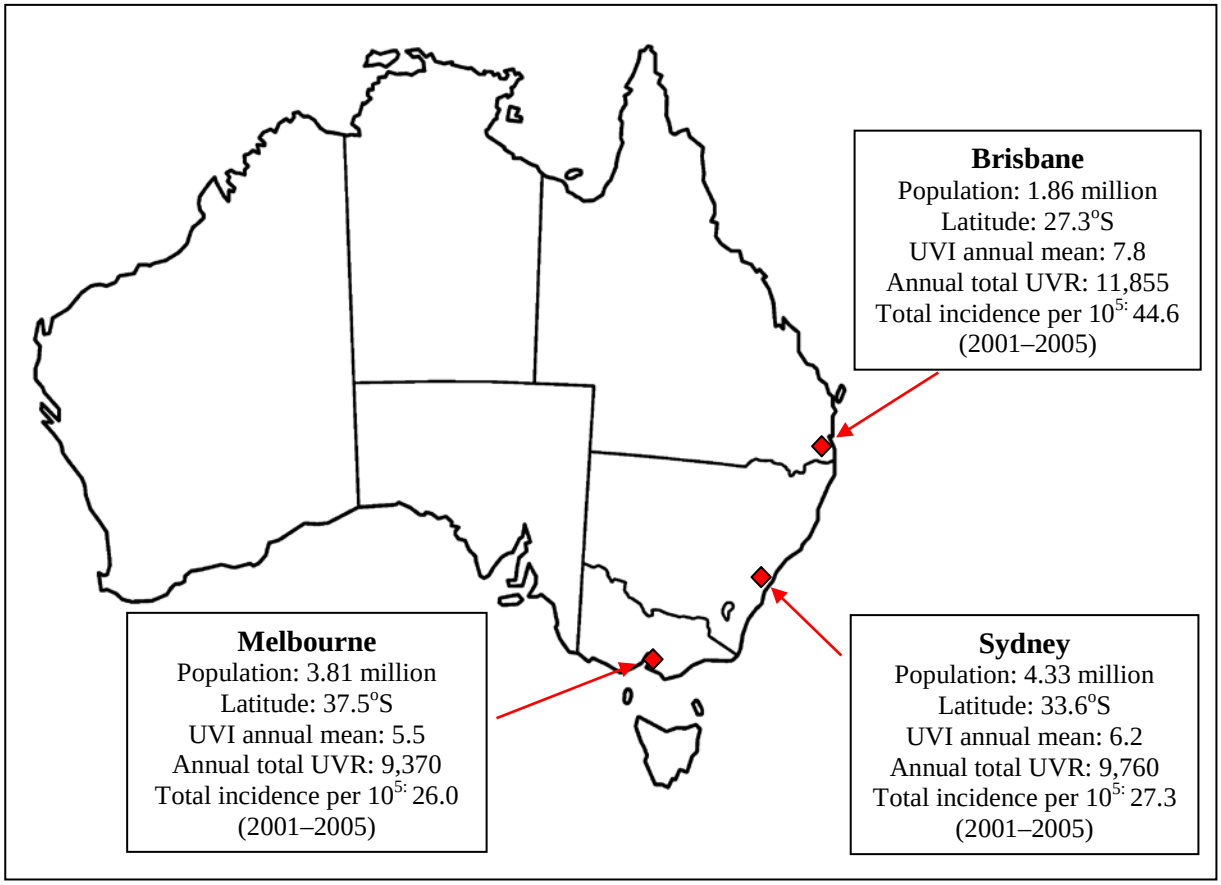


Figure 2: UVR incidence and melanoma rates in Australia. Map of Australia showing the location of major cities Brisbane, Sydney, and Melbourne, their population, latitude, UVR estimates, and standardised melanoma incidence rates. UVI (ultraviolet index) is a measure of UVR intensity. Adapted from Cust et al., 2009.

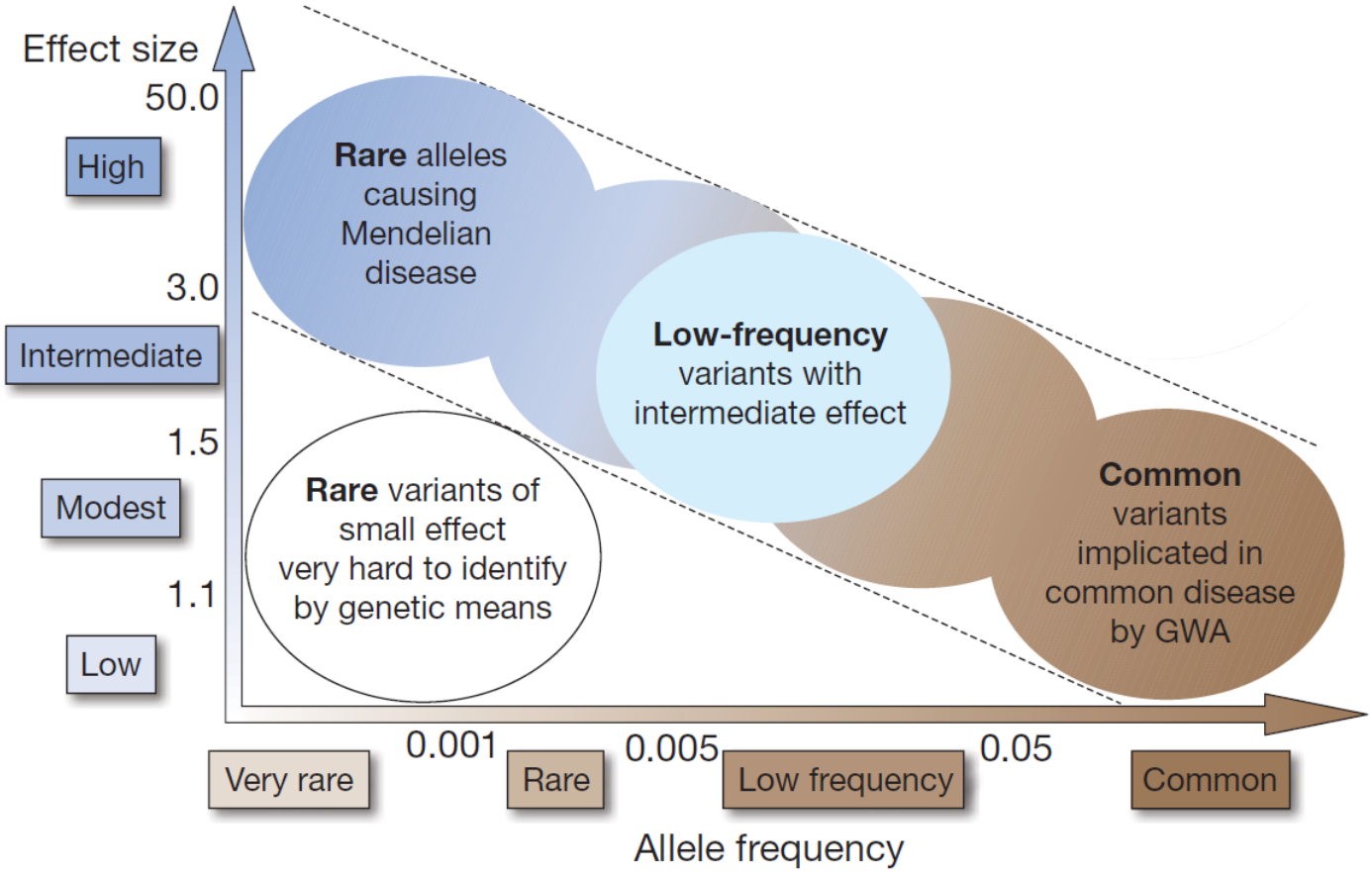


Figure 3: Gradient of effect of predisposition genes. GWA is genome-wide association. Adapted from Manolio et al., 2009.

1.4.1 High-risk genes: *CDKN2A* and *CDK4*

The major melanoma predisposition gene is Cyclin-Dependent Kinase Inhibitor 2A (*CDKN2A*) which is located at chromosome band 9p21 [8, 23-25]. It is a tumour suppressor that inhibits the progression of potential cancer cells by inducing apoptosis or senescence [26]. It is involved in two of the most important tumour suppressor pathways, the p53 and retinoblastoma (RB) pathways, and encodes two different proteins: p16INK4a and p14ARF (Figure 4) [27].

P16INK4a (p16) acts in the RB pathway. When DNA damage occurs, p16 binds to cyclin-dependent kinase 4 and cyclin-dependent kinase 6 (CDK4 and CDK6), inhibiting their activity. This in turn inhibits the phosphorylation of pRB, allowing it to sequester E2F, which is responsible for transcription of genes that allow cells to move from G1 to S phase in the cell cycle. Activating this tumour suppressor pathway leads to senescence and regulation of damaged cells [26]. The second protein encoded by *CDKN2A*, P14ARF (p14), acts in the p53 pathway. In response to DNA damage, p14 raises levels of p53 by interacting with the p53 binding protein, MDM2. This allows levels of p53 to become elevated and in turn inhibits cell proliferation by transcription of several genes capable of inducing apoptosis. P14 sequesters MDM2 in the nucleus thus abrogating its ability to ubiquitinate p53 and target it for proteasomal degradation [30]. Without this process damaged cells may continue progressing towards malignancy.

The contribution of *CDKN2A* to melanoma predisposition was discovered in 1994 through a positional cloning approach following previous linkage studies [8, 23]. Approximately 40% of high-density melanoma families carry inactivating germline mutations in *CDKN2A*. The mutations reported in p16 (predominantly missense mutations) occur across the entire length of the protein [8, 23]. Mutations arising in p14 consist of whole-gene deletions [28], insertions [29], splice mutations [30], and base deletions [31]. No missense variants are reported in conjunction with CMM [24]. In rare instances, a germline *CDKN2A* promoter mutation, which causes an aberrant initiation codon, is shown to predispose to melanoma [32, 33]. An intronic mutation in *CDKN2A* has also been shown to be associated with CMM in an English population [34].

The second melanoma predisposition gene, *CDK4*, was discovered two years after *CDKN2A* through a candidate gene screening approach [6, 7]. It is an oncogene with two activating mutations that predispose to CMM. They result in amino acid substitutions at the same residue encoded in exon 2, p.R24C and p.R24H. This is the position at which p16 binds to CDK4, normally inhibiting the ability of CDK4 to negatively regulate the RB pathway [6, 35].

The largest study to date looking into high-risk melanoma susceptibility genes in high-density melanoma families was conducted by the International Melanoma Genetics Consortium (GenoMEL), a collaboration of major research groups from across four continents. The study found that 41% of familial melanoma was caused by germline variants in *CDKN2A* (encoding either p16INK4a or p14ARF) and *CDK4*. Specifically, they found 178 out of 466 families had a germline p16 mutation, 7 out of 466 families had a p14 mutation and 5 out of 466 families had a *CDK4* mutation. For a complete summary of mutation by geographic region, refer to Table 1. This study highlighted the fact that many different mutations within *CDKN2A* confer an elevated melanoma risk, as they were able to capture 57 different mutations within the 178 p16-positive families. This research also found mutations in *CDKN2A* could predispose to pancreatic cancer, as 49 out of 66 families that presented with both CMM and pancreatic cancer had a mutation in this gene [19]. Collectively these two genes account for approximately 41% of the risk in high-density families. There is a great need to identify the risk alleles in the remainder of these families in order to better understand their disease and provide better clinical outcomes.

To date, there have only been two published studies determining the population-based frequency of *CDKN2A* mutations in CMM cases. The Genes Environment and Melanoma (GEM) study looked at probands from nine different geographical regions in the USA, Canada, Italy and Australia. They discovered 65 *CDKN2A* mutation carriers in a sample of 3,550 affected individuals [36]. This equates to a population frequency of approximately 2%. They also calculated the risk of developing melanoma in the *CDKN2A* mutation carriers they had identified, concluding that carriers had a 14% risk of melanoma by the age of 50, 23% risk by the age of 70 and 28% risk by the age of 80 [36]. Aitken and co-workers also investigated the association between mutations in *CDKN2A* and melanoma in a population-based sample that was derived from families in Queensland. This study also assessed whether there is a link between polymorphisms in *CDKN2A* and melanoma [37]. They used a sample of 482 families that had been categorised into high, intermediate and low risk based on the density of cases and history of disease of each family. They found that *CDKN2A* mutations occurred only in high-risk individuals in this group, with 9 out of 87 probands harbouring a mutation. The occurrence of a Nt500G polymorphism was also found to increase linearly with family risk and was found most often in the families that harboured a *CDKN2A* mutation (though this may be reflective of Celtic ancestry). This study extrapolated this data into a cohort of 7,784 cases that had survey data available, of which 151 fitted the high risk profile. By hypothesising that *CDKN2A* mutations will occur at a similar frequency in the survey group, it was determined that *CDKN2A* mutations occur in 0.2% of melanoma cases in the population-based Queensland sample [37]. This calculation was based on a very small number of individuals screened and thus a broader,

Table 1: Incidence of high-penetrance mutations in *CDKN2A* and *CDK4* in GenoMEL families. Details regarding number of families, melanoma cases, *CDKN2A* and *CDK4* mutations, and large deletions are detailed by country. Genomic regions not tested are abbreviated by 'NT'. This is an abridged table taken from the GenoMEL research paper previously described [19].

city, country	families	total CMM cases	confirmed CMM cases	p16	p14	CDK4	deletions (p16 or p14)	reference
Brisbane, Australia	102	512	450	18	0	1	0	37, 163, 164
Sydney Australia	62	258	258	14	0	0	0	32, 165, 166
Tel Aviv, Israel	1	5	5	0	0	0	NT	
Paris, France	41	146	138	19	0	1	0	7
Genoa, Italy	13	55	35	9	0	0	0	167
Emilia-Romagna, Italy	11	36	33	1	0	0	0	168
Leiden, Netherlands	27	159	159	19	1	0	NT	10, 169
Barcelona, Spain	12	41	40	5	0	0	0	29, 170
Stockholm, Sweden	10	37	36	4	0	0	NT	171-173
Lund, Sweden	15	60	60	8	NT	0	NT	174, 175
Glasgow, Scotland	10	31	31	6	NT	NT	NT	176, 177
Leeds, UK	75	293	169	35	4	1	3	32, 178
Boston, USA	6	23	13	1	0	0	NT	179
Washington, USA	42	219	219	19	2	2	1	8, 180
Philadelphia, USA	9	33	26	2	0	0	NT	
Utah, USA	11	150	150	5	0	NT	NT	181
Toronto, Canada	19	79	45	13	0	0	NT	
TOTAL	**466**	**2137**	**1867**	**178**	**7**	**5**	**4**	
mutation frequency				**0.38**	**0.03**	**0.02**	**0.02**	

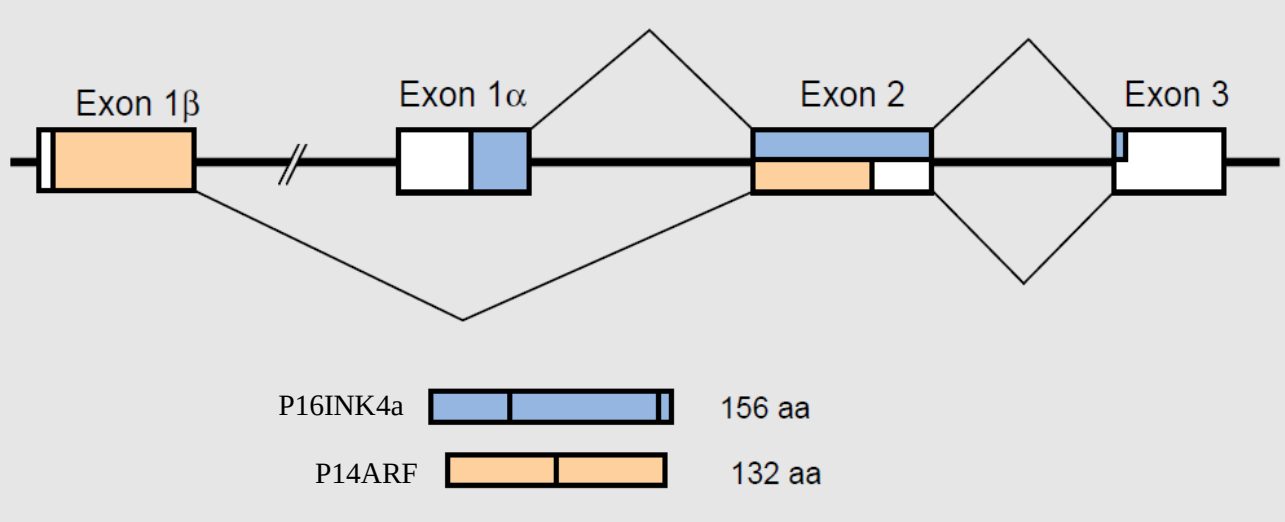


Figure 4: Alternative transcripts and products of *CDKN2A*. The exons that encode p16INK4a are shown in blue. The exons that encode p14ARF are shown in orange. Adapted from Ruas and Peters, 1998.

more comprehensive survey is required to fully estimate the impact of *CDKN2A* mutations in a population-based Australian sample. This forms the basis of the first aim of this thesis.

1.4.2 Medium-risk genes: *MC1R, MITF* and *BRCA2*

Pigmentation traits are one of the phenotypic characteristics associated with melanoma risk [38]. Melanocortin-1 receptor (*MC1R*) is a key regulator of skin pigmentation and is considered to be a medium-penetrance risk allele for melanoma [39]. It is a highly polymorphic gene, with polymorphisms in the gene responsible for much of the variability in skin phenotypes in humans [40, 41]. Melanin is a pigment that protects the skin from the damage caused by UVR. There are two different forms, eumelanin and pheomelanin, and individuals with protein-coding *MC1R* variants show an increased level of pheomelanin compared to eumelanin [42, 43]. It is this increase that results in the red-hair and freckling phenotype and an increased sensitivity to UVR [44]. Eumelanin has a protective role in the skin and aids to protect melanocytes from UVR induced DNA damage [45]. Studies have shown that carriers of an *MC1R* variant have a relative risk of developing melanoma of between 2.2 and 3.9, and that if an individual carries more than one variant the risk increases again to between 4.1 and 4.8 [46-48]. Furthermore, *MC1R* variants can serve as a modifier for *CDKN2A*, causing an increase in the penetrance of *CDKN2A* mutations from 50% to 84% and also decreasing the age of onset of disease by 20 years [46, 47, 49].

Microphthalmia-associated transcription factor (*MITF*) has also been identified as a medium penetrance gene associated with familial and sporadic melanoma. It is located on chromosome 3 and, of the nine gene isoforms, M-MITF is the melanocyte-specific isoform. It is expressed in melanocytes and also in melanoma cells where its function includes regulating the transcription of genes involved in cell survival, differentiation and melanogenesis [50]. Mutations in *MITF* are responsible for the development of Waardenburg Syndrome type 2, which is characterised by pigment abnormalities that involve the skin, hair and eyes, hearing loss, as well as minor defects in the structures that develop from the neural crest [51].

Several studies have shown *MITF* to be somatically dysregulated in melanoma. Cronin et al. found *MITF* to be mutated in 4 out of 50 melanoma cell lines derived from metastatic tumours and amplified in a further 4 out of 50 cell lines. With an additional 3 mutations in SRY (sex determining region Y)-box 10 (*SOX10*), an upstream transcription factor, it was shown that mutations in this pathway existed in over 20% of metastatic melanoma cell lines [52]. Further to this, screening of primary tumours found 2 out of 26 had mutations in *MITF* and 6 out of 55 had mutations in *SOX10*,

so that, overall, 14% of the primary tumours were affected in this pathway. In a separate study by Garraway et al., *MITF* was amplified in 3 out of 30 primary CMM, 7 out of 32 metastatic tumours and 0 out of 10 benign nevi. This is suggestive of *MITF* being a lineage-specific oncogene and emphasises the importance of *MITF* in melanomagenesis [53].

The focus on *MITF* in melanocyte biology and somatic melanoma development led to the discovery of a germline variant associated with melanoma susceptibility in 2011. Two groups reported p.E318K (NM_000248) in families with CMM. In one study, which used whole-genome and exome sequencing of melanoma families, the variant was found to co-segregate with melanoma in some, but not all, cases in these families, consistent with the notion that this is a medium-penetrance variant. They also used case-control studies to show that this is a population variant that confers a slightly greater than two-fold risk of developing CMM (OR 2.7). Interestingly this risk is similar to that attributed to *MC1R* alleles associated with red hair [54]. Other associations were made with family history of disease, multiple primary melanoma, increased naevus count and non-blue eye colour [21]. The second group also described the germline variant increasing the risk of CMM development in a familial setting and showed that it also confers an elevated risk of developing renal cell carcinoma [55].

The role of *MITF* has also been described in relation to tumour properties. Experiments by Hoek et al. show cells *in vivo* that exhibit low levels of MITF expression have an invasive tumour phenotype, while cells exhibiting high levels of MITF expression have a proliferative phenotype [56]. The ability of *MITF* to control the switch between a motile and proliferative state makes it an interesting candidate for melanomagenesis. A candidate gene approach interrogating its role in CMM predisposition is described in Chapter 3.

An essential function required for genome stability is the capacity to repair double-stranded breaks in DNA. Breast cancer 2, early onset (*BRCA2*) is a tumour suppressor gene located on chromosome 13 that is responsible for maintaining genome stability. It does so through its involvement in the homologous recombination pathway that facilitates double stranded DNA repair by binding to recombination enzyme RAD51 [57-60]. Germline mutations in *BRCA2* confer an increased risk of developing breast and ovarian cancer, with approximately one-third of familial breast cancer cases attributed to mutations in this gene [61, 62]. It is estimated that germline *BRCA2* mutations confer a relative risk of melanoma development of 2.6 [62, 63].

1.4.3 Low-risk genes: pigmentation and naevus variants

Melanoma risk can be attributed to phenotypic characteristics relating to pigmentation and naevus variants. These characteristics include fair skin and hair colour, freckling, poor tanning response and high naevus counts [64]. GWAS have been used to determine a large number of genes that can be attributed to these phenotypes. To date, 16 low-penetrance variants have been discovered using this method (Table 2). Genes associated with pigmentation are *SLC45A2*, *TYR*, *HERC2/OCA2*, *MC1R* and *ASIP* [65-69], while the gene *PLAG2G6* is strongly associated with naevus development [70]. GWAS have also linked a number of SNPs to an increased risk of CMM development. These SNPs are located in or near the following genes: *PARP1*, *CASP8*, *TERT*/*CLPTMIL*, *CDKN2A*/*MTAP*, *ATM*, *CCND1*, *MC1R*, *MX2* and a region on locus 1q21.3 [67, 69, 71, 72]. A study by Gudbjartsson et al. used a candidate gene approach combined with GWAS to identify *ASIP* and *TYR* as being associated with increased risk of both CMM and basal cell carcinoma [73]. Many of these GWAS findings are associated with low-penetrance risk and are linked to hair colour, eye colour and naevus count, which is consistent with observations that have been made through epidemiological studies [74]. They also point to some new regions not associated with these traits (e.g. *ATM*, *CCND1*, *MX2*, and *SETDB1*) for consideration when trying to uncover the key genetic events occurring during melanomagenesis [66, 67, 71].

Table 2: Low-risk variants found through genome-wide association studies. SNPs listed have been found to be associated with CMM and/or pigment and naevus traits.

Gene	SNP	Odds ratio	Pigment	Naevus	Melanoma	Reference
PARP1	rs3219090	0.89	No	No	Yes	[71]
1q21.3 Multiple genes	rs7412746	0.9	No	No	Yes	[71]
CASP8	rs13016963	1.14 (1.09, 1.19)	No	Weak	Yes	[67]
SLC45A2	rs35390	0.36 (0.23, 0.53)	Yes	No	No	[67]
TERT/CLPTMIL	rs401681	0.87 (0.81, 0.94)	No	Weak	Yes	[72]
CDKN2A/MTAP	rs7023329	0.83 (0.78, 0.88)	No	Weak	Yes	[69]
TYR	rs1393350	1.30 (1.21, 1.39)	Yes	No	No	[69]
ATM	rs1801516	0.84 (0.79, 0.89)	No	No	Yes	[67]
CCND1	rs1485993	1.11 (1.04, 1.18)	No	No	Yes	[67]
HERC2/OCA2	rs1129038 and rs12913832	0.97	Yes	No	No	[66]
MC1R	rs258322	1.70 (1.54, 1.87)	Yes	No	Yes	[69]
ASIP	rs1015362 and rs4911414	~1.20 to ~1.29	Yes	No	No	[65]
MX2	rs45430	0.88 (0.85, 0.92)	No	No	Yes	[67]
PLA2G6	rs2284063	0.85 (0.79, 0.91)	No	Yes	No	[70]

1.5 Candidate melanoma susceptibility genes

1.5.1 *PALB2*

PALB2 is a BRCA2 interacting protein that is essential for BRCA2 DNA repair functions [75]. It physically binds to both BRCA1 and BRCA2 and is necessary for proper tumour suppressor activity [76]. Germline mutations in *BRCA2* are known to predispose to cancers of the breast, ovary and pancreas, as well as confer an increased risk of CMM (relative risk of 2.6 [62, 63]), suggesting by association that *PALB2* may also play a role in tumour development in these cancer types. A study published in 2007 showed that germline mutations in this gene are associated with an increased breast cancer risk [77]. Probands from breast cancer families that were wild type for *BRCA1* and *BRCA2* were sequenced and results showed that 10 out of 923 affected individuals harboured a truncating mutation in *PALB2* in contrast to 0 out of 1,084 controls. The variants showed imperfect segregation with disease, with approximately half of the pedigrees showing partial segregation. This implies that this is likely to be a medium to low penetrance gene in regards to breast cancer risk. In one family there was a case that had developed both primary melanoma and breast cancer; this is suggestive that *PALB2* mutations may also infer risk of other cancer types [77]. A population-based study into breast cancer risk associated with germline *PALB2* mutations has been carried out by Tischkowitz and colleagues. They identified five cases with pathogenic mutations out of the 559 cases that were screened and found no mutation in the 565 matched controls. From this they were able to infer a 5.3 fold increase in breast cancer risk in mutation carriers [78].

There have been other associations made between *PALB2* mutation and cancer. Several studies have investigated germline variants in regard to increased pancreatic cancer risk. Jones et al. looked at probands from 96 families with pancreatic cancer and identified truncating nonsense mutations in three of the families. In the 1,084 controls that they screened, no *PALB2* mutations were seen [79]. A more recent study, looking at germline mutation in families that have both breast and pancreatic cancer, found no mutations in the 77 families that were screened. This included 22 probands that had both cancer types [79, 80].

Since individuals with mutations in *BRCA2* have an increased risk of developing several cancer types (breast, ovary, pancreas, prostate, melanoma) it is feasible that a mutation in *PALB2*, a gene that physically and functionally interacts with *BRCA2*, could lead to an increased risk of developing each of these other cancer types also [76], and this has indeed been shown for breast cancer and

pancreatic cancer. The first study to examine the association between *PALB2* mutations and melanoma screened 53 families with multiple cases of melanoma that were *CDKN2A* wild type. No mutations were found in any of the 13 exons or exon boundaries of *PALB2*. This led to the conclusion that *PALB2* may not be associated with melanoma, or that alternatively it is a low frequency risk gene, with mutations being rare events that were not detected in this small sample [81]. Thus this gene remains a good candidate for melanoma susceptibility since the published reports have not been able to unambiguously conclude a role for *PALB2* in melanoma predisposition.

1.5.2 *SOX10*

SOX10 encodes a transcription factor that is expressed in tissues derived from the neural crest [82]. It plays a key biological role in melanocyte development and function, which includes the transcriptional control of MITF, the master regulator of pigment cells [83]. This interaction makes *SOX10* a good candidate melanoma risk gene. Researchers have addressed the interaction between this pathway and disease progression. Referring to a paper previously discussed, there has been a link showing that *MITF* and *SOX10* are somatically mutated in a mutually exclusive pattern in metastatic melanoma. The study by Cronin and colleagues identified frameshift or nonsense mutations in 3 out of 50 tumour cell lines from metastatic melanoma. Furthermore, 6 out of 55 primary tumours also harboured mutations in this gene [52]. This makes *SOX10* a candidate for further study as there is an established link to melanocyte function and somatic development of melanoma.

1.6 Somatic mutations in cutaneous malignant melanoma

A spectrum of somatic mutational events occurs in CMM. Mutations in v-raf murine sarcoma viral oncogene homolog B (*BRAF*) are the most common found in primary melanomas [84]. This gene encodes a kinase that acts in the mitogen-activated protein kinase pathway (MAPK) and regulates a variety of processes that include growth, survival and migration of cells [85]. Cells harbouring a *BRAF* mutation have an activation of kinase activity that results in continuous growth signals [86]. Somatic mutations in *BRAF* occur in up to 70% of primary melanomas, with a single point mutation, p.V600E, accounting for approximately 90% of these [87-90]. Melanomas lacking somatically mutated *BRAF* often carry mutations in another component of the MAPK pathway,

NRAS [91, 92]. Between 15% and 30% of primary melanomas lacking *BRAF* mutations have an *NRAS* mutation [87, 93]. *PTEN* is also frequently mutated in melanoma. It is a tumour suppressor that prevents uncontrolled cell growth through regulation of cell migration and apoptosis [94]. The types of somatic mutations acquired by tumours vary according to histological subtype, with *BRAF* and *NRAS* mutations usually occurring in SSM on intermittently sun-exposed areas of the skin. Somatic mutations in the tyrosine receptor, *KIT*, occur predominantly in ALM and mucosal melanoma (Chapter 1.2), with these subtypes often having wild type *BRAF* [95].

1.7 Uveal malignant melanoma

Uveal malignant melanoma (UMM) occurs on the uvea and develops from melanocytes that are located in the eye. There are three types of UMM: choroid, ciliary body and iris [96]. Unlike cutaneous melanoma, UMM is very rare, with the United States reporting an incidence of 5.6 cases per million per year in the years 1973 to 2008 [17]. The number of newly reported cases of UMM over this time frame has remained at a relatively steady rate, suggesting that the influence of UVR on UMM seems smaller than the association that is observed with CMM [96]. Melanoma can also occur on the conjunctiva of the eye and this is the exception to this pattern. Its incidence does follow the UVR gradient and hence it is better described as a cutaneous melanoma as it occurs on the skin, though it is within the structure of the eye [97].

A 2004 study developed an identification system for UMM that clusters primary UMM tumours into two distinct molecular groups based on their gene expression profiles. Tumours that group into the class 1 profile rarely metastasise and are regarded as low-grade primaries, while tumours that group into the class 2 profile are highly metastatic and are regarded as high-grade primaries. This gene expression profile shows that class 2 tumours have clusters of up-regulated genes on chromosome 8q and down-regulated genes on chromosome 3, a region that has been previously implicated in UMM [22, 98]. Class 2 tumours have also been associated with monosomy 3, suggesting that they may have lost a tumour suppressor gene on chromosome 3 and that a mutation on the remaining allele may lead to the development of disease [99].

To date, there have been only a few genes implicated in the development of UMM; this suggests that there are still undiscovered genes that might be involved in its genesis. There is some evidence of familial clustering of UMM and a study by Jönsson et al. found linkage to 9p21-q22 in three Danish UMM families, however the gene responsible for the linkage peak has yet to be determined [100]. Interestingly, *CDKN2A* has been implicated in only a single UMM family. This family also

presented with CMM. No UMM families have mutations in *CDK4*. This indicates that the major predisposition genes in CMM do not play a significant role in familial UMM [19].

1.8 Somatic mutations in uveal melanoma: *GNAQ, GNA11* and *BAP1*

A novel recurrent somatic mutation in *GNAQ*, located at chromosome 9p21, encoding an alpha subunit of hetrotrimeric G proteins, was found in UMM. The mutation occurred in a panel of UMM tumours and blue nevi and were all acquired somatically at a hotspot located at p.Q209 [101]. Subsequent to this study, *GNA11*, a paralogue of *GNAQ* located on chromosome 19p13.3, was sequenced and also found to be mutated in UMM [102]. This later screening revealed that mutations in these two genes were mutually exclusive. Overall, *GNAQ* p.Q209 mutations were found in 55% of blue nevi, 45% of primary UMM and 22% of metastatic tumours. In *GNA11*, mutations at p.Q209 were found in 7% of blue nevi, 32% of primary UMM and 57% of metastatic tumours. A second hotspot was also discovered at position p.R183 in both genes. Combing the data from these investigations showed that 83% of UMM had a somatic mutation in either *GNAQ* or *GNA11*. This work indicates that this pathway plays a major role in the somatic development of UMM and further to this, the presence of these mutations in benign uveal naevi seems to implicate *GNAQ* and *GNA11* in early mutational events that occur in UMM development [96]. It is of note that although *GNAQ* and *GNA11* have been shown to be somatically mutated in UMM they have so far not been associated with predisposition to UMM.

BRCA1 associated protein-1 (*BAP1*) is a tumour suppressor gene located on chromosome 3. A study by Harbour and colleagues demonstrated that 84% of metastasising class 2 UMM had an inactivating somatic mutation in *BAP1*. These mutations were found through a combination of whole exome and Sanger sequencing. Initially, two class 2 tumours that had lost a copy of chromosome 3 were selected for exome sequencing along with their matched normal DNA samples. It was discovered that both tumours had a mutation in *BAP1* (p.W196X and p.Q322fsX100) that was not found in the matched normal sample, indicating they were somatically acquired. To more fully investigate these mutations, a panel of 26 class 1 tumours and 29 class 2 tumours were then genotyped by Sanger sequencing. A single *BAP1* mutation was identified in the class 1 group while in contrast 26 out of 31 metastatic UMM in the high-risk category had a mutation. Fifteen of these variants resulted in the termination of the protein and 6 affected the UCH (ubiquitin carboxy-terminal hydrolase) domains [103].

1.9 Uveal malignant melanoma predisposition gene, *BAP1*

The original study by Harbour and colleagues also identified a single germline *BAP1* mutation that was a frameshift mutation and occurred in a patient with a class 2 UMM tumour [103]. This opened up the possibility that *BAP1* mutations could confer predisposition to UMM. Since this research was completed, there have been a small number of studies published that focus on *BAP1* germline mutation and its association with cancer risk. One of these studies looked at two American mesothelioma families, from Wisconsin and Louisiana, and showed association between germline changes in *BAP1* and mesothelioma in individuals that had no known exposure to asbestos or erionite, the strongest environmental risk factors [104]. Both of these families have developed multiple cancer types, though mesothelioma is the dominant disease. The Wisconsin family had confirmed cases of ovarian, breast and renal cancer, while the Louisiana family harboured two cases of prostate cancer, a pancreatic cancer, two cases of CMM and a case of UMM. By overlaying data from array-comparative genomic hybridization analysis and linkage analysis the researchers uncovered changes that encompassed the *BAP1* locus in two tumour samples, which led to a more specific interrogation of *BAP1* through Sanger sequencing. In the Wisconsin family, a splice variant was found to be responsible for disease as it co-segregated in four mesothelioma cases and two ovarian cases. In the Louisiana family, a truncating mutation, p.Q684X, was found to co-segregate with disease in three individuals with mesothelioma, two of whom also presented with CMM, with the third individual having UMM [104]. This work proved that *BAP1* is associated with predisposition to several cancer types, including UMM. A third study into *BAP1* mutation further supported these findings by interrogating the germline sequence of 53 UMM cases, five of which had a family history of melanoma. They identified three germline mutations in three probands: two variants were non-pathogenic, a synonymous variant c.1026C>T and intronic variant c.123-48T>G, but the third was a truncating mutation, p.Q267X, in a family with mesothelioma, meningioma and UMM [105]. This mutation segregates with disease in individuals from whom DNA is available and is present only in one individual that is cancer free. Genotyping of three tumours from this family also revealed loss of heterozygosity (LOH) on the short arm of chromosome 3 for all samples and in each case a loss of the wild type allele in the tumour. A study of *BAP1* in malignant pleural mesothelioma found inactivating somatic mutations in 23% of samples tested [106]. By interrogating recurrent genomic copy number alterations, 25 candidate genes were identified. *BAP1* was selected from this group as a possible driver mutation as it had the highest rate of non-synonymous variation. This work supports the finding of germline predisposition to mesothelioma found by other researchers.

Another study to document germline *BAP1* mutations shows its alteration in two families presenting with an autosomal dominant syndrome characterised by a variety of melanocytic tumours, including naevi, CMM and also UMM [107]. Both families had a case with UMM and in one family there were also three CMM cases. Segregation analysis has shown that germline BAP1 mutation in these families predisposes to these tumours and co-segregates with the disease phenotype. A frameshift mutation occurred in both families, the first located at p.Q436NfsX135 and the other at p.M687EfsX28. Once again array-comparative genomic hybridization has shown that there are losses affecting chromosome 3 that effectively leaves only a mutated copy of *BAP1* in the tumours. Collectively, these studies have validated *BAP1* as a tumour suppressor, which when mutated can lead to the development of several cancer types, including CMM, UMM and mesothelioma [22]. It is of note that all germline mutations described in the literature are truncating. Refer to Table 3 for a summary of known deleterious *BAP1* germline variants published up to the date this thesis was started. Further literature is outlined in Chapters 6 and 7.

Table 3: Published deleterious germline variants in *BAP1* shown to predispose to cancer.
Information on protein change, cancer type and a reference is given.

Protein change	Cancer type	Reference
p.E402fsX32	UMM	[103]
c.IVS7-2A<G p.I72fsX7	mesothelioma	[104]
p.Q436NfsX135 p. M687EfsX28	melanocytic tumours	[107]
p.Q687X	CMM, UMM, meningioma, mesothelioma	[105]
p.R60X, p.D236fs, p.R385X, p.E611fs, p.A634fs, p.K659X	CMM, UMM	[108]
p.Q684X p.P147fsX48	CMM, UMM, mesothelioma, melanocytic tumours	[109]
p.K25fsX43	UMM	[110]
p.Y241X	CMM, mesothelioma	[111]

1.10 Next-generation sequencing in cancer gene discovery

Whole-genome sequencing captures mutations occurring in the coding and non-coding fractions of the genome that includes promoter regions, exons, introns and intergenic regions. It gives researchers the ability to detect single point mutations as well as insertion/deletions and other more complex chromosomal aberrations. In the broader sense it can be used to detect structural rearrangements including chromosomal translocations and copy number changes (Figure 5) [112]. There has been a lot of success using this method of gene discovery in both the somatic and germline disease settings as researchers are now able to characterise the genetic landscape associated with different disease phenotypes. In the somatic application, a tumour sample is matched to the patient's normal DNA to identify mutational events that have occurred in the tumour. This method of discovery has been a useful tool in many cancer research projects, including studies in acute myeloid leukaemia, prostate cancer and lung cancer [113-115], among a rapidly increasing number of cancer types. In the familial setting, a comparison is made between affected family members to identify common germline mutations that may underlie disease predisposition.

An alternative to whole-genome sequencing is to focus on particular chromosomal regions of interest. This provides cost-advantages as well as smaller volumes of data, which are easier to manipulate at the analysis stage. Exome sequencing is an effective way to screen the entire protein coding fraction of the genome and it typically covers a 50 to 62 Mb region. The limitation of exome sequencing is the inability to detect structural changes across the chromosomes; however, the lower costs and high-sensitivity in detecting mutational events make it a powerful tool for discovering gene variants associated with disease [116]. Exome sequencing is the main gene discovery technology used in this thesis, the outcomes of which are described in Chapters 4, 5 and 7.

Sequencing using Ion Torrent Personal Genome Machine (PGM; Life Technologies, USA) is an even more targeted method of mutation discovery. This technology allows sequencing of multiple gene regions in a high-throughput and cost-effective manner. Targeted sequencing covers up to 1 Mb per pool and allows the user to multiplex up to 64 samples in a single run. Chapter 2 of this thesis outlines the methods and outcomes of the population-based screen of CMM cases from Queensland, performed using the PGM.

As a whole, next-generation sequencing (NGS) represents a rapid and comprehensive method for identifying predisposition genes in order to identify genetic alterations underlying disease development in both the somatic and the familial setting.

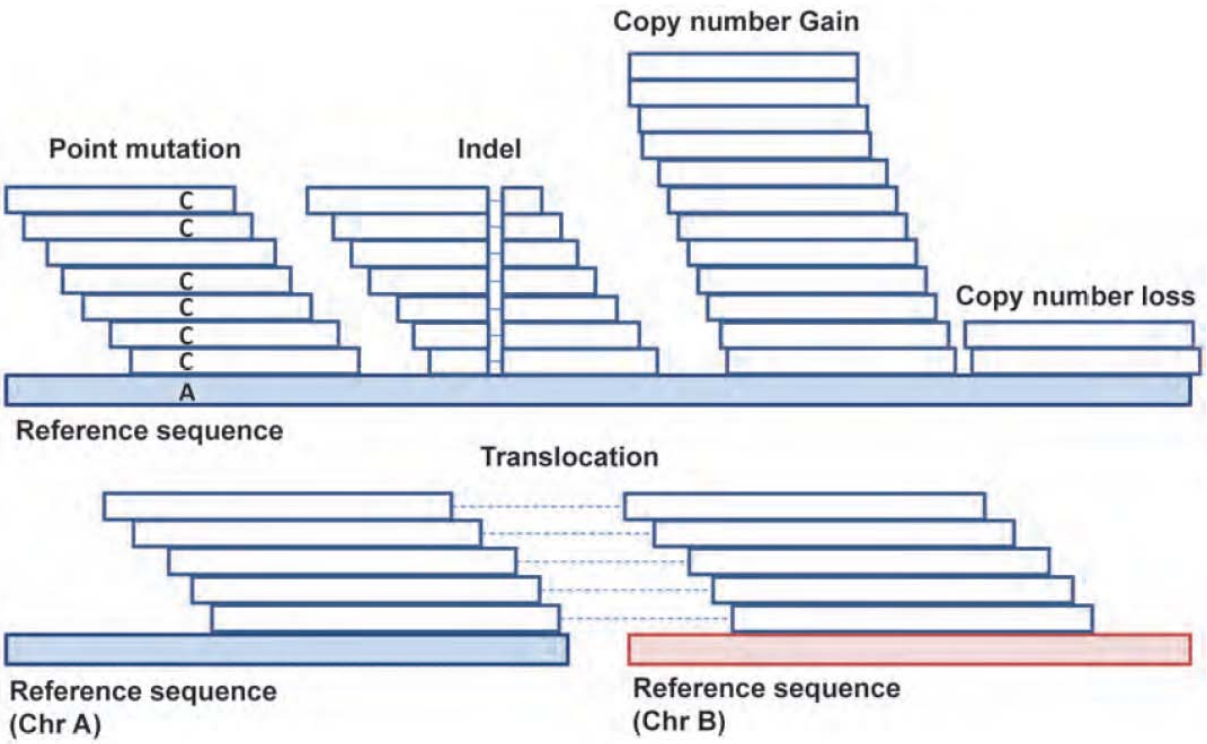


Figure 5: Genetic alterations detected by whole-genome sequencing. Sequence reads are indicated by bars. Point mutations, insertion/deletions and copy number alterations are depicted. Paired-end reads (joined by dotted lines) show identification of translocations. Adapted from Yoshida et al., 2013.

1.10.1 Whole-genome and exome sequencing in melanoma tumours

The first melanoma whole-genome sequence was completed in 2010, on a commonly used melanoma cell line COLO-829 and its matched lymphoblastoid cell line (LCL). This was the first opportunity to catalogue the full complement of somatic mutational events and structural changes occurring in CMM and resulted in the identification of several novel potential melanoma driver genes [117]. This technique was able to identify 88% of previously known mutations in this sample and also uncovered a total of 33,345 somatic base substitutions, with a subset showing a 97% confirmation rate through Sanger sequencing. Analysis of the complete mutation list showed that only 292 variants occurred in a protein-coding region and when this list was further broken down, only 187 caused a non-synonymous change. One of the key features of melanoma biology highlighted by the study is the identification of a mutation signature indicative of DNA damage acquired through UVR exposure. Most of the base substitutions were C>T or G>A. In fact, 92% of mutations occurred with this patterning compared to the 53% that would be expected by chance. Another feature of this study is the demonstration of preferential DNA repair on the transcribed DNA strand as opposed to the non-transcribed strand. There were 2,773 C>T changes on the transcribed strand in contrast to 4,058 on the non-transcribed strand.

The first exome sequencing study in the melanoma field was carried out by Wei et al. The group sequenced DNA from 14 metastatic tumours and the corresponding matched normal DNA samples. They discovered a mutation in the transformation/transcription domain-associated protein, *TRRAP* that occurred in a hotspot (p.S722F) in two samples. When this was expanded into a larger case set, they found this specific mutation in 6 out of 167 samples. This study also found the gene *GRIN2A* (glutamate receptor, ionotropic, N-methyl D-aspartate 2A) mutated in around 30% of melanoma samples. This came out of the initial exome sequencing that identified a mutation event in *GRIN2A* in 6 out of 14 metastatic melanoma exomes [118]. An important finding that came out of this study was the discovery of the high rate of mutation in melanoma, which is much higher than that seen in most other cancers (the exception being lung cancer from smokers) [119]. This demonstrates that gene discovery in melanoma is a complex puzzle and that there is likely to be many passenger mutations to sift through in order to find the important driver mutations that cause disease development and progression. Exome sequencing by other groups has also uncovered other somatic mutations of functional importance, such as the role of the MAPK family of genes in melanoma. In a similar approach using melanoma samples and matched normal DNA samples, somatic non-synonymous mutations in *MAP2K1*, *MAP2K2*, *MAP3K5* and *MAP3K9* have been reported in melanoma cell lines and tumours [120, 121].

1.10.2 Whole-genome and exome sequencing in familial melanoma

Linkage studies of high-density melanoma prone families suggest that there is a great deal of genetic heterogeneity associated with melanoma predisposition. A technique to address this issue of heterogeneity is to consider families individually rather than as a group. By assessing the entire genome or exome of affected family members and searching for variants that are common between cases, genes responsible for increased melanoma risk can be uncovered.

A NGS approach in families is the main discovery tool used in this thesis. High-density families were prioritised as they seemed more likely to carry a genetic predisposition rather than cluster by chance. Between 1 and 6 melanoma cases were selected from each pedigree. Individuals that were the furthest apart genetically were selected as they would share fewer novel variants and would therefore narrow down the gene list. Individuals with multiple primary melanomas, early onset disease, or multiple primary cancers (where CMM is one of the cancers) were also selected as they had a greater disease burden and were again likely to have a greater genetic risk. The majority of the NGS was out-sourced to a South-Korean based company, Macrogen. These samples were sequenced using Agilent SureSelect enrichment kits on the Illumina Hiseq2000 platform. Early whole-genome and exome data was generated by Complete Genomics, or by sequencing facilities at the National Cancer Institute (USA) or the Translational Genomics Research Institute (USA). Refer to Table 4 for a complete list of whole-genomes and exomes sequenced and the centres at which this was carried out.

1.10.3 Next-generation sequencing data analysis

Given the large data sets that are generated by whole-genome and exome sequencing, a robust bioinformatics pipeline is essential to generate usable variant calls that can be filtered to obtain a target list of genes for follow up. For the sequencing projects outlined in this thesis the following bioinformatics were applied to the raw sequence data:

1. The sequence output was mapped to the UCSC human genome reference build 19 using the Burrows-Wheeler alignment algorithm [122].
2. SNPs and insertion/deletions were annotated using bcftools and SAMtools mpileup with disabled base alignment quality computation [123]. Each sample had an average of 90,000-100,000 variants compared to the human genome reference sequence.
3. Quality score of reads had to be greater than or equal to 70.
4. Number of reads of the alternate allele had to be greater than two and equate to at least 20% of the total number of reads.
5. Synonymous and non-protein changing variants were excluded.
6. Common SNPs with minor allele frequency greater than 1% found in dbSNP version 135, The 1000 Genomes Project April 2012 release or the NHLBI Grand Opportunity Sequencing Project (ESP6500) were excluded.
7. Common sequencing artefacts were removed along with any region of duplication in the genome.

After the filtering process samples typically had 100–200 novel (or very rare) germline variants when considering the entire exome (Figure 6).

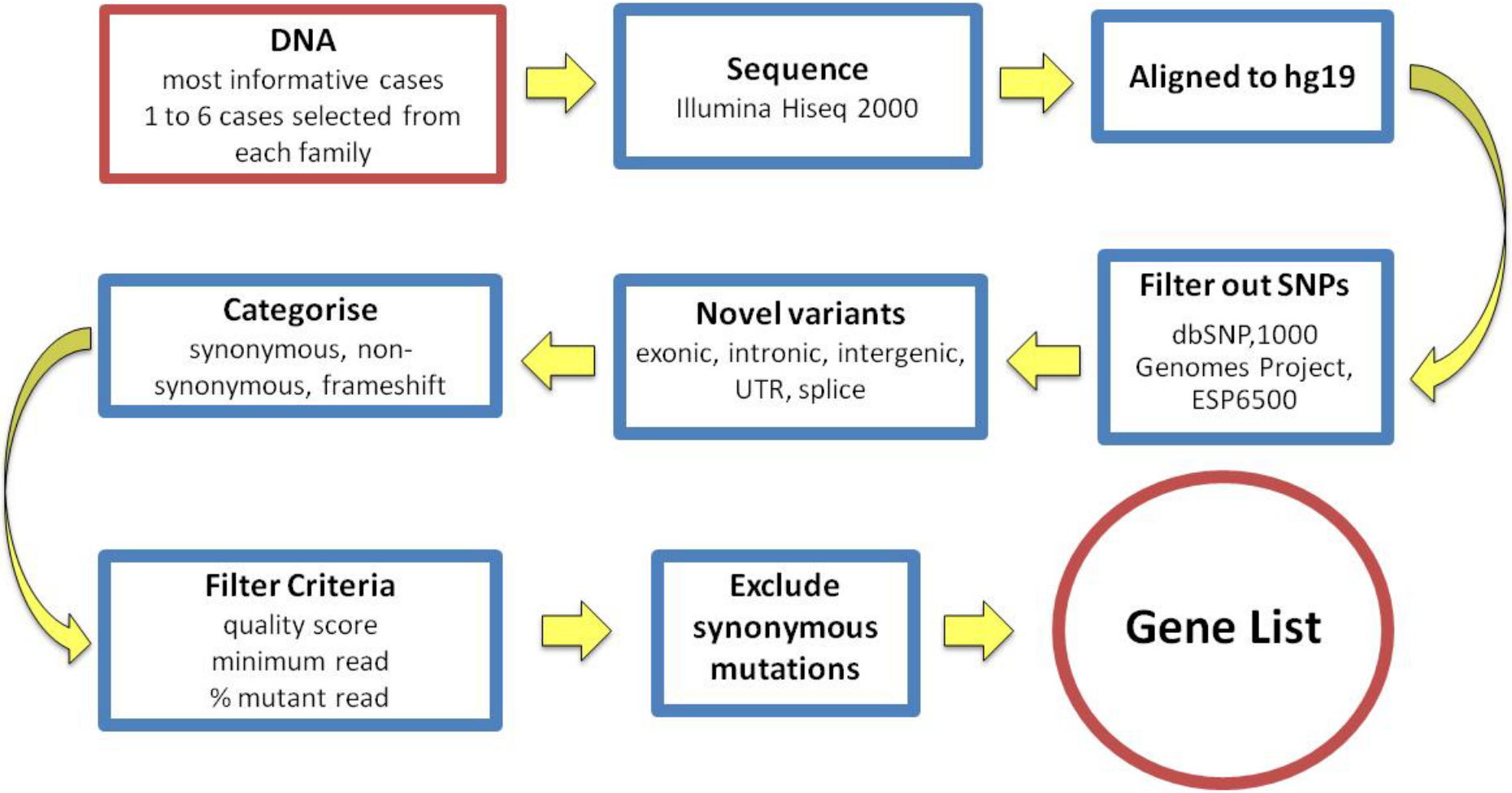


Figure 6: Bioinformatics pipeline for whole-genome and exome sequencing data sets

1.11 QFMP and GenoMEL, resources for melanoma gene discovery

The Queensland Familial Melanoma Project (QFMP) is a family and twin study that originated in Queensland, Australia in 1987. The aim of the study was to better understand the relationship between UVR exposure and individual susceptibility to melanoma [124]. Initially, over 10,000 individuals with CMM were identified through the Queensland Cancer Registry and were asked to participate in this state-wide study. Cases were asked to fill out a questionnaire pertaining to family history, pigmentation, freckling, mole count and likelihood of sunburn. In the instance where an individual had a family history of melanoma, the first degree relatives and affected cases were also ascertained. Overall, 1,897 families took part, 738 of which provided a blood sample for DNA extraction and/or LCL production [125]. A follow-up study in 2002 to 2005 updated data and collected additional blood samples from families. This included low-risk individuals with no family history of CMM, two-case families categorised as intermediate-risk and families with three or more cases categorised as high-risk. The largest family collected in this study has 15 members affected with melanoma. From the total cohort, 15 families presented with both UMM and CMM and 34 families had at least 5 members with CMM; all of these have been genotyped and are wild type for *CDKN2A* and *CDK4*. Along with DNA samples, there is also phenotypic information available on these individuals and disease information derived from the cancer and death registries. There is also linkage data (unpublished) available for the majority of these large families that provides a starting point for prioritising novel variants discovered by NGS.

The International Melanoma Genetics Consortium (GenoMEL) is a valuable resource in melanoma gene discovery. It consists of pooled family samples from across Europe, Asia, North America and Australia and is the largest familial melanoma sample to date [19]. The Australian sample consists of families in the QFMP along with families from the Australian Melanoma Family Study (AMFS) [18], which has been collected in New South Wales, Queensland and Victoria, and includes 9 families with both UMM and CMM cases. The use of these dense melanoma families provides a significant tool for gene discovery and as the price of whole-genome and exome sequencing continues to fall, NGS becomes a feasible option for discovering melanoma risk alleles.

1.12 Objectives

The primary objective of this thesis was to identify and characterise novel (not previously been associated with melanoma predisposition) high-penetrance melanoma predisposition genes by applying NGS techniques and candidate gene sequencing to affected members of cutaneous melanoma-prone families collected as part of the QFMP and through GenoMEL [19, 124]. A total of 231 genomes/exomes from 100 families were sequenced over the course of this study. There were 91 samples from UMM families and 140 samples from CMM families. This formed the core of this project and is the basis for much of the research described in Chapters 4, 5 and 7. For a breakdown of samples sequenced refer to Table 4. Genes selected for the candidate gene sequencing project that used traditional Sanger sequencing techniques are outlined in Chapter 3.

The secondary objective of this thesis was the identification of novel medium or low penetrance variants that may also account for melanoma predisposition in the general population. Case-control analysis was the basis of this study with the cases obtained through the QFMP [124]. In addition to cases collected through the QFMP, we also had access to extensive cohorts of sporadic melanoma patients and age-matched controls from Queensland, in which these variants were further investigated. The control group used in this study were the parents of twins ascertained as part of the Brisbane Twin Naevus Study [126]. The available samples included: 1,550 family probands, 139 sporadic cases and 1,630 controls.

1.13 Hypotheses

I. A small proportion of the Queensland familial melanoma cohort will have mutations in *CDKN2A* and *CDK4*.
II. Some families who do not have germline mutations in *CDKN2A* and *CDK4* will have high penetrance mutations in novel genes.
III. Families who do not have high penetrance mutations will have multiple low or medium penetrance gene variants that in combination contribute to melanoma susceptibility.
IV. Families with a combination of uveal and cutaneous melanoma represent a distinct genetic sub-group of familial melanoma.

1.14 Specific aims of the study

1. To assess the contribution of *CDKN2A, CDK4* and *BAP1* mutations to melanoma susceptibility in a population-based sample of melanoma cases (the QFMP).
2. To use a combination of whole-genome, exome and candidate gene sequencing to identify both novel high and low penetrance melanoma predisposition genes using high density melanoma families.
3. To determine whether variants within candidate genes identified in Aim 2 segregate with melanoma within the 'discovery' families.
4. To assess whether variants within candidate genes identified in Aim 2 segregate in an independent series of families.
5. To assess selected novel candidate genes identified in Aim 2 as possible low penetrance melanoma risk variants in case-control samples.

Table 4: Melanoma family samples with whole-genome or exome sequencing data

Family ID	Sample ID	NGS type	Sequencing centre	CMM or UMM family	Meiosis between cases	Cases sequenced in family	Origin
5	5-6	Exome	Macrogen	UMM	3	3	France
5	5-12	Exome	Macrogen	UMM	3	3	France
5	5-15	Exome	Macrogen	UMM	3	3	France
594	594-1	Exome	Macrogen	UMM	5	2	France
594	594-12	Exome	Macrogen	UMM	5	2	France
908	908-1	Exome	Macrogen	UMM	2	2	France
908	908-8	Exome	NIH	UMM	2	2	France
1374	1374-1	Exome	NIH	UMM	4	2	France
1374	1374-10	Exome	Macrogen	UMM	4	2	France
LM10	mm564	Exome	Macrogen	CMM	3	3	Sweden
LM10	mm615	Exome	Macrogen	CMM	3	3	Sweden
LM10	mm648	Exome	Macrogen	CMM	3	3	Sweden
LM23	mm396	Exome	Macrogen	CMM	2	2	Sweden
LM23	mm641	Exome	Macrogen	CMM	2	2	Sweden
LM48	mm638	Exome	Macrogen	CMM	2	2	Sweden
LM48	mm700	Exome	Macrogen	CMM	2	2	Sweden
LM57	mm56	Exome	Macrogen	CMM	1	2	Sweden
LM57	mm64	Exome	Macrogen	CMM	1	2	Sweden
LM74	mm1611	Exome	Macrogen	UMM	4	2	Sweden
LM74	mm770	Exome	Macrogen	UMM	4	2	Sweden
LM78	mm774	Exome	Macrogen	UMM	3	2	Sweden
LM78	mm784	Exome	Macrogen	UMM	3	2	Sweden
LM82	mm809	Exome	Macrogen	UMM	2	1	Sweden
LM86	mm823	Exome	Macrogen	UMM	1	1	Sweden
Umeå-family	04/2949D	Exome	Macrogen	UMM	3	2	Sweden
Umeå-family	04/3108D	Exome	Macrogen	UMM	3	2	Sweden
Lynch-family	419-01	Exome	Macrogen	UMM	1	3	Sweden
Lynch-family	419-03	Exome	Macrogen	UMM	1	3	Sweden
Lynch-family	419-08	Exome	Macrogen	UMM	1	3	Sweden
1	210349	Exome	Macrogen	UMM	1	2	Denmark
1	90389	Exome	Macrogen	UMM	1	2	Denmark
2	10254	Exome	Macrogen	UMM	1	1	Denmark
3	3-1	Exome	Macrogen	UMM	4	3	Denmark
3	3-2	Exome	Macrogen	UMM	4	3	Denmark
3	3-3	Exome	Macrogen	UMM	4	3	Denmark
F100	110736	Exome	Macrogen	CMM	1	2	Denmark
F100	210949	Exome	Macrogen	CMM	1	2	Denmark
F117	200564	Exome	Macrogen	CMM	3	1	Denmark
F172	190342	Exome	Macrogen	CMM	1	2	Denmark
F172	220839	Exome	Macrogen	CMM	1	2	Denmark

Family ID	Sample ID	NGS type	Sequencing centre	CMM or UMM family	Meiosis between cases	Cases sequenced in family	Origin
F174	070650	Exome	Macrogen	CMM	1	2	Denmark
F174	090146	Exome	Macrogen	CMM	1	2	Denmark
F207	020946	Exome	Macrogen	CMM	1	2	Denmark
F207	160655	Exome	Macrogen	CMM	1	2	Denmark
F250	080545	Exome	Macrogen	UMM	3	2	Denmark
F250	311080	Exome	Macrogen	UMM	3	2	Denmark
F302	270550	Exome	Macrogen	CMM	4	2	Denmark
F302	280944	Exome	Macrogen	CMM	4	2	Denmark
F73	020830	Exome	Macrogen	CMM	1	2	Denmark
F73	221143	Exome	Macrogen	CMM	1	2	Denmark
F77	130948	Exome	Macrogen	CMM	1	2	Denmark
F77	200450	Exome	Macrogen	CMM	1	2	Denmark
33	33-01	Exome	Macrogen	UMM	2	2	Netherlands
33	33-04	Exome	TGen	UMM	2	2	Netherlands
3352	3352-209	Exome	TGen	UMM	3	1	USA
00002	00002-001	Exome	Macrogen	UMM	2	1	QIMR
00016	00016-001	Exome	Macrogen	UMM	1	1	QIMR
00023	00023-001	Exome	Macrogen	UMM	single case	1	QIMR
00029	00029-001	Exome	Macrogen	UMM	single case	1	QIMR
00041	00041-001	Exome	Macrogen	UMM	1	1	QIMR
00052	00052-001	Exome	Macrogen	UMM	1	1	QIMR
00058	00058-001	Exome	Macrogen	UMM	single case	1	QIMR
20014	20014-001	Exome	Macrogen	UMM	4	1	QIMR
20015	20015-002	Exome	Macrogen	UMM	3	2	QIMR
20015	20015-001	Exome	Macrogen	UMM	3	2	QIMR
40094	40094-001	Genome	Macrogen	CMM	spouses	2	QIMR
40094	40094-003	Genome	Macrogen	CMM	spouses	2	QIMR
40138	40138-007	Exome	Macrogen	CMM	3	3	QIMR
40138	40138-009	Genome	Macrogen	CMM	3	3	QIMR
40138	40138-010	Exome	Macrogen	CMM	3	3	QIMR
40354	40354-002	Exome	Macrogen	CMM	5	3	QIMR
40354	40354-007	Exome	Macrogen	CMM	5	3	QIMR
40354	40354-058	Exome	Macrogen	CMM	5	3	QIMR
40374	40374-006	Exome	Macrogen	CMM	2	3	QIMR
40374	40374-008	Exome	Macrogen	CMM	2	3	QIMR
40374	40374-010	Exome	Macrogen	CMM	2	3	QIMR
40449	40449-001	Exome	Macrogen	CMM	4	3	QIMR
40449	40449-003	Genome	CompGen	CMM	4	3	QIMR
40449	40449-004	Exome	Macrogen	CMM	4	3	QIMR
40800	40800-006	Genome	CompGen	CMM	4	3	QIMR
40800	40800-011	Exome	Macrogen	CMM	4	3	QIMR
40800	40800-015	Exome	Macrogen	CMM	4	3	QIMR
40804	40804-002	Exome	Macrogen	CMM	2	3	QIMR

Family ID	Sample ID	NGS type	Sequencing centre	CMM or UMM family	Meiosis between cases	Cases sequenced in family	Origin
40804	40804-005	Exome	Macrogen	CMM	2	3	QIMR
40804	40804-006	Exome	Macrogen	CMM	2	3	QIMR
40849	40849-002	Exome	Macrogen	CMM	3	3	QIMR
40849	40849-007	Exome	Macrogen	CMM	3	3	QIMR
40849	40849-010	Exome	Macrogen	CMM	3	3	QIMR
40851	40851-003	Exome	Macrogen	UMM	3	4	QIMR
40851	40851-005	Exome	QIMR	UMM	3	4	QIMR
40851	40851-006	Exome	Macrogen	UMM	3	4	QIMR
40851	40851-007	Genome	Macrogen	UMM	3	4	QIMR
40872	40872-002	Exome	Macrogen	CMM	4	4	QIMR
40872	40872-007	Exome	Macrogen	CMM	4	4	QIMR
40872	40872-009	Exome	Macrogen	CMM	4	4	QIMR
40928	40928-001	Exome	Macrogen	CMM	3	3	QIMR
40928	40928-003	Exome	Macrogen	CMM	3	3	QIMR
40928	40928-004	Exome	Macrogen	CMM	3	3	QIMR
40929	40929-001	Exome	Macrogen	CMM	4	4	QIMR
40929	40929-004	Exome	Macrogen	CMM	4	4	QIMR
40929	40929-005	Exome	Macrogen	CMM	4	4	QIMR
40929	40929-016	Exome	Macrogen	CMM	4	4	QIMR
40972	40972-007	Genome	CompGen	UMM	3	3	QIMR
40972	40972-012	Genome	Macrogen	UMM	3	3	QIMR
40972	40972-015	Exome	Macrogen	UMM	3	3	QIMR
40993	40993-007	Exome	Macrogen	CMM	1	1	QIMR
40999	40999-001	Exome	Macrogen	CMM	2	3	QIMR
40999	40999-003	Genome	CompGen	CMM	2	3	QIMR
40999	40999-004	Exome	Macrogen	CMM	2	3	QIMR
41015	41015-001	Exome	Macrogen	CMM	3	3	QIMR
41015	41015-009	Exome	Macrogen	CMM	3	3	QIMR
41015	41015-015	Exome	Macrogen	CMM	3	3	QIMR
41070	41070-001	Exome	Macrogen	UMM	2	1	QIMR
41096	41096-002	Exome	Macrogen	CMM	2	3	QIMR
41096	41096-004	Exome	Macrogen	CMM	2	3	QIMR
41096	41096-005	Exome	Macrogen	CMM	2	3	QIMR
41102	41102-002	Exome	Macrogen	CMM	2	3	QIMR
41102	41102-003	Genome	Macrogen	CMM	2	3	QIMR
41102	41102-012	Exome	Macrogen	CMM	2	3	QIMR
41109	41109-001	Exome	Macrogen	CMM	5	4	QIMR
41109	41109-003	Exome	Macrogen	CMM	5	4	QIMR
41109	41109-007	Exome	Macrogen	CMM	5	4	QIMR
41109	41109-016	Exome	Macrogen	CMM	5	4	QIMR
41111	41111-002	Exome	Macrogen	CMM	2	2	QIMR
41111	41111-007	Exome	Macrogen	CMM	2	2	QIMR
41114	41114-001	Exome	Macrogen	CMM	3	3	QIMR

Family ID	Sample ID	NGS type	Sequencing centre	CMM or UMM family	Meiosis between cases	Cases sequenced in family	Origin
41114	41114-003	Exome	Macrogen	CMM	3	3	QIMR
41114	41114-005	Exome	Macrogen	CMM	3	3	QIMR
41116	41116-003	Exome	Macrogen	UMM	4	3	QIMR
41116	41116-004	Exome	TGen	UMM	4	3	QIMR
41116	41116-018	Exome	Macrogen	UMM	4	3	QIMR
41142	41142-002	Exome	Macrogen	CMM	4	2	QIMR
41142	41142-004	Exome	Macrogen	CMM	4	2	QIMR
41155	41155-002	Exome	Macrogen	CMM	4	3	QIMR
41155	41155-003	Genome	Macrogen	CMM	4	3	QIMR
41155	41155-020	Exome	Macrogen	CMM	4	3	QIMR
41164	41164-001	Genome	Macrogen	UMM	2	4	QIMR
41164	41164-002	Genome	CompGen	UMM	2	4	QIMR
41164	41164-003	Exome	Macrogen	UMM	2	4	QIMR
41164	41164-005	Exome	Macrogen	UMM	2	4	QIMR
41167	41167-001	Exome	Macrogen	UMM	2	3	QIMR
41167	41167-003	Genome	Macrogen	UMM	2	3	QIMR
41167	41167-005	Exome	Macrogen	UMM	2	3	QIMR
41168	41168-005	Exome	Macrogen	CMM	1	2	QIMR
41168	41168-006	Exome	Macrogen	CMM	1	2	QIMR
41170	41170-004	Exome	Macrogen	CMM	5	4	QIMR
41170	41170-007	Exome	Macrogen	CMM	5	4	QIMR
41170	41170-008	Exome	Macrogen	CMM	5	4	QIMR
41170	41170-016	Exome	Macrogen	CMM	5	4	QIMR
41192	41192-006	Exome	Macrogen	UMM	2	1	QIMR
41193	41193-001	Exome	Macrogen	UMM	3	3	QIMR
41193	41193-005	Exome	Macrogen	UMM	3	3	QIMR
41193	41193-007	Exome	Macrogen	UMM	2	3	QIMR
41194	41194-001	Exome	Macrogen	CMM	3	4	QIMR
41194	41194-004	Exome	Macrogen	CMM	3	4	QIMR
41194	41194-008	Exome	Macrogen	CMM	3	4	QIMR
41194	41194-009	Exome	Macrogen	CMM	3	4	QIMR
41197	41197-001	Genome	Macrogen	CMM	2	2	QIMR
41197	41197-004	Exome	Macrogen	CMM	2	2	QIMR
41202	41202-001	Genome	Macrogen	CMM	3	3	QIMR
41202	41202-005	Exome	Macrogen	CMM	3	3	QIMR
41202	41202-009	Exome	Macrogen	CMM	3	3	QIMR
41216	41216-001	Exome	Macrogen	CMM	7	4	QIMR
41216	41216-004	Exome	Macrogen	CMM	7	4	QIMR
41216	41216-006	Exome	Macrogen	CMM	7	4	QIMR
41216	41216-013	Exome	Macrogen	CMM	7	4	QIMR
41219	41219-001	Genome	Macrogen	UMM	6	5	QIMR
41219	41219-004	Genome	Macrogen	UMM	3	2	QIMR
41219	41219-015	Exome	Macrogen	UMM	6	5	QIMR

Family ID	Sample ID	NGS type	Sequencing centre	CMM or UMM family	Meiosis between cases	Cases sequenced in family	Origin
41219	41219-020	Exome	Macrogen	UMM	6	5	QIMR
41219	41219-022	Exome	Macrogen	UMM	6	5	QIMR
41219	41219-030	Exome	Macrogen	UMM	3	2	QIMR
41220	41220-003	Exome	Macrogen	CMM	1	3	QIMR
41220	41220-005	Exome	Macrogen	CMM	2	3	QIMR
41220	41220-006	Exome	Macrogen	CMM	2	3	QIMR
41229	41229-001	Genome	Macrogen	CMM	1	4	QIMR
41229	41229-002	Exome	Macrogen	CMM	1	4	QIMR
41229	41229-003	Exome	Macrogen	CMM	1	4	QIMR
41229	41229-004	Exome	Macrogen	CMM	1	4	QIMR
41233	41233-004	Exome	Macrogen	CMM	5	2	QIMR
41233	41233-007	Exome	Macrogen	CMM	5	2	QIMR
41234	41234-001	Genome	Macrogen	CMM	1	2	QIMR
41234	41234-002	Exome	Macrogen	CMM	1	2	QIMR
41248	41248-003	Exome	Macrogen	CMM	7	2	QIMR
41248	41248-014	Genome	Macrogen	CMM	7	2	QIMR
41250	41250-002	Exome	Macrogen	CMM	1	2	QIMR
41250	41250-003	Genome	Macrogen	CMM	1	2	QIMR
41254	41254-001	Exome	Macrogen	UMM	6	3	QIMR
41254	41254-004	Genome	Macrogen	UMM	6	3	QIMR
41254	41254-027	Exome	Macrogen	UMM	6	3	QIMR
41294	41294-001	Genome	Macrogen	CMM	5	3	QIMR
41294	41294-002	Exome	Macrogen	CMM	5	3	QIMR
41294	41294-013	Exome	Macrogen	CMM	5	3	QIMR
41334	41334-003	Exome	Macrogen	CMM	4	3	QIMR
41334	41334-004	Exome	Macrogen	CMM	4	3	QIMR
41334	41334-006	Exome	Macrogen	CMM	4	3	QIMR
41348	41348-001	Genome	Macrogen	UMM	1	2	QIMR
41348	41348-802	Exome	Macrogen	CMM	1	2	QIMR
41397	41397-001	Exome	Macrogen	UMM	1	1	QIMR
41517	41517-001	Exome	Macrogen	UMM	2	2	QIMR
41517	41517-003	Exome	Macrogen	UMM	2	2	QIMR
50003	50003-001	Exome	Macrogen	UMM	2	3	QIMR
50003	50003-004	Exome	Macrogen	UMM	2	3	QIMR
50003	50003-005	Exome	Macrogen	UMM	2	3	QIMR
00-006	1413	Exome	Macrogen	CMM	4	3	QIMR
00-006	1462	Exome	Macrogen	CMM	4	3	QIMR
00-006	2304	Exome	Macrogen	CMM	4	3	QIMR
41185	41185-001	Exome	Macrogen	CMM	7	5	QIMR
41185	41185-007	Exome	Macrogen	CMM	7	5	QIMR
41185	41185-008	Exome	Macrogen	CMM	7	5	QIMR
41185	41185-011	Exome	Macrogen	CMM	7	5	QIMR
41185	41185-022	Exome	Macrogen	CMM	7	5	QIMR

Family ID	Sample ID	NGS type	Sequencing centre	CMM or UMM family	Meiosis between cases	Cases sequenced in family	Origin
1003	1003-017	Exome	TGen	UMM	2	1	UK
20963	975	Exome	Macrogen	UMM	3	1	WICR
21065	1067	Genome	CompGen	CMM	3	1	WICR
22070	2071	Genome	CompGen	UMM	6	2	WICR
22070	2080	exome	Macrogen	UMM	6	2	WICR
22108	2109	Exome	Macrogen	UMM	2	2	WICR
22108	2110	Exome	TGen	UMM	2	2	WICR
22198	2198	Exome	Macrogen	UMM	3	2	WICR
22198	2200	Exome	TGen	UMM	3	2	WICR
22309	2310	Exome	Macrogen	UMM	1	1	WICR
22354	2835	exome	TGen	CMM	6	1	WICR
22412	8683	Exome	Macrogen	UMM	2	1	WICR
22669	2669	Genome	CompGen	CMM	4	3	WICR
22669	2673	Exome	Macrogen	CMM	4	3	WICR
22669	2680	Exome	Macrogen	CMM	4	3	WICR
31662	1755	Exome	Macrogen	UMM	4	3	WICR
31662	1762	Exome	Macrogen	CMM	4	3	WICR
31662	1764	Genome	CompGen	CMM	4	3	WICR
32155	2163	Exome	Macrogen	UMM	2	1	WICR

QIMR is the Queensland Institute of Medical Research, Brisbane, Australia

WICR is the Westmead Institute for Cancer Research, Sydney, Australia

Macrogen is a NGS sequencing service provider

CompGen refers to Complete Genomics

NCI is the National Cancer Institute, USA

TGen is the Translational Genomics Research Institute

Chapter 2

Prevalence of germline *BAP1, CDKN2A* and *CDK4* mutations in an Australian population-based sample of cutaneous melanoma cases

Manuscript 1

Lauren G. Aoude, Michael Gartside, Peter Johansson, Jane M. Palmer, Judith Symmons, Nicholas G. Martin, Grant W. Montgomery, Nicholas K. Hayward. Prevalence of germline *BAP1, CDKN2A* and *CDK4* mutations in an Austrlian population based sample of cutaneous melanoma cases. *Twin Research and Human Genetics* (manuscript submitted 2/10/2014).

2.1 Relevance to thesis aims

This manuscript assesses the contribution of *BAP1, CDKN2A* and *CDK4* mutations to a population-based sample of cutaneous melanoma cases from Queensland. Australia, and in particular Queensland, has the world's highest rate of CMM. Currently, the genetic contribution to the majority of these cases is unexplained. Further characterisation of these risk factors could help better manage the disease in the future. This manuscript describes the use of a targeted pull-down sequencing approach to genotype a large sample of cases in a cost-effective way using Ion Torrent Personal Genome Machine (PGM) platform. This addresses Aim 1 of this thesis by capturing the contribution of know high-risk melanoma genes to CMM development in Queensland. Cases from the QFMP (N=1109) were sequenced using the PGM to look for mutations in *CDKN2A* and *BAP1*. Point mutations (p.R24C, p.R24H) in *CDK4* were assessed using the Sequenom mass spectrometry platform.

There have only been only two population-based studies of germline *CDKN2A* mutation published to date [36, 37]. This report captures a better estimate of *CDKN2A* and *CDK4* contribution to CMM in the Queensland population. Furthermore, this study reports three novel mutations (p.R22W, p.G35R and p.I49F) and shows that the total observed variants in *CDKN2A* accounted for approximately 1.31% of population-based CMM samples in Queensland. *BAP1* has been only recently shown to predispose to melanoma and its contribution to CMM development in the general population is unreported. This is the first study to deduce the overall contribution of *BAP1* to CMM predisposition. We show that mutations in *BAP1* are found in 0.63% of population-based CMM

cases in a geographic area with high ambient UVR. Genotyping showed that no case in the QFMP (N=1550) was a carrier of either of the known pathogenic variants in *CDK4* encoding either p.R24C or p.R24H substitutions. Overall, 1.9% of all probands of Queensland CMM families harboured mutations in a known high-penetrance melanoma susceptibility gene.

2.2 Contribution of candidate

I designed the Ampliseq panel to assess the contribution of *CDKN2A* and *BAP1* with the guidance of Michael Gartside. I sequenced 21 chips (Ion 318v2) on the Ion Torrent platform. This included the library preparation and amplification, binding of the Ion Sphere particles to the library on the One Touch 2 (OT2) and sequencing of samples using the PGM. I also performed all the data analysis for variants reported through the Torrent Suite software. I then performed the Sanger sequencing validation of variants to confirm they were true and not an artefact of the sequencing pipeline. I also performed the genotyping of the *CDK4* p.R24C and p.R24H variants on the Sequenom MassArray platform using iPLEX chemistry. I wrote and edited the manuscript under the supervision of Nicholas Hayward.

2.3 Contribution of other authors

Michael Gartside provided training on the Ion Torrent OT2 and PGM as well as training on the use of the Torrent Suite of analysis software. He also assisted with the library preparation and the sequencing of chips 1 to 6 on the Ion Torrent platform. Samples were collected by Nicholas Hayward, Nicholas Martin and Grant Montgomery. Jane Palmer and Judith Symmons collected the patient data relating to the families described in the manuscript. Peter Johansson provided the bioinformatics support for the Ion Torrent targeted pull-down data. Nicholas Hayward conceived the study design and edited the manuscript. All authors read and approved of the manuscript.

Prevalence of germline *BAP1, CDKN2A* and *CDK4* mutations in an Australian population-based sample of cutaneous melanoma cases

Lauren G. Aoude[1, 2], Michael Gartside[1], Peter Johansson[1], Jane M. Palmer[1], Judith Symmons[1], Nicholas G. Martin[1], Grant W. Montgomery[1], Nicholas K. Hayward[1,2]

[1]QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia
[2]University of Queensland, Brisbane, QLD, Australia

Corresponding author: Lauren G. Aoude, email Lauren.Aoude@qimrberghofer.edu.au, phone +61 7 3362 0323, fax +61 7 3845 3508

Running title: *BAP1*, *CDKN2A* and *CDK4* mutations in melanoma

Abstract

Mutations in *CDKN2A* and *CDK4* contribute to susceptibility in approximately 40% of high-density cutaneous melanoma (CMM) families and about 2% of unselected CMM cases. *BAP1* has been more recently shown to predispose to CMM and uveal melanoma (UMM) in some families; however, its contribution to CMM development in the general population is unreported. We sought to determine the contribution of these genes to CMM susceptibility in a population-based sample of cases from Australia. We genotyped 1,109 probands from Queensland families and found that approximately 1.31% harboured mutations in *CDKN2A*, including some with novel missense mutations (p.R22W, p.G35R and p.I49F). *BAP1* missense variants occurred in 0.63% of cases but no *CDK4* variants were observed in the sample. This is the first estimate of the contribution of *BAP1* and *CDK4* to a population-based sample of CMM and supports the previously reported estimate of *CDKN2A* germline mutation prevalence.

Introduction

Many environmental and genetic factors play a part in melanomagenesis. While exposure to ultraviolet radiation (UVR) plays a significant role in melanoma development, an underlying genetic predisposition can also contribute to an individual's risk. Studies have shown that ~10% of cutaneous malignant melanoma (CMM) cases occur in people that have a family history of melanoma (Gruis et al., 1995; Hussussian et al., 1994; MacGeoch et al., 1994; Soufir et al., 1998; Walker et al., 1995; Zuo et al., 1996). Known high-risk genes account for melanoma susceptibility in a proportion of these families. The major melanoma predisposition locus is Cyclin-Dependent Kinase Inhibitor 2A (*CDKN2A*) on chromosome 9, which encodes two tumour suppressors, p16INK4A and p14ARF, that inhibit progression of potential cancer cells by inducing senescence or apoptosis, respectively (de Snoo & Hayward, 2005; Palmieri et al., 2009). *CDKN2A* is thus involved in two of the most important tumour suppressor pathways, the p53 and the retinoblastoma (RB) pathways.

The largest study of germline mutations in *CDKN2A* to date was conducted by the International Melanoma Genetics Consortium (GenoMEL) in which 466 families from North America, Europe, Asia and Australia were screened for mutations in *CDKN2A* and *CDK4* (Goldstein et al., 2006). The study found that 41% of high-density families (defined by case-load depending on the region of origin) had germline variants, with 38% carrying a p16 mutation, 1.5% carrying a p14 mutation and 1% carrying a *CDK4* mutation. This study reported 57 unique melanoma-associated mutations in p16. In contrast, only two variants in *CDK4* have been attributed to melanoma risk; they occur at the same codon and give rise to amino acid substitutions p.R24C and p.R24H (Soufir et al., 1998; Zuo et al., 1996). This is the position at which p16 binds to CDK4, normally inhibiting the ability of CDK4 to negatively regulate the RB pathway (Zuo et al., 1996). A study into the prevalence of *CDKN2A* and *CDK4* mutations in CMM cases from a Greek hospital-based sample found that 5% of cases (16 of 320) harboured a mutation in one of these genes (Nikolaou et al., 2011). Mutations were most common in those with a family history (2 of 9) or multiple primary melanomas (4 of 7).

To date, there have been only two published studies determining the prevalence of high-risk predisposition genes in population-based samples of CMM cases. The Genes Environment and Melanoma (GEM) study genotyped probands from nine different geographical regions in the USA, Canada, Italy and Australia (Begg et al., 2005). They discovered 65 *CDKN2A* mutation carriers in a sample of 3,550 affected individuals, equating to a population frequency of approximately 2%. A second study looked at the contribution of *CDKN2A* to melanoma to a population-based sample of

cases from Queensland, assessing the link between both novel *CDKN2A* mutations and polymorphisms in CMM cases (Aitken et al., 1999). The sample comprised 482 families and *CDKN2A* mutations only occurred in high risk individuals in this group, with 9 out of 87 high-risk families harbouring a mutation. This study extrapolated this data to a cohort of 7,784 cases that had survey data available, of which 151 fitted the high risk profile. By hypothesising that *CDKN2A* mutations will occur at a similar frequency in the survey group, it was estimated that *CDKN2A* mutations might occur in 0.2% of melanoma cases in the population-based Queensland sample (Aitken et al., 1999). This calculation was based on a very small number of individuals screened and thus a broader more comprehensive study of the Queensland Familial Melanoma Project (QFMP) is required to get a better estimate of prevalence and to more fully understand the prevalence of *CDKN2A* mutations in this population-based sample.

BRCA-1 associated protein-1 (*BAP1*) is a tumour suppressor gene located on chromosome 3. In 2010, a seminal study reported the occurrence of *BAP1* somatic mutations in a panel of sporadic uveal melanoma (UMM) cases; notably, a single UMM case also carried a germline *BAP1* mutation (Harbour et al., 2010). Since then, *BAP1* has been linked to predisposition of a spectrum of cancer types including UMM, CMM, mesothelioma, renal cell carcinoma, basal cell carcinoma, as well as a distinct type of benign melanocytic tumour (Abdel-Rahman et al., 2011; Aoude et al., 2013; Carbone et al., 2012; Cheung et al., 2013; de la Fouchardiere et al., 2014; Harbour et al., 2010; Hoiom et al., 2013; Njauw et al., 2012; Popova et al., 2013; Testa et al., 2011; Wadt et al., 2014; Wiesner et al., 2011). Population-based and clinic-based studies of *BAP1* germline mutation prevalence in UMM cases have been reported and show that *BAP1* contributes to only a small proportion of cases overall (3-4%) (Aoude et al., 2013; Njauw et al., 2012). While CMM is associated with the *BAP1* tumour spectrum, the contribution of germline mutations in this gene to a population based-sample of CMM cases has not been reported.

The primary aim of this study was to more fully quantify the contribution of constitutional *CDKN2A* and *CDK4* mutations in an Australian population-based sample of CMM cases. A secondary aim was to determine the prevalence of germline *BAP1* mutations in this sample.

Methods and Materials

Ethics

Written consent was obtained for each participant in this study. Ethics approval was obtained from the QIMR Berghofer Human Research Ethics Committee (HREC).

Study cohort

Samples were ascertained as part of the Q-MEGA project, a population-based study from Queensland investigating the associations between genes and environment in CMM development (Baxter et al., 2008). Q-MEGA is made up of 4 distinct CMM case sample collections: childhood, adolescent, men over 50 and the Queensland Familial Melanoma Project (QFMP) (Aitken et al., 1996). Individuals that presented with histologically confirmed CMM and were reported to the Queensland Cancer Registry between the years 1982 and 1990 were approached to participate in the study. This accounted for approximately 95% of CMM cases diagnosed in Queensland over this period (Aitken et al., 1996). Cases were asked to fill out a questionnaire pertaining to family history, pigmentation, freckling, mole count and likelihood of sunburn. In the instance where an individual had a family history of melanoma, the first degree relatives and affected cases were also ascertained. A follow up study in 2002 - 2005 collected updated data and additional blood samples. A total of 1,897 individual families were sampled and stratified into three categories according to a standardized family risk index, previously described (Aitken et al., 1994). Generally, although there are a few exceptions, individuals with no family history of CMM were categorised as low-risk, two-case families were categorised as intermediate-risk and families with three or more cases were categorised as high-risk. Additionally twins with CMM collected in Queensland and New South Wales were included (Shekar et al., 2009). For the current study a random selection of cases from across Q-MEGA was used (Table 1).

Sample preparation

Blood samples were obtained from 1,109 family probands ascertained through the QFMP (Aitken et al., 1996). Genomic DNA was extracted using a standard salting out method (Miller et al., 1988).

Targeted Sequencing of CDKN2A and BAP1 using the Ion Torrent PGM

Melanoma cases were assessed for variants in *BAP1* and *CDKN2A* in a targeted sequencing approach. Using Ion AmpliSeq library kits (Life Technologies, CA, USA), 10 ng of genomic DNA from each proband were amplified using custom designed primer pools. The AmpliSeq panel was designed with a coverage of 40X across all regions with amplicon lengths of 150 bp to 250 bp.

BAP1 and *CDKN2A* had coverage 96% and 97% respectively. Ion Xpress barcode adapters 1-64 were used to pool samples. Unamplified libraries were purified using Agencourt Ampure XP reagent (Beckman Coulter, CA, USA) in order to eliminate fragments < 100 bp and increase the proportion of on-target reads. Libraries were equalised to ~100 pM using Ion Library Equalizer kits then combined into a single sample. A portion of the library (4 µl) was then diluted in 21 µl nuclease-free water to create a working stock. Using OT2 200 Kits (Life Technologies, CA, USA), clonal amplification and template enrichment of the Ion Sphere particles was performed. The template quality was assessed using a Qubit 2.0 (Life Technologies, CA, USA). Finally, Ion 318v2 chips were run on a Personal Genome Machine (Life Technologies, CA, USA) with 500 run flows per chip.

Overall, 1,109 samples had mean sequence coverage of 30X and therefore the sequencing of these samples was considered to be of sufficient depth to give accurate mutation data. Sequence data were analysed using Torrent Suite software (Life Technologies, CA, USA). SAMTOOLS was used to detect SNPs (Li et al., 2009). In order to minimise the false positive rates, filtering criteria were applied to the output. Firstly, variants were required to have minimum of 4 reads for the reference and 4 reads for the alternative alleles. The variant allele also had to compromise at least 20% of the total read count. Quality score had to be > 40. Synonymous variants were excluded as well as variants occurring commonly in the NHLBI Exome Sequencing Project (ESP6500) (minor allele frequency >0.001). This sculpted the final list of variants which were verified using Sanger sequencing. As a previous study has reported the *CDKN2A* mutations in the QFMP high-risk cohort (Aitken et al., 1999), *CDKN2A* genotyping of these samples was not repeated here but the published results were included in the overall prevalence statistics. Sanger sequencing was used to validate the *CDKN2A* and *BAP1* variants identified through targeted sequencing. See Supplementary Table S1 for the list of primers used. Supplementary Material is available on the Cambridge Journals Online website.

CDK4 genotyping using the Sequenom MassArray

The two known familial melanoma variants in *CDK4* (p.R24C and p.R24H) were genotyped in the QFMP probands using a Sequenom iPLEX gold assay (Sequenom Bioscience, CA, USA). The MassArray designer software was used to design the forward, reverse and extension primers for p.R24C (acgttggatgagtggctgaaattggtgtcg, acgttggatgtcacactcttgagggccac and gcactgtggggatcac) and p.R24H (acgttggatgagtggctgaaattggtgtcg, acgttggatgtcacactcttgagggccac and ttggccactgtggggatca). IPLEX Gold PCR amplification reactions were set up according to standard manufacturer protocols. Cluster plots were analysed using Typer Analyzer software 4.0.

Results

Targeted sequencing of an unselected population of CMM cases from Queensland, Australia revealed 6 of 1,055 cases harbour a missense variant in *CDKN2A* (Table 2). We report a single incidence of each of the following variants occurring in p16 (NM_000077): p.L16P, p.R22W, p.G35R and p.I49F. We also report two incidences of p.G101W (rs104894094), a common founder mutation in European melanoma populations (Goldstein et al., 2006). We found a recurrent mutation, p.A121T (rs199888003), in p14 (NM_058195) in two cases. As this residue is not overly conserved in primates (Rhesus, Figure 1), this is likely to be a non-deleterious mutation and is therefore not included in the overall prevalence statistics. Seventy-two cases carried the well-documented *CDKN2A* p.A148T polymorphism, giving it a minor allele frequency (MAF) of 0.0353. The MAF reported for this variant in the European American cohort (N = 4300) of the ESP6500 is 0.0225, which correlates well with the occurrence in our sample. CDKN2A p.A148T has also been omitted from the prevalence statistics. Of the variants seen here, three are novel (not previously reported in dbSNP, ESP6500 or the 1000 Genomes project) and have not been previously reported to the Leiden Open Variation Database (p.R22W, p.G35R and p.I49F). All occur at evolutionarily conserved amino acids (Figure 1), and although none of these variants have been reported as somatic mutations in CMM in the COSMIC (catalogue of somatic mutations in cancer) database, p.G35R has been seen in ovarian and pancreatic cancers (Table 3).

The incidence of *CDKN2A* mutation in the QFMP high-risk probands has been reported previously (Supplementary Table S2) (Aitken et al., 1999). We did not re-genotype *CDKN2A* in these samples here, as the results for over 95% of the cases are in the report by Aitken et al., 1999. In that study, Sanger sequencing found that 9 of 87 probands had a *CDKN2A* mutation. When these results are combined with those of the current study, we find that 1.31% (15 of 1142) of the Australian population-based CMM sample carries a *CDKN2A* mutation affecting p16. The childhood cohort has also been partly reported on previously, with the p.L16P mutation being found in a case presenting with multiple primary melanomas by the age of 12 (Whiteman et al., 1997), the results of which have been replicated here.

Genotyping of *CDK4* p.R24C and p.R24H showed that no proband in this population-based sample is a carrier of a melanoma-associated *CDK4* mutation.

Targeted sequencing of this unselected population of CMM cases from Queensland, Australia found 7 out of 1,109 cases harbour a missense variant in *BAP1* (Table 2). Two novel variants occurred in the ubiquitin carboxy-terminal hydrolase domain (p.G121R and p.R150C), one novel variant

occurred in the BARD1 interacting domain (p.P222T), two novel variants occurred outside of any known domain (p.N446I and p.P519A) and a recurrent variant (p.V604M) was seen in two probands in the BRCA1 interacting domain (Figure 2). In our study the minor allele frequency for p.V604M is 0.0018, compared to the MAF reported in the ESP6500 of 0.000154. Overall, 0.63% of cases harboured a missense variant in *BAP1*.

Since some missense mutations in BAP1 have been shown to affect alternative splicing (Wadt et al., 2012; Popova et al., 2013), the Automated Splice Site and Exon Definition Analyses (ASSEDA) online tool (https://splice.uwo.ca), was used to assess whether *BAP1* mutations in this study might create cryptic acceptor/donor sites (Rogan et al., 2003). No splice site alterations were predicted. Mutations in BAP1 occurred in highly conserved residues across species (Figure 1) and the five novel mutations (p.G121R, p.R150C, p.P222T, p.N446I and p.P519A) were predicted to be damaging by SIFT and/or Polyphen 2 (Table 2).

Discussion

We identified *CDKN2A* missense mutations in 0.57% of melanoma cases in the intermediate, low, twin, childhood, adolescent and men over 50 cohorts of the QFMP. When this data is combined with that for the QFMP high-risk group published by Aitken et al. 1999 (Aitken et al., 1999), 1.31% of the overall Queensland population-based sample of CMM cases carried a *CDKN2A* mutation. The estimate for *CDKN2A* we report here is in keeping with that reported (~2%) in the GEM study (Begg et al., 2005). As might be expected, since *CDKN2A* mutation has been associated with an increased risk of development of cancer in general (Mukherjee et al., 2012), *CDKN2A* mutation-positive families are often enriched for other cancers (Tables 3 and 4), although absolute numbers are too low in the QFMP families to show statistically significant association with any particular cancer type other than melanoma.

The *CDKN2A* p.G35R mutation has been reported to occur somatically in the COSMIC (catalogue of somatic mutations in cancer) database, but here we report for the first time its occurrence in the germline of an individual presenting with melanoma and colorectal cancer. A study using both functional and computational prediction of p16 mutations has found that the pathogenicity of this mutation is uncertain (Scaini et al., 2014). The p.A148T variant was observed at a frequency of 6.8% in the Queensland melanoma cases. This substitution is generally not thought to be deleterious, but the functional effect of this variant has been debated in the literature (Debniak et al., 2005; Spica et al., 2006). It has been shown to occur more frequently in Celtic populations and therefore its potential association with CMM risk may be due to its prevalence in a more melanoma-prone population (Aitken et al., 1999).

We found that 0.63% of CMM cases harboured germline missense mutations in *BAP1*. This mutation rate is similar to that reported by Njauw et al (Njauw et al., 2012) when they screened a hospital-based sample of 193 CMM families for *BAP1* mutation and reported 1 truncating mutation (0.5%). To date, all disease-associated variants in *BAP1* have been shown to truncate the protein. The variants we report here are all missense mutations that occur at highly conserved residues in mammals (Figure 2). None were predicted to alter splicing, thus at this stage it is not clear whether they are responsible for melanoma susceptibility in these individuals, or if they represent rare, benign polymorphisms.

We did not observe any mutations in *CDK4*. This result is in keeping with the low frequency of mutations of this gene reported in the literature. To date, only 17 families worldwide have been documented to carry *CDK4* mutations (Puntervoll et al., 2013; Soufir et al., 1998; Zuo et al., 1996).

All reported pathogenic mutations in p14 have been splice mutations, whole gene deletions or insertions (Harland et al., 2005; Mistry et al., 2005; Randerson-Moor et al., 2001; Rizos et al., 2001). Currently, no pathogenic missense mutations are known to occur in p14.

In summary, we report 3 novel variants in *CDKN2A* and 3 previously reported mutations, along with 7 novel missense variants in *BAP1*. Germline mutations that we observe for the two genes combined (~2%) thus account for only a small proportion of all CMM cases in this population. This suggests that genes other than the high-penetrance familial melanoma genes are responsible for the bulk of CMM susceptibility in the general population.

Acknowledgements

The authors would like to thank the participants of this study. This project was funded by the National Health and Medical Research Council of Australia (NHMRC). LGA receives support from an Australia and New Zealand Banking Group Limited Trustees PhD scholarship. NKH and GWM are supported by fellowships from the NHMRC.

References

Abdel-Rahman, M. H., Pilarski, R., Cebulla, C. M., Massengill, J. B., Christopher, B. N., Boru, G., . . . Davidorf, F. H. (2011). Germline BAP1 mutation predisposes to uveal melanoma, lung adenocarcinoma, meningioma, and other cancers. *J Med Genet, 48*(12), 856-859.

Aitken, J., Welch, J., Duffy, D., Milligan, A., Green, A., Martin, N., & Hayward, N. (1999). CDKN2A variants in a population-based sample of Queensland families with melanoma. *J Natl Cancer Inst, 91*(5), 446-452.

Aitken, J. F., Duffy, D. L., Green, A., Youl, P., MacLennan, R., & Martin, N. G. (1994). Heterogeneity of melanoma risk in families of melanoma patients. *Am J Epidemiol, 140*(11), 961-973.

Aitken, J. F., Green, A. C., MacLennan, R., Youl, P., & Martin, N. G. (1996). The Queensland Familial Melanoma Project: study design and characteristics of participants. *Melanoma Res, 6*(2), 155-165.

Aoude, L. G., Vajdic, C. M., Kricker, A., Armstrong, B., & Hayward, N. K. (2013). Prevalence of germline BAP1 mutation in a population-based sample of uveal melanoma cases. *Pigment Cell Melanoma Res, 26*(2), 278-279.

Baxter, A. J., Hughes, M. C., Kvaskoff, M., Siskind, V., Shekar, S., Aitken, J. F., . . . Whiteman, D. C. (2008). The Queensland Study of Melanoma: environmental and genetic associations (Q-MEGA); study design, baseline characteristics, and repeatability of phenotype and sun exposure measures. *Twin Res Hum Genet, 11*(2), 183-196.

Begg, C. B., Orlow, I., Hummer, A. J., Armstrong, B. K., Kricker, A., Marrett, L. D., . . . Berwick, M. (2005). Lifetime risk of melanoma in CDKN2A mutation carriers in a population-based sample. *J Natl Cancer Inst, 97*(20), 1507-1515.

Carbone, M., Korb Ferris, L., Baumann, F., Napolitano, A., Lum, C. A., Flores, E. G., . . . Yang, H. (2012). BAP1 cancer syndrome: malignant mesothelioma, uveal and cutaneous melanoma, and MBAITs. *J Transl Med, 10*(1), 179.

Cheung, M., Talarchek, J., Schindeler, K., Saraiva, E., Penney, L. S., Ludman, M., & Testa, J. R. (2013). Further evidence for germline BAP1 mutations predisposing to melanoma and malignant mesothelioma. *Cancer Genet, 206*(5), 206-210.

de la Fouchardiere, A., Cabaret, O., Savin, L., Combemale, P., Schvartz, H., Penet, C., . . . Bressac-de Paillerets, B. (2014). Germline BAP1 mutations predispose also to multiple basal cell carcinomas. *Clin Genet*.

de Snoo, F. A., & Hayward, N. K. (2005). Cutaneous melanoma susceptibility and progression genes. *Cancer Lett, 230*(2), 153-186.

Debniak, T., Scott, R. J., Huzarski, T., Byrski, T., Rozmiarek, A., Debniak, B., . . . Lubinski, J. (2005). CDKN2A common variants and their association with melanoma risk: a population-based study. *Cancer Res, 65*(3), 835-839.

Goldstein, A. M., Chan, M., Harland, M., Gillanders, E. M., Hayward, N. K., Avril, M. F., . . . Yakobson, E. (2006). High-risk melanoma susceptibility genes and pancreatic cancer, neural system tumors, and uveal melanoma across GenoMEL. *Cancer Res, 66*(20), 9818-9828.

Gruis, N. A., van der Velden, P. A., Sandkuijl, L. A., Prins, D. E., Weaver-Feldhaus, J., Kamb, A., . . . Frants, R. R. (1995). Homozygotes for CDKN2 (p16) germline mutation in Dutch familial melanoma kindreds. *Nat Genet, 10*(3), 351-353.

Harbour, J. W., Onken, M. D., Roberson, E. D., Duan, S., Cao, L., Worley, L. A., . . . Bowcock, A. M. (2010). Frequent mutation of BAP1 in metastasizing uveal melanomas. *Science, 330*(6009), 1410-1413.

Harland, M., Taylor, C. F., Chambers, P. A., Kukalizch, K., Randerson-Moor, J. A., Gruis, N. A., . . . Bishop, J. A. (2005). A mutation hotspot at the p14ARF splice site. *Oncogene, 24*(28), 4604-4608.

Hoiom, V., Edsgard, D., Helgadottir, H., Eriksson, H., All-Ericsson, C., Tuominen, R., . . . Hansson, J. (2013). Hereditary uveal melanoma: a report of a germline mutation in BAP1. *Genes Chromosomes Cancer, 52*(4), 378-384.

Hussussian, C. J., Struewing, J. P., Goldstein, A. M., Higgins, P. A., Ally, D. S., Sheahan, M. D., . . . Dracopoli, N. C. (1994). Germline p16 mutations in familial melanoma. *Nat Genet, 8*(1), 15-21.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25: 2078–2079

MacGeoch, C., Bishop, J. A., Bataille, V., Bishop, D. T., Frischauf, A. M., Meloni, R., . . . Spurr, N. K. (1994). Genetic heterogeneity in familial malignant melanoma. *Hum Mol Genet, 3*(12), 2195-2200.

Miller, S. A., Dykes, D. D., & Polesky, H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res, 16*(3), 1215.

Mistry, S. H., Taylor, C., Randerson-Moor, J. A., Harland, M., Turner, F., Barrett, J. H., . . . Bishop, D. T. (2005). Prevalence of 9p21 deletions in UK melanoma families. *Genes Chromosomes Cancer, 44*(3), 292-300.

Mukherjee, B., Delancey, J. O., Raskin, L., Everett, J., Jeter, J., Begg, C. B., . . . Investigators, G. E. M. S. (2012). Risk of non-melanoma cancers in first-degree relatives of CDKN2A mutation carriers. *J Natl Cancer Inst, 104*(12), 953-956.

Nikolaou, V., Kang, X., Stratigos, A., Gogas, H., Latorre, M. C., Gabree, M., . . . Tsao, H. (2011). Comprehensive mutational analysis of CDKN2A and CDK4 in Greek patients with cutaneous melanoma. *Br J Dermatol, 165*(6), 1219-1222.

Njauw, C. N., Kim, I., Piris, A., Gabree, M., Taylor, M., Lane, A. M., . . . Tsao, H. (2012). Germline BAP1 Inactivation Is Preferentially Associated with Metastatic Ocular Melanoma and Cutaneous-Ocular Melanoma Families. *PLoS One, 7*(4), e35295.

Palmieri, G., Capone, M., Ascierto, M. L., Gentilcore, G., Stroncek, D. F., Casula, M., . . . Ascierto, P. A. (2009). Main roads to melanoma. *J Transl Med, 7*, 86.

Popova, T., Hebert, L., Jacquemin, V., Gad, S., Caux-Moncoutier, V., Dubois-d'Enghien, C., . . . Stern, M. H. (2013). Germline BAP1 Mutations Predispose to Renal Cell Carcinomas. *Am J Hum Genet*, 92(6), 974-80.

Puntervoll, H. E., Yang, X. R., Vetti, H. H., Bachmann, I. M., Avril, M. F., Benfodda, M., . . . Molven, A. (2013). Melanoma prone families with CDK4 germline mutation: phenotypic profile and associations with MC1R variants. *J Med Genet, 50*(4), 264-270.

Randerson-Moor, J. A., Harland, M., Williams, S., Cuthbert-Heavens, D., Sheridan, E., Aveyard, J., . . . Bishop, D. T. (2001). A germline deletion of p14(ARF) but not CDKN2A in a melanoma-neural system tumour syndrome family. *Hum Mol Genet, 10*(1), 55-62.

Rizos, H., Puig, S., Badenas, C., Malvehy, J., Darmanian, A. P., Jimenez, L., . . . Kefford, R. F. (2001). A melanoma-associated germline mutation in exon 1beta inactivates p14ARF. *Oncogene, 20*(39), 5543-5547.

Rogan, P. K., Svojanovsky, S., & Leeder, J. S. (2003). Information theory-based analysis of CYP2C19, CYP2D6 and CYP3A5 splicing mutations. *Pharmacogenetics, 13*(4), 207-218.

Scaini, M. C., Minervini, G., Elefanti, L., Ghiorzo, P., Pastorino, L., Tognazzo, S., . . . Tosatto, S. C. (2014). CDKN2A Unclassified Variants in Familial Malignant Melanoma: Combining Functional and Computational Approaches for Their Assessment. *Hum Mutat*, 35(7), 828-40.

Shekar, S. N., Duffy, D. L., Youl, P., Baxter, A. J., Kvaskoff, M., Whiteman, D. C., . . . Martin, N. G. (2009). A population-based study of Australian twins with melanoma suggests a strong genetic contribution to liability. *J Invest Dermatol, 129*(9), 2211-2219.

Soufir, N., Avril, M. F., Chompret, A., Demenais, F., Bombled, J., Spatz, A., . . . 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*(2), 209-216.

Spica, T., Portela, M., Gerard, B., Formicone, F., Descamps, V., Crickx, B., . . . Melan, C. (2006). The A148T variant of the CDKN2A gene is not associated with melanoma risk in the French and Italian populations. *J Invest Dermatol, 126*(7), 1657-1660.

Testa, J. R., Cheung, M., Pei, J., Below, J. E., Tan, Y., Sementino, E., . . . Carbone, M. (2011). Germline BAP1 mutations predispose to malignant mesothelioma. *Nat Genet, 43(10), 1022-1025.*

Wadt, K., Choi, J., Chung, J. Y., Kiilgaard, J., Heegaard, S., Drzewiecki, K. T., . . . Brown, K. M. (2012). A cryptic BAP1 splice mutation in a family with uveal and cutaneous melanoma, and paraganglioma. *Pigment Cell Melanoma Res, 25(6), 815-8.*

Wadt, K. A., Aoude, L. G., Johansson, P., Solinas, A., Pritchard, A., Crainic, O., . . . Hayward, N. K. (2014). A recurrent germline BAP1 mutation and extension of the BAP1 tumor predisposition spectrum to include basal cell carcinoma. *Clin Genet*.

Walker, G. J., Hussussian, C. J., Flores, J. F., Glendening, J. M., Haluska, F. G., Dracopoli, N. C., . . . Fountain, J. W. (1995). Mutations of the CDKN2/p16INK4 gene in Australian melanoma kindreds. *Hum Mol Genet, 4*(10), 1845-1852.

Whiteman, D. C., Milligan, A., Welch, J., Green, A. C., & Hayward, N. K. (1997). Germline CDKN2A mutations in childhood melanoma. *J Natl Cancer Inst, 89*(19), 1460.

Wiesner, T., Obenauf, A. C., Murali, R., Fried, I., Griewank, K. G., Ulz, P., . . . Speicher, M. R. (2011). Germline mutations in BAP1 predispose to melanocytic tumors. *Nat Genet, 43*(10), 1018-1021.

Zuo, L., Weger, J., Yang, Q., Goldstein, A. M., Tucker, M. A., Walker, G. J., . . . Dracopoli, N. C. (1996). Germline mutations in the p16INK4a binding domain of CDK4 in familial melanoma. *Nat Genet, 12*(1), 97-99.

Table 1

Number of samples sequenced for *CDKN2A* and *BAP1* variants

study group	total number available	number sanger sequenced	number Ion torrent sequenced
high	91	87~	54^
intermediate	414	194~	243
low	1392	201~	565
men over 50	178	none	65
childhood	101	31*	48
adolescents	298	none	105
twins	125	none	29

~ Aitken et al., (1999) JNCI

* Whiteman et al., (1997) JNCI

^ *BAP1* sequencing only

Table 2

Variants in *BAP1* and *CDKN2A* identified through targeted sequencing

sample	risk group	gene	exon	position	reference	NT change	AA change	rsID	ESP6500 MAF*	SIFT	PolyPhen 2	GERP ++
006090	low	BAP1	5	52441988	NM_004656	c.361G>C	p.G121R	-	-	damaging	damaging	5.14
002203	low	BAP1	7	52441322	NM_004656	c.448C>T	p.R150C	-	-	damaging	damaging	5.38
009341	intermediate	BAP1	9	52440388	NM_004656	c.664C>A	p.P222T	-	-	damaging	damaging	5.69
011937	high	BAP1	13	52437824	NM_004656	c.1337A>T	p.N446I	-	-	tolerated	damaging	4.76
052809	intermediate	BAP1	13	52437606	NM_004656	c.1555C>G	p.P519A	-	-	damaging	damaging	5.7
050611	intermediate	BAP1	14	52437234	NM_004656	c.1810G>A	p.V604M	-	0.000154	tolerated	benign	1.49
051964	low	BAP1	14	52437234	NM_004656	c.1810G>A	p.V604M	-	0.000154	tolerated	benign	1.49
006528	low	CDKN2A (p14)	2	21971040	NM_058195	c.361G>A	p.A121T	rs199888003	0.000154	tolerated	na	2.33
002507	low	CDKN2A (p14)	2	21971040	NM_058195	c.361G>A	p.A121T	rs199888003	0.000154	tolerated	na	2.33
012039	childhood	CDKN2A (p16)	1	21974780	NM_000077	c.47T>C	p.L16P	-	-	damaging	damaging	4.14
050731	intermediate	CDKN2A (p16)	1	21974763	NM_000077	c.64C>T	p.R22W	-	-	damaging	damaging	-0.43
010255	low	CDKN2A (p16)	1	21974724	NM_000077	c.103G>A	p.G35R	-	-	damaging	probably damaging	4.84
051362	intermediate	CDKN2A (p16)	1	21974682	NM_000077	c.145A>T	p.I49F	-	-	damaging	damaging	4.22
008197	low	CDKN2A (p16)	2	21971057	NM_000077	c.301G>T	p.G101W	rs104894094	-	damaging	probably damaging	5.49
061542	adolescent	CDKN2A (p16)	2	21971057	NM_000077	c.301G>T	p.G101W	rs104894094	-	damaging	probably damaging	5.49

* European American population; na is not available; GERP++ is an estimate of the constrained elements in the human genome.

Table 3

CDKN2A mutations reported in publically available databases LOVD and COSMIC

AA change	present in LOVD	present in COSMIC CMM	present in COSMIC in other cancers
p.L16P	yes	no	biliary tract carcinoma, upper aerodigestive tract
p.R22W	no	no	no
p.R24P	yes	yes	CNS (glioma), soft tissue sarcoma
p.G35R	no	no	ovary, pancreas
p.I49F	no	no	no
p.G101W	yes	yes	lung carcinoma

LOVD is the Leiden Open Variation Database 3.0 which lists published germline *CDKN2A* variants in cancer; COSMIC is the Catalogue of Somatic Mutations in Cancer v68.

Table 4

Summary of cancers in families with germline *BAP1* and *CDKN2A* mutations

proband	gene	AA change	risk group	ages of onset of CMM	other cancers in proband (age)	other cancers in family (age)
6090	BAP1	p.G121R	low	45, 72		
2203	BAP1	p.R150C	low	32, 57	breast (57)	
9341	BAP1	p.P222T	intermediate	36		CMM (37)
1599	BAP1	p.L342F	Intermediate	47, 49		
11937	BAP1	p.N446I	high	53	lung (65), lung (68)	CMM (60)
52809	BAP1	p.P519A	intermediate	41		CMM (36)
50611	BAP1	p.V604M	intermediate	64		CMM (33)
51964	BAP1	p.V604M	low	59		lung (65)
6528	CDKN2A (p14)	p.A121T	low	69		
2507	CDKN2A (p14)	p.A121T	low	38		lung (65), colorectal (64)
12039	CDKN2A (p16)	p.L16P	childhood	12, 18, 22		CMM (19), CMM (21), CMM (23), CMM (25), CMM (43), CMM (44), Lung (40)
50731	CDKN2A (p16)	p.R22W	intermediate	45,		CMM (26), pancreas (52), oesophageal (76), prostate (58), stomach (63)
10255	CDKN2A (p16)	p.G35R	low	78	colorectal (86)	stomach (45), breast (66), oesophagus (60)
51362	CDKN2A (p16)	p.I49F	intermediate	26, 26, 27, 29, 45		CMM (38), breast (38), breast (48), lung (84)
8197	CDKN2A (p16)	p.G101W	low	47, 71		CMM (28), CMM (44)
61542	CDKN2A (p16)	p.G101W	adolescent	18, 30, 30, 35		CMM (36), cervix (31), lung (69)

CMM is cutaneous malignant melanoma; MPM is multiple primary melanomas

Figure 1 Conservation across species.

Conservation of protein altering missense variants in BAP1 and CDKN2A.

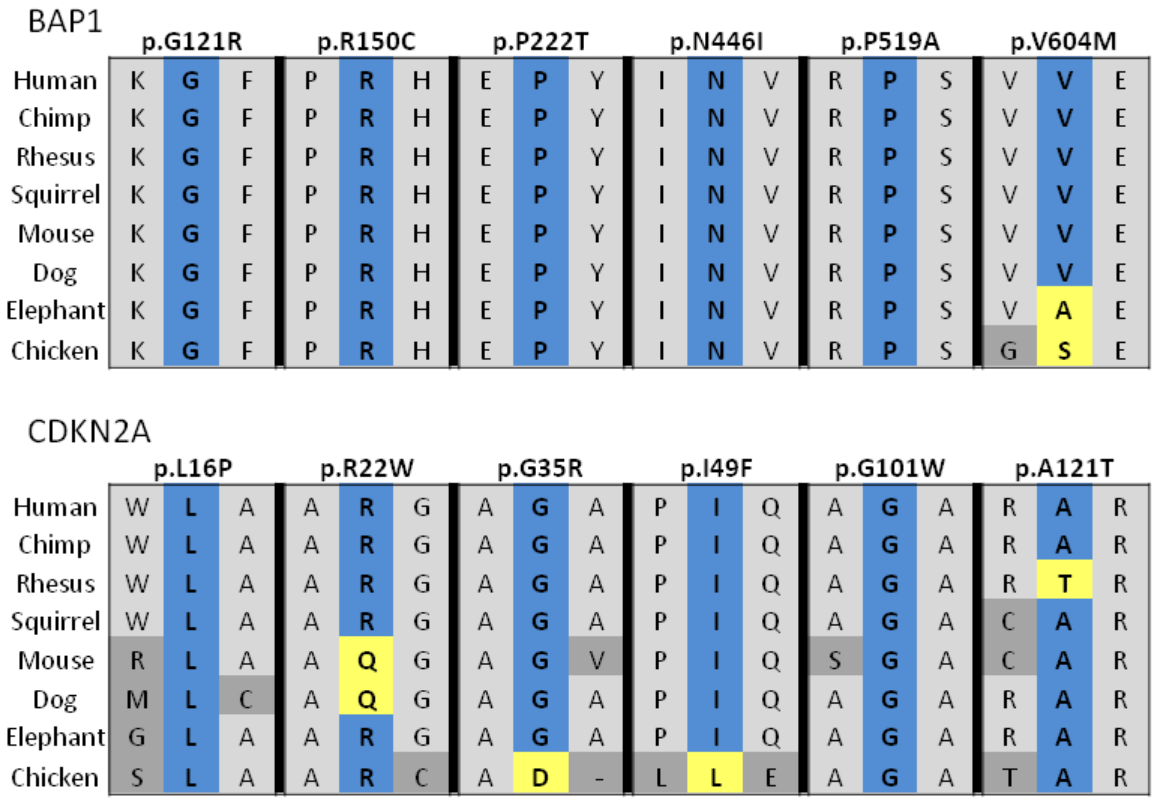
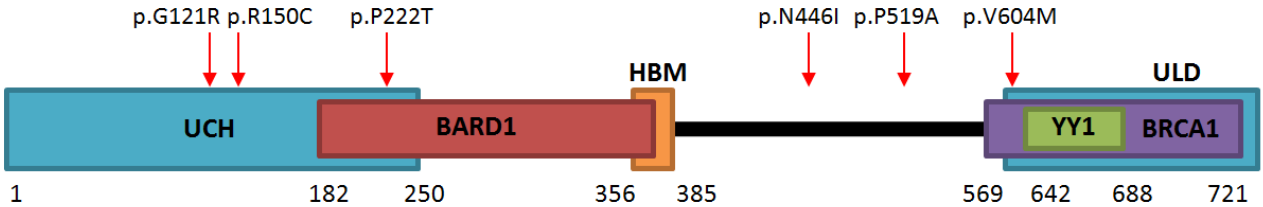


Figure 2 Germline variants in functional domains and protein interaction regions of BAP1.

All germline variants found in the population-based sample of cutaneous melanoma cases are shown in relation to their position in the protein. The arrows show the position of germline variants.UCH is the ubiquitin carboxy-terminal hydrolase domain; HBM is the HCFC1 binding motif; ULD is the UCH37-like domain. Binding sites for genes BARD1, BRCA1 and YY1 are depicted by their gene symbols.



Supplementary Material

Supplementary Table S1

Primers for Sanger sequencing validation

Amplicon	Primer Sequence
BAP1 exon 5 forward	TGGGTATTTGGTAGGTGCTTG
BAP1 exon 5 reverse	CCCGCAACTGCATCTAAAAAC
BAP1 exon 6 and 7 forward	TCCACCCATAGTCCTACCTGA
BAP1 exon 6 and 7 reverse	GGGCAATATGGTGTAGGGTGA
BAP1 exon 9 forward	TTCCAGATAGGCCCCTCATAC
BAP1 exon 9 reverse	GTGGTTAGCTGAAGCCCAGAT
BAP1 exon 11 forward	GGGGAGACTGTGAGCTTTTCT
BAP1 exon 11 reverse	ATCAGGCAGAGGAACCTAGCA
BAP1 exon 13 forward	TTCTGGGTACTGCTGGGTATG
BAP1 exon 13 reverse	GGACACTTTGTGGTCACTTGG
BAP1 exon 14 forward	CGTGTTGACTGCATACGCTAC
BAP1 exon 14 reverse	CCAATCTTCACACCAAAGTTCC
P16 exon 1 forward	ACGCACTCAAACACGCCTTTG
P16 exon 1 reverse	CAAACTTCGTCCTCCAGAGTC
P16 exon 2 forward	CAGAAGTTCGGAGGATATAATG
P16 exon 2 reverse	GGGCTGAACTTTCTGTGCTG
P14 forward	CCTCGCTTTCCTTTCTTCCT
P14 reverse	CGGTTATCTCCTCCTCCTCC

Supplementary Table S2

Summary of cancers in high-risk families reported by Aitken et al., 1999

Proband	Gene	Variant	Ages of onset of CMM	Other cancers in proband (age)	Other cancers in family (age)
001790	CDKN2A (p16)	p.R24P	28, 29, 36, 40		MPM (4 additional cases), CMM (42), CMM (65), CMM (65), CMM (72), CMM (72), breast (84), bladder (43), uterus (76)
006581	CDKN2A (p16)	p.L32P	30, 43, 51, 56	pancreas (54)	CMM (33), MPM (3 cases), bronchogenic (59), bowel (50), lung (65)
004197	CDKN2A (p16)	p.G35A	31, 31, 31, 32, 35		CMM (39), CMM (40), CMM (41), CMM (45), CMM (51), CMM (54), CMM (70), breast (39), lymphoma (62), Hogkins lymphoma (37), non-Hogkins lymphoma (52)
005526	CDKN2A (p16)	p.M53I	37		CMM (34), CMM (53), MPM (2 cases), ovary (41), carcinomatosis (49), caecum (83), liver (56)
003301	CDKN2A (p16)	p.M53I	26, 27, 27, 37		CMM (23), CMM (26), CMM (27), CMM (52), MPM (9 cases), rectum (43), prostate (72), breast (43), breast (48), breast (56), bile duct (64)
004478	CDKN2A (p16)	p.D108N	36, 38, 39, 46, 49, 58, 59, 63, 63, 64, 66, 66		CMM (19), CMM (46), MPM (28)
002622	CDKN2A (p16)	24 bp deletion	28		CMM (37), CMM (44)
000397	CDKN2A (p16)	24 bp duplication	38, 43, 43, 44, 44, 48, 49, 53		CMM (27), CMM (30), CMM (33), CMM (40), CMM (41), MPM (2 cases), lung (60), carcinomatosis (79)
002899	CDKN2A (p16)	Nt46delC	47		CMM (27), CMM (34), CMM (40), MPM (3 cases), stomach (52)

CMM is cutaneous malignant melanoma

MPM is multiple primary melanomas

Chapter 3

Assessment of *PALB2, SOX10* and *MITF* as candidate melanoma susceptibility genes

3.1 *SOX10* and *MITF*

3.1.1 Introduction

In order to address the second aim of this study, a candidate gene approach was used to search for novel melanoma predisposition genes. Genes for this study were selected based on a survey of the literature to identify candidates that may have a role in familial melanoma but are as yet undefined in the familial melanoma research landscape.

The first gene investigated as a candidate for predisposition encoded the transcription factor SRY (sex determining region Y)-box 10 (*SOX10).* It has been associated with Waardenburg syndrome, a rare genetic disorder for which characteristics include deafness, neural crest defects and pigmentation abnormalities [127]. *SOX10* acts in the pigmentation pathway and has a key role in the in the embryonic development of melanocytes. It also controls the transcription of microphthalmia-associated transcription factor (*MITF*). Cronin and colleagues investigated the role of *SOX10* in somatic melanoma development and found that 20% of metastatic cell lines harbour mutations in either *MITF* or *SOX10* [52]. It is yet to be ascertained whether *SOX10* could also be involved in melanoma predisposition, which makes it a good candidate for this study.

The next logical gene to interrogate was *MITF*. It also plays a role in melanocyte development and is the master regulator of melanocyte function [83]. Like *SOX10,* mutations in *MITF* can also lead to the development of Waardenburg syndrome [128]. Reports of germline *MITF* mutations in CMM were published in 2011. Two groups reported a novel germline variant, p.E318K, which increases the risk of melanoma development in the familial and sporadic settings [21, 55]. I was a lead author on one of the manuscripts describing p.E318K (rs149617956) in melanoma predisposition (*Nature.* 2011 Nov 13;480 (7375):99-103). This study identified 31 melanoma-prone families that carried the p.E318K variant, which segregated in some but not all of the families identified (log of odds score of 2.7). The variant was analysed in an Australian case-control study, which showed a significantly higher frequency in the cases (34 out of 2059) compared to the controls (14 out of 1953). Among cases, the risk allele occurred more frequently in those with multiple primary melanoma (OR 4.22,

95% CI 1.52–10.91), family history of melanoma (OR 2.95, 95% CI 1.23–6.92), or a combination of both (OR 8.37, 95% CI 2.58–23.80). These findings were also replicated in a UK case-control sample. The variant allele was also shown to be associated with increased naevus count and non-blue eye-colour. The manuscript was published in the first year of my PhD but has not been included as a chapter of this thesis as the majority of the experimental work was done prior to commencing my candidature. Refer to the appendix for the full manuscript. It is clear that p.E318K plays an important role in melanoma predisposition; whether other variants in *MITF* may also lead to an increased risk of CMM development is yet to be ascertained.

3.1.2 Methods

Ethics

Written consent was obtained for each participant in this study. Ethics approval for this project was obtained from the Human Research Ethics Committees (HREC) of the QIMR Berghofer Medical Research Institute.

Samples

Probands from dense melanoma families were selected for mutational screening. They were selected on the basis of case load (a minimum of 3 cases in the families) and/or having cases with multiple primary melanomas. As described by Yokoyama et al. [21] these are factors strongly associated with *MITF* p.E318K mutation and therefore could be a starting point for a screen of *SOX10* as a candidate predisposition gene or the discovery of further *MITF* variants. Families with mutations in *CDKN2A* or *CDK4* were excluded from the panel. A total of 182 probands were selected with a family profile that adhered to study boundaries.

Sanger sequencing

Sanger sequencing was used to detect pathogenetic mutations in *SOX10* and *MITF* in the selected probands. *SOX10* genotyping was carried out on 92 samples while 182 samples were interrogated for mutations in *MITF*. PCR reactions were run using ABI goldTaq PCR reagents. For a complete list of primers refer to Table 5.

Exome sequencing

CMM families (N=23) with a combination of whole-genome and whole-exome sequencing data were also included in the study. Each of these families had data available on up to three affected members. The sequencing was performed on the Illumina Hiseq 2000 platform using Illumina TrueSeq enrichment kits. A BWA alignment algorithm was used to map the sequencing output to the UCSC human genome reference build 19 [122]. SNPs and indels were annotated using SAMTOOLS [123]. Before analysis, the data set was filtered using a stringent set of cut-offs. Quality score was required to be greater than 70; number of alternate reads had to be greater than 2; and at least 20% of all reads had to be called with the alternate allele. This gave us confidence that the remaining data set would have a minimal false positive rate.

Case-control analysis

Novel and rare non-synonymous variants that were identified using Sanger sequencing methods and whole-genome/exome sequencing methods were analysed in a case-control cohort to determine whether they might be low to medium penetrance CMM predisposition variants. The case-control cohort was derived from an Australian population and includes a wide cross-section of cases from the QFMP[124] (N=1,630) with varied personal and family histories of melanoma. A portion of probands in this set were obtained from dense melanoma families while others are sporadic melanoma cases. This cohort also includes cases with diverse age of onset of CMM ranging from childhood disease to late onset melanoma. Controls (N=1,690) used in this panel were parents of twins ascertained as part of the Brisbane Twin Naevus Study [126].

Sequenom MassArray genotyping using iPLEX chemistry was run on this case-control sample. Variants were multiplexed on the Sequenom iPLEX gold system using 10ng of genomic DNA. The results were analysed using Typer Analyzer software 4.0.

3.1.3 Results

Sanger sequencing of 92 probands combined with exome sequencing data (N=23) for the three exons of *SOX10* (NM_006941) found only known single nucleotide polymorphisms within the coding region of the gene (Table 6), all of which were synonymous and therefore did not result in a protein change. The variants observed were p.D6D (rs149435516), p.N275N (rs200437243) and p.H309H (rs139884). Case-control analysis for *SOX10* variants was not performed as these all resulted in a synonymous change.

Exome sequence data (N=23) along with Sanger sequencing data of 182 probands for disease-causing mutations in *MITF* revealed three non-synonymous variants (Table 7). One of the variants, p.R72H was novel, that is, not previously reported in dbSNP or ESP6500. A case-control analysis showed that it was present only in the discovery family. The second variant p.T91M (rs201247895) has been previously reported. A case-control analysis identified two CMM cases and no controls with the variant. One individual had early onset CMM (at age 18years), and their mother, who was also a carrier, developed UMM, aged 68 years. The second individual with a p.T91M variant was a sporadic CMM case that similarly had early onset melanoma, at the age of 35 years. The third variant that was observed in this gene was the p.E318K variant previously identified as predisposing to melanoma (see Appendix).

3.1.4 Discussion

The data presented here shows no evidence that *SOX10* is involved in CMM predisposition as all variants that were observed resulted in a synonymous change to the protein and hence is not likely to be disease causing. There is also no evidence of further mutations in *MITF* causing melanoma predisposition since the two observed variants, p.R72H and p.T91M, do not occur in the melanocyte specific MITF-M isoform. Overall, candidate gene analysis showed no evidence that germline *SOX10* mutations contribute to melanoma susceptibility and did not find evidence of other relevant variants in *MITF* predisposing to CMM.

Table 5: Primers for Sanger sequencing of *SOX10* and *MITF*

Amplicon	Primer sequence
SOX10 exon 1 forward	GTGGGCGTTGGACTCTTTG
SOX10 exon 1 reverse	CCTCTAGCTTCGGGTGGATT
SOX10 exon 2 forward	TCTGAGGCTGTTATTCCTTGG
SOX10 exon 2 reverse	GTCAGTGGAGACAGG
SOX10 exon 3.1 forward	CAACAAGAGCGAAACTCCATC
SOX10 exon 3.1 reverse	AAAGCCCAGGTGAAGACAGAG
SOX10 exon 3.2 forward	ACCTTTGATGTGGCTGAGTTG
SOX10 exon 3.2 reverse	CTGTGTGCCCTGTTCCTTG
MITF-M exon 1 forward	TTATAGAAAGTAGAGGGAGGGATAGTC
MITF-M exon 1 reverse	GCTGTTTATTATTTGATGCCATAAG
MITF-M exon 2 forward	TTTGTGCCTGAAGGAAGAGC
MITF-M exon 2 reverse	GAGGTGTGGGACATGCTGTT
MITF-M exon 3 forward	GCCATCAGCTTTGTGTGAAC
MITF-M exon 3 reverse	GGATCACACCTTCCTGAAAAC
MITF-M exon 4 forward	GTGGAAAGAGGACAGTTACTTCTTAG
MITF-M exon 4 reverse	CTTGAGGTGTTTCCAGGGTT
MITF-M exon 5 forward	TCAAAGGGAACTGGTTGAGG
MITF-M exon 5 reverse	CTGTCCATTTCCTGATAAGACAAA
MITF-M exon 6 forward	GCTTCTGTATGTTTGGGAAATG
MITF-M exon 6 reverse	GGAGAGTTGATTCCTACAGCTG
MITF-M exon 7 forward	GCTAAATGCATACATGGCACTG
MITF-M exon 7 reverse	GTAAAGAAGTCTCCCCTCTCCC
MITF-M exon 8 forward	TCCATGTAACCAAGCACCAC
MITF-M exon 8 reverse	GGATATAATGAGCCATAGGGGA
MITF-M exon 9 forward	GGCTTAAAAGTCCTCTGTGCTC
MITF-M exon 9 reverse	CTTACCTGAAGGGGTTTTCTTG

Table 6: ***SOX10*** **variants**

Nucleotide change	Amino acid	rs ID	Proportion of probands with variant	MAF in probands	MAF*
c.18C>T	p.D6D	rs149435516	6/90	0.0333	0.031
c.825C>T	p.N275N	rs200437243	1/92	0.0054	na
c.927T>C	p.H309H	rs139884	het- 45/92; hom- 35/92	T- 0.375; C- 0.625	0.366

na is not available

* minor allele frequency reported by ESP6500

Table 7: ***MITF*** **variants**

Nucleotide change	Amino acid	Transcript	rs ID	Frequency in cases^	MAF*	Comment
c.215G>A	p.R72H	NM_006722	na	1/1690	na	not isoform M
c.272C>T	p.T91M	NM_197178	rs201247895	2/1690	0.0116	not isoform M
c.952G>A	p.E318K	NM_000248	rs149617956	not typed	0.314	already published

na is not available

^ refers to Australian case-control sample

* minor allele frequency reported by ESP6500

3.2 *PALB2*

Manuscript 2

Lauren G. Aoude, Mai Xu, Zhen Zhen Zhao, Michael Kovacs, Jane M. Palmer, Peter Johansson, Judith Symmons, Jeffrey M. Trent, Nicholas G. Martin, Grant W. Montgomery, Kevin M. Brown and Nicholas K. Hayward. Assessment of *PALB2* as a candidate melanoma susceptibility gene. *PLOS One*. 2014 June; 9(6):e100683.

3.2.1 Relevance to thesis aims

The final gene that was selected through the candidate gene approach was partner and localizer of BRCA2 (*PALB2*). The selection process was based on evidence from the literature that the PALB2 protein interacts with BRCA2 to enable double-strand break repair through homologous recombination. The BRCA2 interaction makes *PALB2* a good candidate melanoma susceptibility gene as it has been documented that aside from breast cancer, germline mutations in *BRCA2* can lead to CMM and other cancer types, including those of the ovary, stomach and prostate. Anecdotal evidence has shown that *PALB2* mutation carriers have presented with melanoma, though predominantly breast and pancreatic cancers have been reported.

The aim of this study was to assess *PALB2* as a candidate melanoma predisposition gene. In doing so I sought to determine whether *PALB2* functions as a general tumour suppressor, for which melanoma is part of the disease spectrum, or whether certain high-penetrance germline mutations of *PALB2* are responsible for melanoma development in some high-density melanoma families. This addresses Aim 2 of this study by searching for novel melanoma predisposition genes in high-density melanoma families from Queensland.

Here we report a germline missense mutation in *PALB2* (p.V78I) that incompletely segregates with disease in a family with cutaneous melanoma and breast cancer. We also found a known deleterious *PALB2* mutation (rs118203998) causing a premature truncation of the protein (p.Y1183X) in an individual with four different cancer types, including melanoma. Neither of these variants was found in 1,630 melanoma probands analysed as part of an Australian case-control set. Overall, the impact of these extremely rare mutations on melanoma susceptibility is uncertain and larger studies are needed to assess families with these and potentially other rare *PALB2* mutations.

3.2.2 Contribution of candidate

I performed the analysis of the exome sequencing data and Sanger sequencing validation of variants that were part of this data set. I also performed all of the co-segregation analysis to determine whether variants segregated with disease in the given families. I performed the case-control analysis on 1,630 cases and 1,690 controls to determine the contribution of the non-synonymous variants on melanoma predisposition. I performed the quantitative PCR to assess copy number variation observed in exome data. I wrote and compiled the manuscript under the guidance of my principal supervisor, Nicholas Hayward.

3.2.3 Contribution of other authors

Sanger sequencing of the protein coding region of *PALB2* in melanoma probands was performed by Mai Xu and Michael Kovacs. Zhen Zhen Zhao performed the case-control analysis of two of the missense variants. Jane Palmer and Judith Symmons collected the patient data relating to the families described in the manuscript. Peter Johansson performed the copy number analysis using the whole-genome and exome data. Kevin Brown and Jeffery Trent helped design the experiments, Nicholas Martin and Grant Montgomery contributed samples to this study. Nicholas Hayward helped design the experiments and edited the manuscript. All authors read and approved of the manuscript.

OPEN ACCESS Freely available online

PLOS ONE

Assessment of *PALB2* as a Candidate Melanoma Susceptibility Gene

Lauren G. Aoude[1,2]*, Mai Xu[3], Zhen Zhen Zhao[1], Michael Kovacs[3], Jane M. Palmer[1], Peter Johansson[1], Judith Symmons[1], Jeffrey M. Trent[4], Nicholas G. Martin[1], Grant W. Montgomery[1], Kevin M. Brown[3,4], Nicholas K. Hayward[1]

1 QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia, **2** University of Queensland, Brisbane, QLD, Australia, **3** National Cancer Institute, Bethesda, Maryland, United States of America, **4** Translational Genomics Research Institute, Phoenix, Arizona, United States of America

Abstract

Partner and localizer of BRCA2 (PALB2) interacts with BRCA2 to enable double strand break repair through homologous recombination. Similar to *BRCA2*, germline mutations in *PALB2* have been shown to predispose to Fanconi anaemia as well as pancreatic and breast cancer. The PALB2/BRCA2 protein interaction, as well as the increased melanoma risk observed in families harbouring *BRCA2* mutations, makes *PALB2* a candidate for melanoma susceptibility. In order to assess *PALB2* as a melanoma predisposition gene, we sequenced the entire protein-coding sequence of *PALB2* in probands from 182 melanoma families lacking pathogenic mutations in known high penetrance melanoma susceptibility genes: *CDKN2A, CDK4,* and *BAP1*. In addition, we interrogated whole-genome and exome data from another 19 kindreds with a strong family history of melanoma for deleterious mutations in *PALB2*. Here we report a rare known deleterious *PALB2* mutation (rs118203998) causing a premature truncation of the protein (p.Y1183X) in an individual who had developed four different cancer types, including melanoma. Three other family members affected with melanoma did not carry the variant. Overall our data do not support a case for *PALB2* being associated with melanoma predisposition.

Citation: Aoude LG, Xu M, Zhao ZZ, Kovacs M, Palmer JM, et al. (2014) Assessment of *PALB2* as a Candidate Melanoma Susceptibility Gene. PLoS ONE 9(6): e100683. doi:10.1371/journal.pone.0100683

Editor: Alvaro Galli, CNR, Italy

Received March 19, 2014; **Accepted** May 27, 2014; **Published** June 20, 2014

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. Whole-genome and exome sequencing data are unsuitable for deposit. All patients have consented to genetic analysis but have not specifically consented to having their data made publically available. Whole-genome and exome sequencing data may be made available for research purposes upon request to the authors.

Funding: This work was supported by the Melanoma Research Alliance and the National Health and Medical Research Council (NHMRC) of Australia. NKH and GWM receive salary support through a NHMRC Senior Principal Research Fellowship. LGA receives salary support by the Australia and New Zealand Banking Group Limited Trustees PhD scholarship. KMB is supported by the Division of Cancer Epidemiology and Genetics, National Cancer Institute, USA. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: LGA receives salary support by the Australia and New Zealand Banking Group Limited Trustees PhD scholarship. There are no patents, products in development or marketed products to declare. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials, as detailed online in the guide for authors.

* Email: Lauren.Aoude@qimrberghofer.edu.au

Introduction

Familial melanoma represents approximately 5–10% of all cutaneous malignant melanoma (CMM) cases and it is estimated that approximately 40% of familial cases can be attributed to known high penetrance genes. The cause of increased risk of melanoma development in the remaining families is largely unknown but might be due to a combination of low or medium penetrance gene mutations or to rare high penetrance mutations that are as yet undiscovered. A plausible candidate that might contribute to the melanoma risk landscape in certain individuals is partner and localizer of *BRCA2* (*PALB2*). PALB2 has a role in DNA repair, which it does by binding to BRCA1 and BRCA2 to facilitate homologous recombination for repair of double-strand breaks. It also has a role in facilitating a DNA checkpoint response and disruption to this complex can lead to instability of the DNA damage response pathway [1]. Germline mutations in its interacting partner BRCA2 are known to predispose to cancers of the breast, ovary and pancreas, as well as confer a moderate increased risk of CMM (relative risk of 2.6 [2,3]), suggesting by association that PALB2 may also play a role in tumour development in these cancer types.

There have been several diseases attributed to germline mutation in *PALB2*, the first reports were in Fanconi anemia (FA) patients with the FA-N subtype [4,5]. Carriers were affected by bi-allelic mutation in *PALB2* that led to the onset of early childhood cancers, along with other FA disease traits including growth retardation and congenital malformation. Looking at the family histories of these cases, 4/8 families had a history of breast cancer, suggesting that like *BRCA2*, mono-allelic mutations in *PALB2* may lead to the development of breast cancer.

From this observation, germline *PALB2* mutations were subsequently shown to be associated with breast cancer risk [6,7]. Rahman and colleagues investigated a cohort of 923 breast cancer cases of which 10 individuals presented with mono-allelic truncating *PALB2* mutations, while a set of 1,084 controls harboured no truncating mutations. Segregation analysis showed that in approximately half of these families the variants did not completely co-segregate with disease, suggesting that *PALB2* acts as a medium-penetrance rather than high-penetrance gene for

breast cancer risk [6]. A study published at the same time by Erkko et al., reported a founder mutation, c.1592delT, present in 0.2% of the Finnish population, but 1% of breast cancer cases unselected for family history and 2.7% of families with multiple cases of breast and/or ovarian cancer [7]. This shows a clear role for *PALB2* in breast cancer development. Since these initial reports, there has been a wide array of population-based studies describing the prevalence of *PALB2* in breast cancer cases with incidence ranging from 0.5% to 2.6% [8–18].

An association between germline *PALB2* mutation and increased pancreatic cancer risk has also been established. After identifying bi-allelic inactivation of *PALB2* in a tumour from a familial pancreatic cancer patient, Jones and colleagues investigated germline *PALB2* variation in 96 probands with a family history of pancreatic cancer and identified truncating mutations in three individuals [19]. In another study, the screening of 254 pancreatic cancer cases, including 101 with a family history of the disease, led to the report of a 6.7 kb deletion of *PALB2* in an individual with both breast and pancreatic cancer. This was the only conclusively inactivating mutation found in this sample set [20]. A study looking at European familial pancreatic cancer discovered truncating mutations in 3/81 cases, all of which were from families with cases of breast cancer, suggesting that these mutations may preferentially occur in families with both of these cancer types [21]. Consistent with this, another study reported that 4.8% of families with both cancer types harboured truncating *PALB2* mutations [22], higher than the rate previously reported in studies investigating breast cancer alone. Although a study that followed on from this looked at 77 families with breast cancer, pancreatic cancer, or a combination of both, but found no mutations [23]. Because *PALB2* germline truncating mutations are relatively rare, the full spectrum of cancer predisposition associated with these mutations has yet to be fully characterized. This is clear from the differing reports relating PALB2 to breast and pancreatic cancer development.

Multiple studies have reported melanoma in families harbouring inactivating *PALB2* mutations, including individuals with diagnoses of both melanoma and breast or pancreatic cancer [6] [24]. These data implicate *PALB2* as a possible melanoma susceptibility gene, although Sabbaghian and colleagues found no association between germline *PALB2* mutation and CMM risk in a screen of 53 probands from multi-case *CDNK2*A mutation-negative melanoma families, and Yang failed to identify *PALB2* mutations in 23 CMM families that are *CDKN2A* mutation-positive and contain a subset of families (n = 11) with pancreatic cancer also [25]. Given the small size of these studies, the association between *PALB2* and melanoma risk remains unclear. We therefore sought to determine the incidence of germline *PALB2* mutations in a larger series of 201 *CDKN2A* and *CDK4* mutation-negative melanoma families, including 63 with confirmed cases of breast, pancreatic, or multiple other types of cancer.

Methods

Ethics

Written consent was obtained from each participant in this study. Ethics approval was obtained from the QIMR Berghofer Human Research Ethics Committee (HREC).

Sample collection

Samples were ascertained as part of the Queensland Familial Melanoma Project (QFMP), a population based study of melanoma in Queensland, Australia [26]. Genomic DNA was extracted from whole blood using standard salting out methods. In some instances DNA was extracted from transformed lymphoblastoid cell lines.

Samples

Selection criteria for inclusion of families were those with: CMM plus breast and/or pancreatic cancer (n = 52); individuals who had developed three or more different cancer types, where CMM was one of the cancers (n = 3); a minimum of three CMM cases (n = 127). A total of 182 families met these criteria. No additional criteria relating to age of onset of the cancers, or degree of relationship between affected members were imposed. All families have previously been shown to lack pathogenic mutations in known high risk melanoma susceptibility genes, *CDKN2A*, *CDK4* and *BAP1* [*27,28*]. In the instance where multiple DNA samples were available for sequencing, the youngest available CMM case was chosen as the proband for each family. Where affected individuals within a family were of a comparable age and an individual presented with multiple primary melanomas, they were then selected as the proband.

In addition to this, whole-genome and whole-exome sequencing data from some QFMP families (n = 19) was interrogated. These families were selected for sequencing as they had a minimum of three affected members. Within this group, 16/19 families had cases with multiple primary melanoma, and 8/19 families had occurrences of breast cancer. Six families also included individuals that had developed three different cancer types. A total of 24 exomes and 15 genomes were interrogated, with up to 3 cases sequenced within a family. None of the cases sequenced carried deleterious variants in breast cancer susceptibility genes *BRCA1* or *BRCA2*.

Sanger sequencing

In order to look for pathogenic mutations in *PALB2*, Sanger sequencing was used to screen the 13 exons as well the exon boundaries of the 182 selected probands for protein altering variants. The complete list of M13-tagged primers and sequencing methods can be found in Methods S1.

Whole-genome and exome sequencing

Whole-genome sequencing for 12 samples and exome sequencing for 25 samples was performed using the Illumina Hiseq 2000 platform combined with the Agilent SureSelect Human All Exon V4+UTRs enrichment kits (Methods S1). 100 bp paired-end reads were generated with samples having a mean coverage of 96X. A further three genomes were sequenced by Complete Genomics. Using the BWA alignment algorithm, the sequence output was mapped to the UCSC human genome reference build 19 [29]. SNPs and indels were detected using bcftools and SAMtools mpileup with disabled BAQ computation [30]. Each sample had on average 90,000–100,000 variants compared to the human genome reference sequence. Variants were filtered for stringency using a quality score (>40), alternate reads (>2 and >20% of all reads at a given position). Variants from next-generation sequencing data sets were validated using Sanger sequencing methods. Whole-genome and exome sequencing data may be made available for research purposes upon request to the authors.

Copy number analysis at the *PALB2* locus

To assess the possibility that some probands may have partial or complete gene deletions of *PALB2* we interrogated the whole-genome or exome data from those patients where data was available. Briefly, to normalise for different coverage per sample,

Table 1. *PALB2* variants detected through Sanger sequencing and exome sequencing.

Location	protein change	nucleotide change	rs ID	freq. of genotype in CMM probands (n = 201)	MAF in CMM probands	MAF in ESP6500* (n = 4300)	SIFT prediction	protein domain function	reference
chr16:23647635	p.V78I	c.232G>A	–	1/201	0.002	na	tolerated	interacts with BRCA1 & RAD51; required for oligomerization	[33]
chr16:23646857	p.L337S	c.1010T>C	rs45494092	8/201	0.020	0.020	tolerated		[17,18,20,24,33,35,36]
chr16:23646673	p.V398V	c.1194G>A	rs61755173	2/201	0.005	0.002	na		[12]
chr16:23646295	p.S524S	c.1572A>G	rs45472400	3/201	0.007	0.005	na		[12,18]
chr16:23646191	p.Q559R	c.1676A>G	rs152451	33/201	0.082	0.091	tolerated		[17,18,20,24,33,35–37]
chr16:23641461	p.E672Q	c.2014G>C	rs45532440	13/201	0.032	0.031	tolerated		[17,18,20,24,33,35,36]
chr16:23637715	p.P864S	c.2590C>T	rs45568339	3/201	0.007	0.003	tolerated	WD1; interacts with BRCA2 & RAD51	[17,18,33,35,38]
chr16:23635370	p.V932M	c.2794G>A	rs45624036	2/201	0.005	0.005	tolerated	WD2; interacts with BRCA2 & RAD51	[20,33,35]
chr16:23635348	p.L939W	c.2816T>G	rs45478192	2/201	0.005	0.002	damaging	WD2; interacts with BRCA2 & RAD51	[33,35]
chr16:23634293	p.G998E	c.2993G>A	rs45551636	12/201	0.030	0.023	damaging	WD3; interacts with BRCA2 & RAD51	[17,18,20,24,33,35,36]
chr1 6:23619235	p.T1100T	c.3300T>G	rs45516100	11/201	0.027	0.031	na	WD4; interacts with BRCA2 & RAD51	[18,20,35,36]
chr16:23614846	p.S1165S	c.3495G>A	rs45439097	1/201	0.002	0.001	na	WD7; interacts with BRCA2 & RAD51	[35]
chr16:23614792	p.Y1183X	c.3549C>G	rs118203998	1/201	0.002	na	na	WD7; interacts with BRCA2 & RAD51	[5,6,8,34]

*European American population.
na is not available.
doi:10.1371/journal.pone.0100683.t001

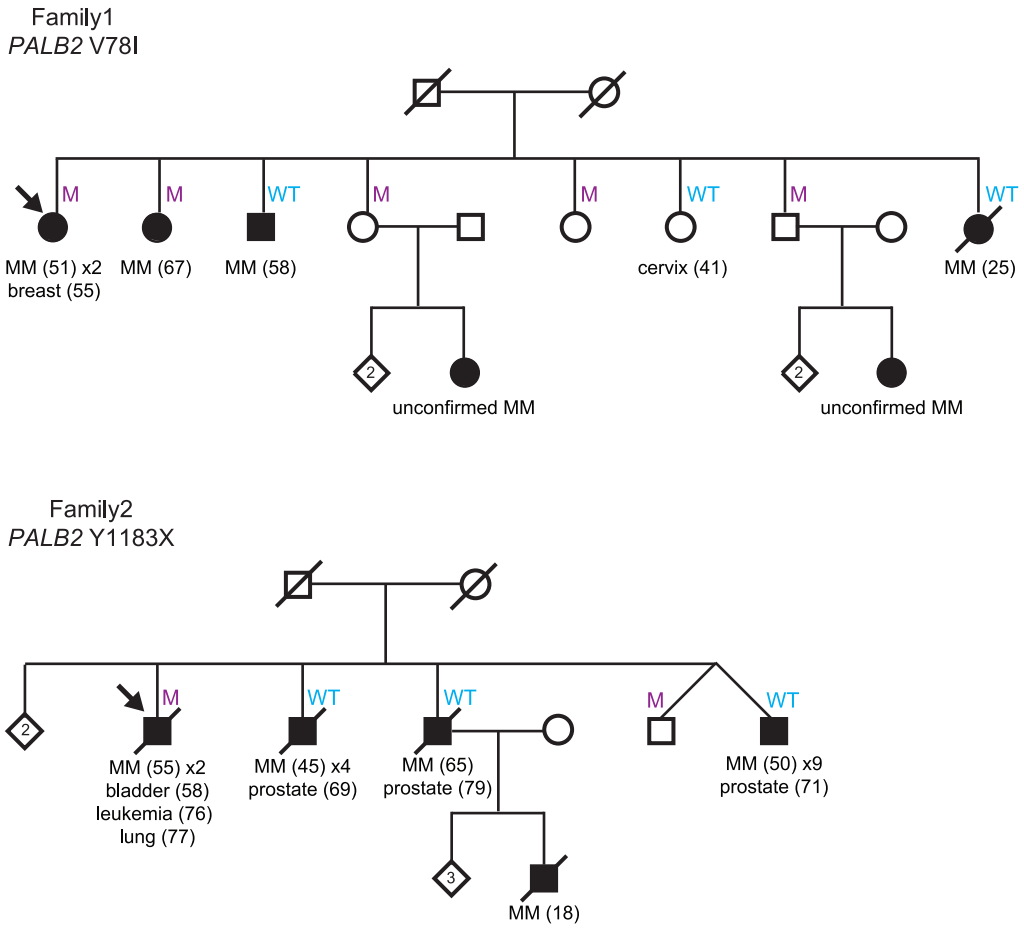


Figure 1. Co-segregation analysis of *PALB2* variants in two high-risk CMM families. Individuals that have melanoma (MM) are represented by black circles (female) and black boxes (male). The age of diagnosis of each cancer is indicated in brackets. A line through a symbol indicates that the person is deceased. Individuals carrying a *PALB2* mutation are indicated by an 'M', while those wild-type for the variant are indicated by 'WT'. Other cancer types are also indicated on the pedigree. Unaffected siblings are represented by a diamond with the number indicating the number of siblings. The arrow indicates the proband in each family.
doi:10.1371/journal.pone.0100683.g001

the number of reads for each exon was divided by the number of reads across the entire sample. The median value for each exon was then calculated across all samples. This value was used to estimate copy-number variation per exon of *PALB2*. In instances where a copy number value for a given exon in an individual was estimated to be less than half of the median value across all patients, quantitative PCR using SYBR Green and exon-specific primers was then used to assess the validity of the bioinformatics output.

Case-control analysis

All protein-changing variants that were identified by Sanger sequencing and exome sequencing were analysed in a case-control set derived from two Australian studies to determine whether they might be low to medium penetrance CMM predisposition variants. A Sequenom iPLEX was run on 3320 individuals which included 1,690 probands derived from the QFMP [26]. These individuals were sampled from Queensland, Australia and include both cases with a first degree relative with CMM (n = 1551); and also sporadic cases with no family history of melanoma (n = 139). The cases included individuals with a wide spectrum of age of disease onset that ranges from childhood disease to late onset melanoma.

The control sample (n = 1630) were parents of twins ascertained as part of the Brisbane Twin Naevus Study [31]. They were asked to self report and had no history of melanoma at the time of sample collection.

In order to analyse the significance of observed protein altering variants found through sequencing methods, a chi-squared test was used to compare the cases and controls.

Results

We sequenced the entire protein coding sequence of *PALB2* in the panel of CMM cases described above. We identified eight missense variants, four synonymous variants and one nonsense variant. Of these, 11 were present in NHLBI Exome Sequencing Project (ESP6500) [32], dbSNP or the 1000 Genomes Project (see Table 1 for a full list of variants). No proband for which whole-genome (n = 12) or exome (n = 25) sequence data were available was found to carry a partial or complete deletion of *PALB2*.

A missense variant, p.V78I (c.G432A, NCBI accession NM_024675), not listed in dbSNP but reported previously in a study of families with breast/ovarian and pancreatic cancers [33], was identified via Sanger sequencing and found in a female who presented with two primary CMM, at ages 51 and 71, and breast cancer at age 55. Co-segregation analysis showed incomplete

Table 2. Case-control analysis on non-synonymous *PALB2* variants.

location	protein change	nucleotide change	rs ID	Frequency in cases (n = 1690)		Frequency in controls (n = 1630)		chi-square	p-value
				het carriers	hom carriers	het carriers	hom carriers		
Exon 4	p.V78I	c.232G>A	–	1	0	0	0	–	–
Exon 4	p.L337S	c. 1010T>C	rs45494092	79	2	71	0	0.565	0.452
Exon 4	p.Q559R	c. 1676A>G	rs152451	289	12	243	10	2.981	0.084
Exon 5	p.E672Q	c. 2014G>C	rs45532440	105	0	94	1	0.148	0.700
Exon 7	p.P864S	c. 2590C>T	rs45568339	8	0	18	0	4.202	0.040
Exon 8	p.V932M	c. 2794G>A	rs45624036	13	0	13	0	0.009	0.924
Exon 8	p.L939W	c. 2816T>G	rs45478192	6	0	9	0	0.715	0.398
Exon 9	p.G998E	c. 2993G>A	rs45551636	73	0	66	0	0.148	0.700
Exon 13	p.Y1183X	c. 3549C>G	rs118203998	1	0	0	0	–	–

doi:10.1371/journal.pone.0100683.t002

segregation with disease in the remainder of the family (Figure 1). Only two of the four individuals affected with CMM in this family carry *PALB2* p.V78I, the second carrier being a sibling who had developed CMM at age 67. A protein-truncating variant, p.Y1183X, was found in a five case CMM family that was analysed via exome sequencing. This variant, rs118203998, was originally identified in a breast cancer family by Rahman and colleagues [6] and has since been reported by several other groups in patients diagnosed with breast cancer, pancreatic cancer, and Fanconi anaemia [5,6,8,34]. The individual carrying this mutation in our study has been diagnosed with four different primary cancer types, including melanoma (twice, initially diagnosed at age 55), bladder cancer (at age 58), leukaemia (at age 76), and non-small cell lung cancer (at age 77). This variant did not co-segregate with melanoma in this family as no other affected family member was found to be a carrier.

Genotyping a collection of 1,690 CMM probands, as well as 1630 Australian population controls, showed that *PALB2* p.Y1183X was not observed in any other individual. This is consistent with results found by Rahman [6]. This variant is also not seen in the 1000 Genomes Project, nor is it seen in any population of the ESP6500, and is listed in dbSNP as having unknown population frequency.

The p.V78I was observed in two controls. Personal history of non-melanoma cancers was not collected at the time of the study setup. It is unclear whether these individuals may have been affected by any other form cancer. The other seven rare missense variants we identified in the family collection were found at low frequency in the case-control panel (Table 2). None were significantly over-represented in melanoma cases.

Discussion

To assess the contribution of *PALB2* to melanoma predisposition we sequenced the protein-coding region of *PALB2* in probands from 201 melanoma families lacking pathogenic mutations in know melanoma susceptibility genes. This is the largest study reported to date to assess the relationship between germline *PALB2* mutation and melanoma risk. We have identified a missense mutation (chr16:23647635, p.V78I) that incompletely segregates with disease in a family with cutaneous melanoma and breast cancer. We have also identified a further 7 previously reported missense variants for which the PALB2 protein function is undetermined. Genotyping of these variants did not reveal any significant differences in allele frequency between cases and controls, with the exception of one variant which occurred more frequently in controls (rs45568339). These data do not support a role for these rare PALB2 variants in melanoma susceptibility.

We also found a known deleterious mutation (rs118203998) causing a premature truncation of the protein (Y1183X) in an individual with four different cancer types, including melanoma. Neither this nor p.V78I variant has a population frequency reported in the 1000 Genomes Project or the ESP6500. Interestingly, the truncating mutation, p.Y1183X, has been reported by four previous studies. A report discusses three children with Fanconi anaemia that carried a *PALB2* p.Y1183X mutation. Each of these cases presented with an early childhood cancer at ages 0.7, 1.0 and 2.3 years with neuroblastoma, Wilms' tumour and medulloblastoma respectively [5]. Rahman and colleagues also reported this mutation in three individuals with breast cancer [6]. Interestingly, in one of the families an individual with this genotype was diagnosed with melanoma at age 47, prior to developing breast cancer. In a third study, *PALB2* p.Y1183X was associated with breast cancer in a person presenting with two

primary breast cancers. The mother of this proband was also a mutation carrier and had developed both breast cancer and pancreatic cancer [34]. Pancreatic and breast cancers have been previously associated with a *PALB2* mutation but this study was the first instance of it occurring in an individual with both cancer types. Lastly a population-based study of breast cancer predisposition by Tischowitz et al. found a family in which three cases had this genotype [8]. In our study we present an individual with two primary melanomas, bladder cancer, leukaemia and non-small cell lung cancer who is a carrier of the *PALB2* p.Y1183X variant. Taking into consideration that this specific mutation has been seen in one other individual with melanoma, this suggests that *PALB2* may play a role as a melanoma susceptibility gene and that this mutation could predispose to other cancers whose spectrum of disease extends beyond Fanconi anaemia, breast cancer or pancreatic cancer.

When we consider truncating mutations alone, as Rahman and colleagues did in relation to breast cancer susceptibility [6], the data overall are not strongly supportive of *PALB2* being a melanoma susceptibility gene. Larger sample sizes will clearly be needed to better characterize the spectrum of cancers associated with deleterious *PALB2* mutations along with a larger sample of melanoma-only families to unambiguously determine the role of *PALB2* in melanoma susceptibility.

Supporting Information

Methods S1 Additional information on whole-genome and exome sequencing, Sanger sequencing, and iPLEX methods.
(DOCX)

Acknowledgments

The Authors would like to thank the families for their participation in the study.

Author Contributions

Conceived and designed the experiments: LGA NKH. Performed the experiments: LGA MX ZZZ MK. Analyzed the data: LGA MX ZZZ PJ. Contributed reagents/materials/analysis tools: NKH KMB GWM NGM JMT. Contributed to the writing of the manuscript: LGA NKH. Contacted families and obtained consent: JMP JS.

References

1. Xia B, Sheng Q, Nakanishi K, Ohashi A, Wu J, et al. (2006) Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2. Mol Cell 22: 719–729.
2. Moran A, O'Hara C, Khan S, Shack L, Woodward E, et al. (2012) Risk of cancer other than breast or ovarian in individuals with BRCA1 and BRCA2 mutations. Fam Cancer 11: 235–242.
3. Breast Cancer Linkage C (1999) Cancer risks in BRCA2 mutation carriers. J Natl Cancer Inst 91: 1310–1316.
4. Xia B, Dorsman JC, Ameziane N, de Vries Y, Rooimans MA, et al. (2007) Fanconi anemia is associated with a defect in the BRCA2 partner PALB2. Nat Genet 39: 159–161.
5. Reid S, Schindler D, Hanenberg H, Barker K, Hanks S, et al. (2007) Biallelic mutations in PALB2 cause Fanconi anemia subtype FA-N and predispose to childhood cancer. Nat Genet 39: 162–164.
6. Rahman N, Seal S, Thompson D, Kelly P, Renwick A, et al. (2007) PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. Nat Genet 39: 165–167.
7. Erkko H, Xia B, Nikkila J, Schleutker J, Syrjakoski K, et al. (2007) A recurrent mutation in PALB2 in Finnish cancer families. Nature 446: 316–319.
8. Tischkowitz M, Capanu M, Sabbaghian N, Li L, Liang X, et al. (2012) Rare germline mutations in PALB2 and breast cancer risk: A population-based study. Hum Mutat.
9. Foulkes WD, Ghadirian P, Akbari MR, Hamel N, Giroux S, et al. (2007) Identification of a novel truncating PALB2 mutation and analysis of its contribution to early-onset breast cancer in French-Canadian women. Breast Cancer Res 9: R83.
10. Cao AY, Huang J, Hu Z, Li WF, Ma ZL, et al. (2009) The prevalence of PALB2 germline mutations in BRCA1/BRCA2 negative Chinese women with early onset breast cancer or affected relatives. Breast Cancer Res Treat 114: 457–462.
11. Garcia MJ, Fernandez V, Osorio A, Barroso A, Llort G, et al. (2009) Analysis of FANCB and FANCN/PALB2 fanconi anemia genes in BRCA1/2-negative Spanish breast cancer families. Breast Cancer Res Treat 113: 545–551.
12. Sluiter M, Mew S, van Rensburg EJ (2009) PALB2 sequence variants in young South African breast cancer patients. Fam Cancer 8: 347–353.
13. Papi L, Putignano AL, Congregati C, Piaceri I, Zanna I, et al. (2010) A PALB2 germline mutation associated with hereditary breast cancer in Italy. Fam Cancer 9: 181–185.
14. Dansonka-Mieszkowska A, Kluska A, Moes J, Dabrowska M, Nowakowska D, et al. (2010) A novel germline PALB2 deletion in Polish breast and ovarian cancer patients. BMC Med Genet 11: 20.
15. Casadei S, Norquist BM, Walsh T, Stray S, Mandell JB, et al. (2011) Contribution of inherited mutations in the BRCA2-interacting protein PALB2 to familial breast cancer. Cancer Res 71: 2222–2229.
16. Ghadirian P, Robidoux A, Zhang P, Royer R, Akbari M, et al. (2009) The contribution of founder mutations to early-onset breast cancer in French-Canadian women. Clin Genet 76: 421–426.
17. Zheng Y, Zhang J, Niu Q, Huo D, Olopade OI (2012) Novel germline PALB2 truncating mutations in African American breast cancer patients. Cancer 118: 1362–1370.
18. Teo ZL, Park DJ, Provenzano E, Chatfield CA, Odefrey FA, et al. (2013) Prevalence of PALB2 mutations in Australasian multiple-case breast cancer families. Breast Cancer Res 15: R17.
19. Jones S, Hruban RH, Kamiyama M, Borges M, Zhang X, et al. (2009) Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. Science 324: 217.
20. Tischkowitz MD, Sabbaghian N, Hamel N, Borgida A, Rosner C, et al. (2009) Analysis of the gene coding for the BRCA2-interacting protein PALB2 in familial and sporadic pancreatic cancer. Gastroenterology 137: 1183–1186.
21. Slater EP, Langer P, Niemczyk E, Strauch K, Butler J, et al. (2010) PALB2 mutations in European familial pancreatic cancer families. Clin Genet 78: 490–494.
22. Peterlongo P, Catucci I, Pasquini G, Verderio P, Peissel B, et al. (2011) PALB2 germline mutations in familial breast cancer cases with personal and family history of pancreatic cancer. Breast Cancer Res Treat 126: 825–828.
23. Stadler ZK, Salo-Mullen E, Sabbaghian N, Simon JA, Zhang L, et al. (2011) Germline PALB2 mutation analysis in breast-pancreas cancer families. J Med Genet 48: 523–525.
24. Sabbaghian N, Kyle R, Hao A, Hogg D, Tischkowitz M (2011) Mutation analysis of the PALB2 cancer predisposition gene in familial melanoma. Fam Cancer 10: 315–317.
25. Yang XR, Jessop L, Myers T, Amundadottir L, Pfeiffer RM, et al. (2011) Lack of germline PALB2 mutations in melanoma-prone families with CDKN2A mutations and pancreatic cancer. Fam Cancer 10: 545–548.
26. Aitken JF, Green AC, MacLennan R, Youl P, Martin NG (1996) The Queensland Familial Melanoma Project: study design and characteristics of participants. Melanoma Res 6: 155–165.
27. Wiesner T, Fried I, Ulz P, Stacher E, Popper H, et al. (2012) Toward an improved definition of the tumor spectrum associated with BAP1 germline mutations. J Clin Oncol 30: e337–340.
28. Goldstein AM, Chan M, Harland M, Gillanders EM, Hayward NK, et al. (2006) High-risk melanoma susceptibility genes and pancreatic cancer, neural system tumors, and uveal melanoma across GenoMEL. Cancer Res 66: 9818–9828.
29. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754–1760.
30. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079.
31. McGregor B, Pfitzner J, Zhu G, Grace M, Eldridge A, et al. (1999) Genetic and environmental contributions to size, color, shape, and other characteristics of melanocytic naevi in a sample of adolescent twins. Genet Epidemiol 16: 40–53.
32. Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA (URL: http://evs.gs.washington.edu/EVS/) Accessed 27 February 2014.
33. Blanco A, de la Hoya M, Osorio A, Diez O, Miramar MD, et al. (2013) Analysis of PALB2 gene in BRCA1/BRCA2 negative Spanish hereditary breast/ovarian cancer families with pancreatic cancer cases. PLoS One 8: e67538.
34. Hofstatter EW, Domchek SM, Miron A, Garber J, Wang M, et al. (2011) PALB2 mutations in familial breast and pancreatic cancer. Fam Cancer 10: 225–231.
35. Hellebrand H, Sutter C, Honisch E, Gross E, Wappenschmidt B, et al. (2011) Germline mutations in the PALB2 gene are population specific and occur with low frequencies in familial breast cancer. Hum Mutat 32: E2176–2188.
36. Catucci I, Milgrom R, Kushnir A, Laitman Y, Paluch-Shimon S, et al. (2012) Germline mutations in BRIP1 and PALB2 in Jewish high cancer risk families. Fam Cancer 11: 483–491.

37. Phuah SY, Lee SY, Kang P, Kang IN, Yoon SY, et al. (2013) Prevalence of PALB2 mutations in breast cancer patients in multi-ethnic Asian population in Malaysia and Singapore. PLoS One 8: e73638.

38. Tischkowitz M, Sabbaghian N, Ray AM, Lange EM, Foulkes WD, et al. (2008) Analysis of the gene coding for the BRCA2-interacting protein PALB2 in hereditary prostate cancer. Prostate 68: 675–678.

Supplementary Methods

Whole-genome and whole-exome Sequencing

Whole-genome (N=12) and exome sequencing (N=25) was outsourced to Axeq Technologies in Seoul, South Korea. Sequencing was performed on the Illumina HiSeq 2000 platform using the Agilent SureSelect Human All Exon V4+UTRs enrichment kits. The data output was 100bp paired end reads with a median coverage of 96x. A further 3 genomes were sequenced using Complete Genomics. The sequencing data for each samples was mapped to the Human Genome build 19 (hg19) using the BWA alignment algorithm [1]. SNPs and indels were detected using bcftools and SAMtools mpileup with disabled BAQ computation [2]. Any variant that was found in dbSNP or the 1000 Genomes project was flagged to generate both a data set of novel mutations and also a data set of known mutations. Each sample had on average 90,000-100,000 variants compared to the human genome reference sequence. Variants were filtered for stringency using a quality score of >40, >2 alternate reads and >20% of all reads at a given position being the alternate read.

Sanger Sequencing

Sanger sequencing was used to identify mutation events in *PALB2* in a cohort of familial melanoma cases. PCR reactions were run using ABI goldTaq PCR reagents. Standard manufacturer's protocol was followed. Refer to Supplementary table S1 for a list of the primers used.

Sequenom iPLEX

The *PALB2* variants were multiplexed using the Sequenom iPLEX gold system which allows up to 36 variants to be genotyped concurrently using 10ng of genomic DNA. The iPLEX Gold PCR amplification reactions were carried out in four 384-well plates according to supplier protocol and then transferred onto SpectroCHIP arrays using the MassARRAY Nanodispenser. The chips were then transferred to the Sequenom mass spectrometer where the data was generated. The results were analysed using the Typer Analyzer software 4.0.

References

1. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754-1760.

2. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079.

Supplementary Table S1: Primers used for Sanger sequencing of *PALB2*

Exon	Forward primer	Reverse primer
exon 1	AACTGGGTCCCGGTGTCG	GCCTAAAACCCTGGGAAAGC
exon 2 & 3	TGACTCCACCTTTCCACTTGC	AAGAACAATAGCCAAAATATACCTGGG
exon 4-1	ATTCATCTGCCTGAATGAAATG	TGCTACTGATTTCTTCCTGTTCC
exon 4-2	AGCTGCCAAGCAGAAGAAAG	ACCTTTAGGAGGAATGTGTTCAAG
exon 4-3	CACCAGGGCGACTACAGTTC	AAGAGGAGAGGTTGCTTCCAG
exon 4-4	CCCAGTGACACTCTTGATGG	AGGAAGTGCCAGGCAAATAG
exon 5-1	TTGTCATCAGTGAAACAGATTGTC	GAAAGGCCCGTCTTTGTATG
exon 5-2	AAACCAGTGGAGCCCTTTG	CACTTGCAGGGTGGTATGTG
exon 5-3	GACTCAGTCTGTCTTGCCAGTG	AAGCAAGTCATGCTGTTTACATTC
exon 6	AGTGGGTAATGCAGGCAGAC	GCCAATAGGTTGGCATAGAAAC
exon 7	CCACAAAGCTCTTTCTTTTCACC	GCCTTGCATGGTCATAGCTC
exon 8	GATTAAACAAAAATGAAACAACCAAGC	GGTTATTACCTGCACTTAAAACCAGC
exon 9	TAATATTAAAAGGTTACTCCTCACATCACC	ACAGAAAAACGAGATCCTAGTTACCC
exon 10	TCAACAATGCGGAGAAGGG	CTCTCTTATTTAATCTTCACAACAACCC
exon 11	GGCAAAATTAACCCACAGTTC	ACTGCTTATGACTTACTGCTCTCAC
exon 12	TCAGAGCCTATCGGTCATTGC	GCCTTTCAGAATGTCCCACC
exon 13-1	CAATAGCCAACAGACCTCTAAGGC	TCATTTTAAGTGTCATTCAGATATTCTCC
exon 13-2	AGGTTCCTGGAAGGTGACG	GCCATTTGAAGCTTTATGTACACC
M13 tag	TGTAAAACGACGGCCAGT	CAGGAAACAGCTATGACC

Chapter 4

POT1 loss-of-function variants predispose to familial melanoma

Manuscript 3

Carla Daniela Robles-Espinoza*, Mark Harland*, Andrew J. Ramsay*, **Lauren G. Aoude***, Víctor Quesada, Zhihao Ding, Antonia L. Pritchard, Jessamy C. Tiffen, Mia Petljak, Jane M. Palmer, Judith Symmons, Peter Johansson, Mitchell S. Stark, Michael G. Gartside, Helen Snowden, Grant W. Montgomery, Nicholas G. Martin, Jimmy Z. Liu, Jiyeon Choi, Matthew Makowski, Kevin M. Brown, Thomas M. Keane, Carlos López-Otín, Nelleke A. Gruis, Nicholas K. Hayward†, D. Timothy Bishop†, Julia A. Newton-Bishop†, David J. Adams†. *POT1* loss-of-function variants predispose to familial melanoma. *Nature Genetics* 2014 May; 46(5):478-81.

*joint first author; †joint last author

4.1 Relevance to thesis aims

This chapter describes an unbiased sequencing approach to search for novel melanoma predisposition genes in high-density CMM families. It uses whole-genome and exome data to identify novel variants that are segregating with disease in these families and identifies mutations in a novel gene that predisposes to CMM in high-density families. Unlike the candidate gene approach in Chapter 3, this method considers all variants in the data set in order to find genes that have not been previously associated with melanoma development but may play a role in melanomagenesis. Variants were selected if they segregated with disease in families, had an obvious deleterious effect on the protein or were mutated in several families. This approach has proven effective by uncovering mutations in the Protection of Telomeres 1 gene (*POT1*) in melanoma families. Given that this gene has been recently described to have a high frequency of somatic mutations in chronic lymphocytic leukaemia [129], it was a strong contender for a cancer predisposition gene.

This chapter addresses Aims 2 to 4 of my thesis, as NGS was used to identify loss-of-function mutations in *POT1*, occurring in high-density melanoma families from Australia and from the UK. Three missense mutations (p.Y89C, p.Q94E and p.R273L) occurred in the highly conserved oligonucleotide/oligosaccharide-binding (OB) domains of POT1, which facilitate the binding of POT1 to single-stranded telomeric repeats. Structural modelling predicted that the missense mutations would disrupt the binding of the OB-domains with telomeric DNA, which was confirmed

by an electromobility shift assay. A splice acceptor mutation (c.1687-1G>A) was also seen in the data set. RT-PCR showed that this resulted in the truncation of the protein 12 amino acids after the mutation site. Furthermore, experimental data has shown that carriers of *POT1* missense mutations have longer telomeres, which may be a major contribution to CMM development.

This study has shown that *POT1* is involved in a CMM predisposition in high-density families. Given that mutations in the *TERT* promoter have been shown in both tumours and sporadic CMM [130, 131], it demonstrates that telomere dysregulation is a newly found pathway contributing to CMM susceptibility. In addition, the pedigrees we describe are enriched for other cancers, suggesting that *POT1* mutations may predispose to cancers other than CMM.

4.2 Contribution of candidate

This was a large collaborative effort between international research groups and contained elements that were outside the scope of this thesis. My role was in the analysis of the whole-genome and exome data from Australia, Sweden and Denmark, as well as the validation of variants using Sanger sequencing methods. I carried out the RT-PCR experiments to determine the effect of the splice mutation on the POT1 protein. I was also responsible for the preparation of Australian samples including extraction of DNA and RNA from whole blood as well as the synthesis of cDNA. I wrote methods that pertained to my experiments and compiled the sequencing and histology data of Queensland CMM cases. I also compiled tables and edited the manuscript.

4.3 Contribution of other authors

Nicholas K. Hayward, David J. Adams, Carla Daniela Espinoza-Robles, Julia A. Newton Bishop, Timothy Bishop and Mark Harland designed the study and wrote the manuscript. Samples were provided by Nelleke A. Gruis, Grant W. Montgomery, Nicholas G. Martin and Helen Snowden. NGS data analysis and experiments were performed by Antonia L. Pritchard, Mitchell S. Stark, Michael G. Gartside, Jessamy C. Tiffen Mia Petljak, Jimmy Z. Liu, Jiyeon Choi, Matthew Makowski and Kevin M. Brown. Carla Daniela Espinoza-Robles performed NGS data analysis for the UK families and the controls. Peter Johansson, Zhihao Ding and Thomas M. Keane provided the bioinformatics support. Zhihao Ding performed the analysis of exome data to determine telomere length. Protein modelling and EMSAs were performed by Víctor Quesada, Andrew J. Ramsay and

Carlos López-Otín. Jane M. Palmer and Judith Symmons contacted the families and collected the patient data and samples for the Australian cases. All authors read and approved of the manuscript.

nature genetics

POT1 loss-of-function variants predispose to familial melanoma

Carla Daniela Robles-Espinoza[1,12], Mark Harland[2,12], Andrew J Ramsay[3,12], Lauren G Aoude[4,12], Víctor Quesada[3], Zhihao Ding[1], Karen A Pooley[5], Antonia L Pritchard[4], Jessamy C Tiffen[1], Mia Petljak[1], Jane M Palmer[4], Judith Symmons[4], Peter Johansson[4], Mitchell S Stark[4], Michael G Gartside[4], Helen Snowden[2], Grant W Montgomery[6], Nicholas G Martin[7], Jimmy Z Liu[8], Jiyeon Choi[9], Matthew Makowski[9], Kevin M Brown[9], Alison M Dunning[10], Thomas M Keane[1], Carlos López-Otín[3], Nelleke A Gruis[11], Nicholas K Hayward[4,13], D Timothy Bishop[2,13], Julia A Newton-Bishop[2,13] & David J Adams[1,13]

Deleterious germline variants in *CDKN2A* account for around 40% of familial melanoma cases[1], and rare variants in *CDK4*, *BRCA2*, *BAP1* and the promoter of *TERT* have also been linked to the disease[2–5]. Here we set out to identify new high-penetrance susceptibility genes by sequencing 184 melanoma cases from 105 pedigrees recruited in the UK, The Netherlands and Australia that were negative for variants in known predisposition genes. We identified families where melanoma cosegregates with loss-of-function variants in the protection of telomeres 1 gene (*POT1*), with a proportion of family members presenting with an early age of onset and multiple primary tumors. We show that these variants either affect *POT1* mRNA splicing or alter key residues in the highly conserved oligonucleotide/oligosaccharide-binding (OB) domains of POT1, disrupting protein-telomere binding and leading to increased telomere length. These findings suggest that *POT1* variants predispose to melanoma formation via a direct effect on telomeres.

Cutaneous malignant melanoma accounts for around 75% of skin cancer deaths, with around 10% of cases having one first-degree relative and 1% of cases having two or more first-degree relatives who have had a diagnosis of this disease[6]. As only around half of familial melanoma cases can be attributed to variants in known predisposition genes, principally *CDKN2A*, a substantial proportion of genetic risk for melanoma remains elusive. To identify new mediators of germline genetic risk, we sequenced 184 melanoma cases from 105 pedigrees. The cases sequenced came from pedigrees with between 2 and 11 cases of melanoma (169 cases) or were single cases that presented with either multiple primary melanoma (MPM), multiple primary cancers (one of which was melanoma) and/or an early age of onset (before the fourth decade of life; 15 cases) (Online Methods and **Supplementary Tables 1** and **2**). Sequencing of two-case pedigrees was preferentially performed for those families enriched with cases of MPM. All cases were previously found to be negative for pathogenetic variants in *CDKN2A* and *CDK4*.

After performing exome (168 samples) or whole-genome (16 samples) sequencing, we called and filtered variants, keeping only those predicted to affect protein structure or function. Notably, we found no known pathogenetic variants in *BAP1* or *BRCA2*, and we confirmed that all samples had wild-type *CDKN2A* and *CDK4*. We further filtered the calls, taking forward only non-polymorphic variants (Online Methods). When we sequenced more than one member of a pedigree, we retained only cosegregating variants, whereas all variants were considered from pedigrees in which only one affected family member was sequenced. As a result, a total of 23,051 variants remained for downstream analysis. Focusing on the 28 pedigrees for which sequence data were available for 3 or more family members, we found 320 genes carrying cosegregating protein-changing variants (**Supplementary Table 3**). Of particular interest were five genes that showed previously unreported variants in more than one of these pedigrees (*POT1*, *MPDZ*, *ACD*, *SMG1* and *NEK10*). Analysis of the missense and disruptive variants (nonsense, splice acceptor or donor, and frameshift) in these genes led us to identify a five-case pedigree (UF20) carrying a *POT1* variant encoding a p.Tyr89Cys change (GRCh37 chromosome 7, g.124503684T>C) in the highly conserved N-terminal OB domain of the protein[7,8] and a six-case family (AF1) carrying a splice-acceptor variant between

[1]Experimental Cancer Genetics, Wellcome Trust Sanger Institute, Hinxton, UK. [2]Section of Epidemiology and Biostatistics, Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, UK. [3]Departamento de Bioquímica y Biología Molecular, Instituto Universitario de Oncología del Principado de Asturias (IUOPA), Universidad de Oviedo, Oviedo, Spain. [4]Oncogenomics Laboratory, QIMR Berghofer Medical Research Institute, Herston, Brisbane, Queensland, Australia. [5]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [6]Molecular Epidemiology Laboratory, QIMR Berghofer Medical Research Institute, Herston, Brisbane, Queensland, Australia. [7]Genetic Epidemiology Laboratory, QIMR Berghofer Medical Research Institute, Herston, Brisbane, Queensland, Australia. [8]Statistical Genetics, Wellcome Trust Sanger Institute, Hinxton, UK. [9]Laboratory of Translational Genomics, National Cancer Institute, Bethesda, Maryland, USA. [10]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK. [11]Department of Dermatology, Leiden University Medical Centre, Leiden, The Netherlands. [12]These authors contributed equally to this work. [13]These authors jointly directed this work. Correspondence should be addressed to D.J.A. (da1@sanger.ac.uk).

Received 29 May 2013; accepted 7 March 2014; published online 30 March 2014; doi:10.1038/ng.2947

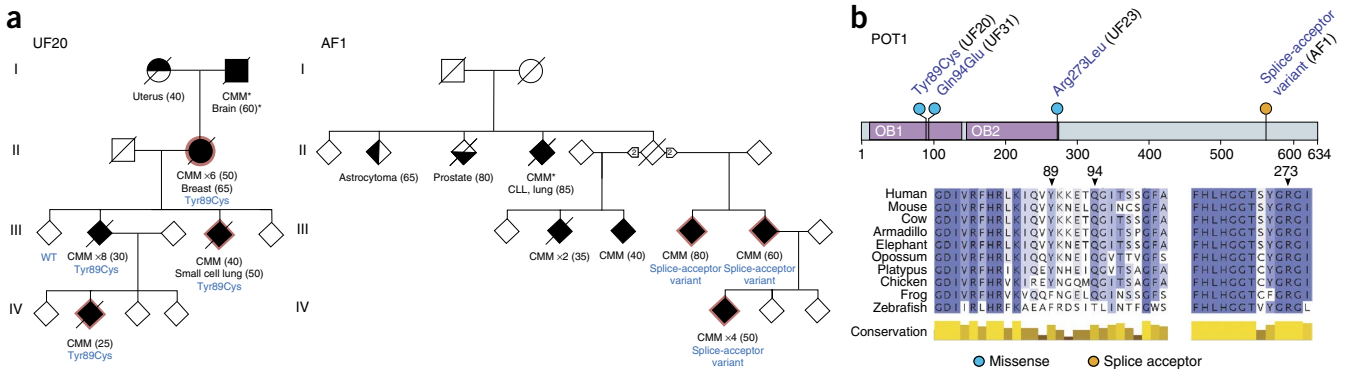


Figure 1 Rare variants in *POT1* found in familial melanoma pedigrees. (**a**) We identified four pedigrees carrying deleterious variants in *POT1*. Shown are a five-case pedigree (UF20) and a six-case pedigree (AF1) carrying the disruptive p.Tyr89Cys OB domain variant and a splice-acceptor variant, respectively. Note that pedigrees have been adjusted to protect the identity of the families without loss of scientific integrity. POT1 genotypes for all samples available for testing are shown in blue, and other types of cancer are indicated. CMM, cutaneous malignant melanoma; CLL, chronic lymphocytic leukemia; WT, wild type. Diamonds represent individuals of undisclosed sex. The cases that were sequenced have a red outline. All melanomas were confirmed by histological analysis, with the exception of two cases (marked by asterisks). The number of primary melanomas in each subject is indicated; age of onset in years is shown in parentheses. Half-filled symbols represent other cancers. (**b**) Highly conserved residues of POT1 are altered in familial melanoma. Shown are the positions of the variants identified on the POT1 protein (top) and on an amino acid alignment (missense variants; bottom).

exons 17 and 18 (g.124465412C>T) of *POT1*, which was scored as deleterious by the MaxEntScan algorithm[9] and was shown to affect transcript splicing by RT-PCR and sequencing (**Fig. 1**, **Table 1**, **Supplementary Figs. 1a–c** and **2**, and **Supplementary Tables 4** and **5**). Scrutiny of the remaining sequenced pedigrees, including those in which only one family member had been sequenced, identified two additional individuals from distinct families with nonsynonymous changes in *POT1* affecting the OB domains (UF31: g.124503670G>C, p.Gln94Glu; UF23: g.124493077C>A, p.Arg273Leu) (**Fig. 1**, **Table 1**, **Supplementary Figs. 1d,e** and **3**, and **Supplementary Tables 4** and **5**). Remarkably, the *POT1* codon for the Gln94 residue affected in pedigree UF31 has been found to be a target for recurrent somatic alteration (p.Gln94Arg) in chronic lymphocytic leukemia (CLL), where ~5% of cases carry *POT1* mutations that cluster in the sequences encoding the OB domains[10].

To gather further evidence for an association between *POT1* variants and familial melanoma, we compared the representation of *POT1* variants in our familial melanoma cases with variants in controls. Notably, the presence of *POT1* variants in 4 of 105 families with melanoma represented a statistically significant enrichment of variants ($P = 0.016$, excluding a discovery pedigree) compared with a control data set of 520 exomes from individuals sequenced as part of the UK10K project (see URLs) in which we found only 1 missense variant located outside the OB domains of POT1 (Online Methods, **Supplementary Fig. 4** and **Supplementary Table 6**). Furthermore, none of the 4 *POT1* variants identified in our melanoma pedigrees were found by genotyping 2,402 additional population-matched controls (Online Methods and **Supplementary Table 6**). Interestingly, genotyping of these positions across a matched cohort of 1,739 population-based sporadic melanoma cases identified 1 individual who carried the *POT1* variant encoding p.Arg273Leu who presented with early-onset MPM similar to the phenotype presented by the familial cases (Online Methods and **Supplementary Table 5**).

All of the missense variants we identified in *POT1* disrupt amino acids that are completely conserved throughout eutherians (**Fig. 1b**) and are more evolutionarily conserved than the average for other OB domain residues (Online Methods). Analysis of carrier individuals identified through genome and/or exome sequencing or by targeted PCR resequencing of additional family members showed that all nine carriers of *POT1* variants from the familial cohort had developed melanoma, presenting with one primary (four cases) to eight melanomas at 25 to 80 years of age (**Fig. 1** and **Supplementary Fig. 3**). One variant carrier from these familial cases also developed breast cancer at 65 years, and another developed small cell lung cancer at 50 years (pedigree UF20). Other malignancies in the untested first- or second-degree relatives of variant carriers included melanoma (pedigrees UF20 and UF31), endometrial cancer (pedigree UF20) and brain tumors (pedigrees UF20 and UF23). Intriguingly, the pedigree with the splice-acceptor variant (AF1) had a member with a history of melanoma and CLL, in keeping with a role for *POT1* in CLL development[10,11]. Collectively, these data suggest a possible role for germline *POT1* variants in susceptibility to a range of cancers in addition to melanoma.

To test whether the identified missense variants disrupted telomere binding as was observed for somatic mutations found in CLL, we examined the structure of POT1 protein bound to a telomere-like polynucleotide (dTUdAdGdGdGdTdTdAdG) (Protein Data Bank (PDB) 3KJP)[12,13]. According to this model, all 3 altered residues (Tyr89, Gln94 and Arg273) were among 24 residues located in close proximity (<3.5 Å) to the telomeric polynucleotide[10] (**Fig. 2a**). Arg273 interacts with the oxygen at position 2 of telomeric deoxythymidine 7, whereas Gln94 and Tyr89 both interact with the G deoxynucleotide at position 4. Therefore, as described for the somatic mutations in CLL, the POT1 variants we identified are expected to weaken or abolish the interaction of POT1 with telomeres. Analysis of the nucleotides coding for these 24 OB domain residues identified 1 nonsynonymous change in 6,498 control exomes[14] compared with 3 in 105 families with melanoma, emphasizing a highly significant enrichment of variants in the melanoma cohort ($P = 1.54 \times 10^{-5}$) (Online Methods and **Supplementary Tables 6** and **7**).

To further test whether the OB domain variants we identified disrupted POT1 function, we assessed the ability of *in vitro*-translated POT1 Tyr89Cys, Gln94Glu and Arg273Leu proteins to bind to $(TTAGGG)_3$ sequences. Electrophoretic mobility shift assays showed a complete abolition of POT1-DNA complex formation with mutant POT1 (**Fig. 2b** and **Supplementary Fig. 5**). Notably, the POT1 p.Tyr36Asn and p.Tyr223Cys alterations recently described in CLL[10], which seem to be functionally analogous to the variants we describe

Table 1 ***POT1*** **variants identified in familial melanoma pedigrees**

Pedigree	Number of cases in pedigree	Number of carriers/tested cases	Genomic change	Coding mutation[a]	Exon	Amino acid change	Variant type	Bioinformatic prediction tools[23–25]		
								SIFT	PolyPhen-2	CAROL
UF20	5	4/4	g.124503684T>C	c.266A>G	8	p.Tyr89Cys	Missense	Deleterious	Probably damaging	Deleterious
AF1	6	3/3	g.124465412C>T	c.1687–1G>A	–	–	Splice acceptor (intronic, between exons 17 and 18)	–	–	–
UF31	2	1/1	g.124503670G>C	c.280C>G	8	p.Gln94Glu	Missense	Tolerated	Probably damaging	Deleterious
UF23	2	1/2[b]	g.124493077C>A	c.818G>T	10	p.Arg273Leu[c]	Missense	Deleterious	Probably damaging	Deleterious

[a]The reference transcript, taken from the Ensembl database (release 70), is POT1-001 (ENST00000357628). [b]A second case within this pedigree had a different clinical presentation (solitary melanoma *in situ*) in the sixth decade of life and did not carry the p.Arg273Leu variant. [c]This variant was also detected in a melanoma case from a population-based case-control series that presented with MPMs and an early age of onset (Online Methods).

here, promote uncapping of telomeres, telomere length extension and chromosomal aberrations and thereby promote tumorigenesis[10].

Given the role of POT1 in telomere length maintenance, we next asked whether melanoma cases from pedigrees with mutated *POT1* had telomere lengths that differed from those of non-carrier melanoma cases. Using exome sequence data from 41 cases, including 3 members of pedigree UF20, we estimated the telomere length of each subject by counting TTAGGG repeats[15]. This analysis showed that all three members of pedigree UF20 had telomeres that were significantly longer than those in melanoma cases with wild-type *POT1* ($P < 0.0002$; **Fig. 2c** and **Supplementary Fig. 6**). This result was confirmed by telomere-length PCR, which also showed longer telomeres for subjects carrying the p.Gln94Glu and p.Arg273Leu variants compared to melanoma cases without *POT1* variants ($P = 3.62 \times 10^{-5}$; **Fig. 2d** and **Supplementary Fig. 7**). Thus, missense variants in the OB domains of POT1 not only abolish telomere binding but are also associated with increased telomere length, a key factor influencing melanoma risk[16]. Notably, OB domain variants that disrupt the interaction of POT1 with telomeric single-stranded DNA are thought to function as dominant-negative alleles[10,17], yet, as we show here, they are compatible with life, suggesting that additional somatic events are required to promote tumorigenesis.

The identification of *POT1* mutations in CLL and the probable susceptibility of our *POT1*-mutated familial melanoma pedigrees to other tumor types suggest that *POT1* might have a more general role in tumorigenesis. To investigate this possibility, we examined pan-cancer data from the Catalogue of Somatic Mutations in Cancer (COSMIC)[18] and IntOGen[19] databases (data from The Cancer Genome Atlas

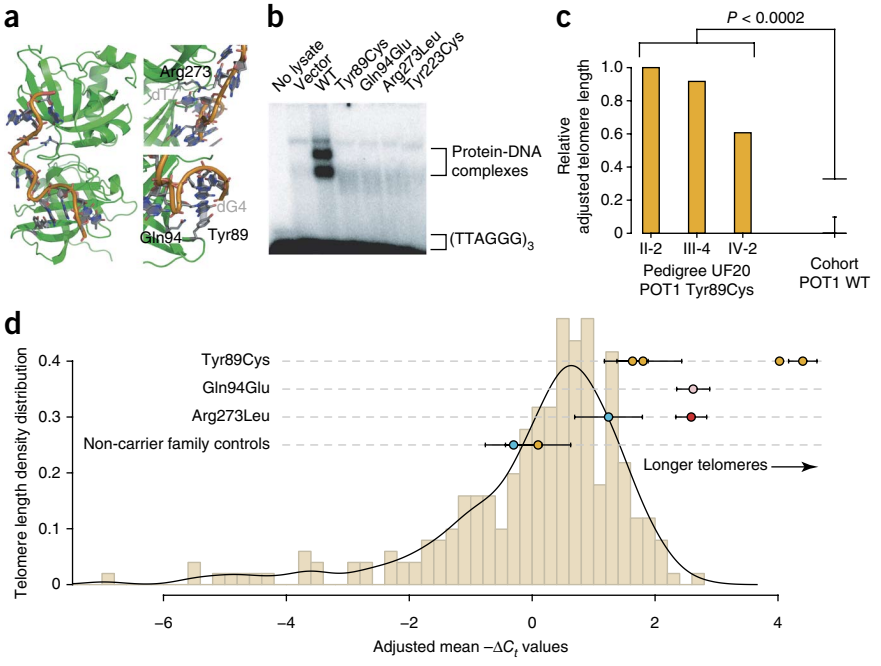


Figure 2 Missense variants in *POT1* disrupt the interaction between POT1 and single-stranded DNA and lead to elongated telomeres. (**a**) Shown are the locations of the POT1 Tyr89, Gln94 and Arg273 residues in the two N-terminal OB domains (green). A telomere-like polynucleotide sequence is shown in orange. Interacting nucleotides in the telomeric sequence are labeled in gray. All three substitutions are predicted to disrupt the association of POT1 with telomeres. (**b**) Mutant Tyr89Cys, Gln94Glu and Arg273Leu POT1 proteins are unable to bind telomeric $(TTAGGG)_3$ sequences as shown by an electrophoretic mobility shift assay. The Tyr223Cys POT1 mutant was used as a positive control representing a known disruptive alteration[10]. (**c**) Calculation of telomere length from exome sequence data. The method used has been described previously[15]. Relative adjusted telomere lengths for the 3 sequenced members of pedigree UF20 are shown alongside the mean telomere length of 38 (all other) melanoma cases that were sequenced alongside them but were wild type for *POT1*. Error bars, 1 s.d. A Wilcoxon rank-sum test was performed comparing the telomere lengths of the 3 Tyr89Cys cases to that for the 38 non-carrier controls. (**d**) PCR-based estimates of telomere length. Adjusted mean $-\Delta C_t$ values, which correlate positively with telomere length, for *POT1* missense variant carriers and non-carrier family controls are shown against a distribution of values from 252 melanoma cases recruited from the Leeds Melanoma cohort that are wild type at the above-mentioned positions (Online Methods). All measurements have been adjusted for age at blood draw and sex. The black line represents a Gaussian kernel density estimate for this set using Silverman's rule of thumb[22] for bandwidth smoothing. Orange dots, members of pedigree UF20; pink dots, members of pedigree UF31; blue dots, members of pedigree UF23; red dots, individual CT1663 from the Leeds Melanoma case-control study carrying the p.Arg273Leu variant (**Supplementary Table 5**). The number of biological replicates for each individual ranged from one to four, each with two technical replicates, for the *POT1* missense variant carriers and non-carrier family controls. Two technical replicates were performed for the 252 *POT1* non-carrier cases. Error bars, s.e.m.

(TCGA) and the International Cancer Genome Consortium (ICGC)) from 14 cancer types and found that somatic *POT1* mutations were more likely to be missense ($P < 0.03$), to alter residues in close proximity to DNA ($P < 0.02$) and to have a higher functional bias ($P < 0.03$) than expected by chance (Online Methods). These results suggest that, although they are rare, somatic *POT1* mutations may drive tumorigenesis across multiple histologies.

Here we describe germline variants in the gene encoding the telomere-associated protein POT1 in almost 4% of familial melanoma pedigrees negative for mutations in *CDKN2A* and *CDK4* and in 2 of 34 pedigrees (5.8%) with ≥5 cases, making *POT1* the second most frequently mutated high-penetrance melanoma gene reported thus far. This work and a companion study describing germline *POT1* variants in Italian, French and US families with melanoma[20], together with a recent report of a *TERT* promoter variant[2], substantially extend understanding of a newly discovered mechanism predisposing to the development of familial melanoma. As the dysregulation of telomere protection by POT1 has recently been identified as a target for potential therapeutic intervention[21], it may be possible that the early identification of families with *POT1* variants might facilitate better management of their disease in the future.

URLs. UK10K Sequencing Project, http://www.uk10k.org/; European Genome-phenome Archive (EGA), https://www.ebi.ac.uk/ega/; dbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP/; PyMOL Molecular Graphics System, Version 0.99, http://www.pymol.org/.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Sequence data have been deposited in the European Genome-phenome Archive (EGA), hosted at the European Bioinformatics Institute, under accession EGAS00001000017.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank the UK10K Consortium (funded by the Wellcome Trust; WT091310) for access to control data. D.J.A., C.D.R.-E., Z.D., J.Z.L., J.C.T., M.P. and T.M.K. were supported by Cancer Research UK and the Wellcome Trust (WT098051). C.D.R.-E. was also supported by the Consejo Nacional de Ciencia y Tecnología of Mexico. K.A.P. and A.M.D. were supported by Cancer Research UK (grants C1287/A9540 and C8197/A10123) and by the Isaac Newton Trust. N.K.H. was supported by a fellowship from the National Health and Medical Research Council of Australia (NHMRC). L.G.A. was supported by an Australia and New Zealand Banking Group Limited Trustees PhD scholarship. A.L.P. is supported by Cure Cancer Australia. The work was funded in part by the NHMRC and Cancer Council Queensland. The work of N.A.G. was in part supported by the Dutch Cancer Society (UL 2012-5489). M.H., J.A.N.-B. and D.T.B. were supported by Cancer Research UK (programme awards C588/A4994 and C588/A10589 and the Genomics Initiative). C.L.-O., A.J.R. and V.Q. are funded by the Spanish Ministry of Economy and Competitiveness through the Instituto de Salud Carlos III (ISCIII), the Red Temática de Investigación del Cáncer (RTICC) del ISCIII and the Consolider-Ingenio RNAREG Consortium. C.L.-O. is an investigator with the Botín Foundation.

AUTHOR CONTRIBUTIONS

C.D.R.-E., M.H., J.A.N.-B., D.T.B., N.K.H. and D.J.A. designed the study and wrote the manuscript. C.D.R.-E., M.H., L.G.A., J.C.T., M.M., J.C., M.P., A.J.R., Z.D., V.Q., A.L.P., J.M.P., J.S., M.S.S., N.G.M., M.G.G., A.M.D., K.A.P., P.J., J.Z.L., K.M.B., C.L.-O. and T.M.K. performed experiments or analysis. N.A.G., G.W.M., H.S. and N.G.M. provided vital biological resources.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Goldstein, A.M. *et al.* Features associated with germline *CDKN2A* mutations: a GenoMEL study of melanoma-prone families from three continents. *J. Med. Genet.* **44**, 99–106 (2007).
2. Horn, S. *et al. TERT* promoter mutations in familial and sporadic melanoma. *Science* **339**, 959–961 (2013).
3. Breast Cancer Linkage Consortium. Cancer risks in *BRCA2* mutation carriers. *J. Natl. Cancer Inst.* **91**, 1310–1316 (1999).
4. Wiesner, T. *et al.* Germline mutations in *BAP1* predispose to melanocytic tumors. *Nat. Genet.* **43**, 1018–1021 (2011).
5. Zuo, L. *et al.* Germline mutations in the p16INK4a binding domain of *CDK4* in familial melanoma. *Nat. Genet.* **12**, 97–99 (1996).
6. Law, M.H., Macgregor, S. & Hayward, N.K. Melanoma genetics: recent findings take us beyond well-traveled pathways. *J. Invest. Dermatol.* **132**, 1763–1774 (2012).
7. Loayza, D. & De Lange, T. POT1 as a terminal transducer of TRF1 telomere length control. *Nature* **423**, 1013–1018 (2003).
8. Baumann, P. & Cech, T.R. Pot1, the putative telomere end–binding protein in fission yeast and humans. *Science* **292**, 1171–1175 (2001).
9. Yeo, G. & Burge, C.B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* **11**, 377–394 (2004).
10. Ramsay, A.J. *et al. POT1* mutations cause telomere dysfunction in chronic lymphocytic leukemia. *Nat. Genet.* **45**, 526–530 (2013).
11. Speedy, H.E. *et al.* A genome-wide association study identifies multiple susceptibility loci for chronic lymphocytic leukemia. *Nat. Genet.* **46**, 56–60 (2014).
12. Nandakumar, J., Podell, E.R. & Cech, T.R. How telomeric protein POT1 avoids RNA to achieve specificity for single-stranded DNA. *Proc. Natl. Acad. Sci. USA* **107**, 651–656 (2010).
13. Lei, M., Podell, E.R. & Cech, T.R. Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection. *Nat. Struct. Mol. Biol.* **11**, 1223–1229 (2004).
14. Fu, W. *et al.* Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* **493**, 216–220 (2013).
15. Ding, Z. *et al.* Estimating telomere length from whole genome sequence data. *Nucleic Acids Res.* doi:10.1093/nar/gku181 (7 March 2014).
16. Burke, L.S. *et al.* Telomere length and the risk of cutaneous malignant melanoma in melanoma-prone families with and without *CDKN2A* mutations. *PLoS ONE* **8**, e71121 (2013).
17. Kendellen, M.F., Barrientos, K.S. & Counter, C.M. POT1 association with TRF2 regulates telomere length. *Mol. Cell. Biol.* **29**, 5611–5619 (2009).
18. Forbes, S.A. *et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39**, D945–D950 (2011).
19. Gonzalez-Perez, A. *et al.* IntOGen-mutations identifies cancer drivers across tumor types. *Nat. Methods* **10**, 1081–1082 (2013).
20. Shi, J. *et al.* Rare missense variants in *POT1* predispose to familial cutaneous malignant melanoma. *Nat. Genet.* doi:10.1038/ng.2941 (30 March 2014).
21. Jacobs, J.J. Loss of telomere protection: consequences and opportunities. *Front. Oncol.* **3**, 88 (2013).
22. Silverman, B.W. *Density Estimation* (Chapman and Hall, London, 1986).
23. Ng, P.C. & Henikoff, S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).
24. Adzhubei, I.A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
25. Lopes, M.C. *et al.* A combined functional annotation score for non-synonymous variants. *Hum. Hered.* **73**, 47–51 (2012).

ONLINE METHODS

Case samples and DNA extraction. The families included in this study were recruited to a UK Familial Melanoma Study directed by the Section of Epidemiology and Biostatistics, University of Leeds (Leeds, UK); the Leiden University Medical Center (Leiden, The Netherlands); and the Queensland Familial Melanoma Project (QFMP)[26]. Informed consent was obtained under the Multicentre Research Ethics Committee (UK): 99/3/045 (UK Familial Melanoma Study cases), Protocol P00.117-gk2/WK/ib (Leiden cases) and from the Human Research Ethics Committee of the QIMR Berghofer Medical Research Institute for QFMP cases. Genomic DNA was extracted from peripheral blood using standard methods.

Pedigrees and clinical presentation. The pedigrees in this study are listed by institute and by sequencing center in **Supplementary Table 1**. All pedigrees, the number of cases of melanoma in each pedigree and the number of cases that were whole-genome sequenced are listed in **Supplementary Table 2**. To help anonymize the pedigrees, ages were rounded up to the nearest 5-year tier.

Sequence alignment and analysis. DNA libraries were prepared from 5 µg of genomic DNA, and exonic regions were captured with the Agilent SureSelect Target Enrichment System, 50 Mb Human All Exon kit. Whole-genome libraries were prepared using the standard Illumina library preparation protocol. Paired-end reads of between 75 and 100 bp were generated on the HiSeq 2000 platform and mapped to the reference GRCh37/hg19 human genome assembly using the Burrows-Wheeler Aligner (BWA)[27]. Reads were filtered for duplicates using Picard[28] and were recalibrated and realigned around indels using the Genome Analysis Toolkit (GATK) package[29] (Familial Melanoma Study and Leiden data). Exome capture and sequencing resulted in an average of 84% of target bases being covered by ≥10× across the autosomes and sex chromosomes. Whole genomes were sequenced to at least 27× mapped coverage. Data for *POT1* variant carriers have been released (EGAS00001000017). Variants were then called using SAMtools mpileup[30] and filtered for quality. The variant collection was filtered to remove positions found in Phase 1 of the 1000 Genomes Project October 2011 release[31] and the dbSNP 135 release (see URLs). Variants were also filtered for positions found in a collection of 805 in-house control exomes. Only variants in exonic regions, as defined in Ensembl release 70, were taken forward for analysis. Positions resulting in protein-altering changes were then identified using the Ensembl Variant Effect Predictor, version 2.8 (Ensembl release 70)[32], a combination of VCFTools[33] and custom scripts. Variants marked as 'transcript_ablation', 'splice_donor_variant', 'splice_acceptor_variant', 'stop_gained', 'frameshift_variant', 'stop_lost', 'initiator_codon_variant', 'inframe_insertion', 'inframe_deletion', 'missense_variant', 'transcript_amplification', 'splice_region_variant', 'incomplete_terminal_codon_variant', 'mature_miRNA_variant', 'TFBS_ablation', 'TFBS_amplification', 'TF_binding_site_variant', 'feature_elongation' and 'feature_truncation' were kept for further analyses. We retained only those variants found in all affected cases of a single pedigree (to reduce the impact of systematic mapping errors). An identity-by-descent (IBD) analysis was performed to confirm that cases from different pedigrees within the study were not related.

Genes with cosegregating variants from the 28 pedigrees for which we had sequence data for 3 or more family members are shown in **Supplementary Table 3** with their Gene Ontology (GO) terms. Variants in *POT1* identified from this analysis were confirmed by capillary sequencing (**Supplementary Fig. 1**). Several low-penetrance variants in *MC1R* were also identified in the pedigrees with *POT1* variants; all of these are common variants associated with freckling and sun sensitivity (**Supplementary Table 8**). We also identified a variant at the +6 intronic splice site of introns 17 and 18 of *POT1* (g.124467262A>C) in one melanoma pedigree, but our analyses suggested that this variant was unlikely to be deleterious (**Supplementary Note**), although the telomeres of the subject carrying this variant at +6 appeared longer than the telomeres of controls, suggesting some effect on telomere regulation (**Supplementary Fig. 8**).

MaxEntScan scoring of splice-site acceptor variant g.124465412C>T. We used the MaxEntScan algorithm[9], which yielded scores of −3.22 for the mutated splice-acceptor site (g.124465412C>T) and 5.53 for the wild-type splice-site sequence. To put these values in context, we retrieved 10,000 splice-acceptor and splice-donor sites from random genes (choosing always the second exon) and obtained a distribution of their scores. The splice-acceptor variant lowered the score of the wild-type sequence from the 9.2 to the 0.57 percentile compared to the score distribution of real splice acceptors (**Supplementary Fig. 2**) and is thus predicted to be highly deleterious.

RT-PCR sequencing of the *POT1* product in two individuals carrying the splice-acceptor variant g.124465412C>T. RNA extracted from the whole blood of two carriers of the splice-acceptor variant was converted to cDNA using SuperScript III Reverse Transcriptase (Invitrogen). RT-PCR was then performed to confirm that *POT1* g.124465412C>T (c.1687–1G>A; ENST00000357628) was indeed disruptive to splicing. M13-tagged forward primer and reverse primer were designed to flank the spliced region (primer sequences available upon request). The product was visualized on a 3% NuSieve Agarose gel, and the sequence was verified using standard Sanger sequencing methods. Sequencing traces for one control and one carrier sample are shown in **Supplementary Figure 1c**.

Frequency of *POT1* variants in a control data set. All exomes from the UK10K sequencing project (REL 14/03/12) cohorts UK10K_NEURO_MUIR, UK10K_NEURO_IOP_COLLIER and UK10K_NEURO_ABERDEEN (see URLs) were selected as controls because these exomes were captured with the same Agilent SureSelect exome probes as those used for the melanoma cases described above and were also sequenced on the Illumina HiSeq 2000 platform ($n = 546$). One exome was discarded at random from each of three pairs of relatives within this set. UK10K exomes were aligned, filtered for duplicates, and recalibrated and realigned around indels as described above. Variants were then called and filtered for base quality with the same tools and parameters as the melanoma cohort. For 104 of 105 families, we had exome data for at least 1 individual in the pedigree; for 1 melanoma family, we used whole-genome sequence.

To ensure that the controls were matched by ancestry to the melanoma cohort, we performed a principal-component analysis (PCA) using 1,092 individuals across 14 populations from the 1000 Genomes Project Phase 1 data set[31]. A subset of high-quality variant positions (quality score >10, minimum mapping quality >10, strand bias P value >0.0001, end distance bias P value >0.0001) that were common to the melanoma cohort and the UK10K controls, as well as the 1000 Genomes Project data set, were taken forward for analysis. SNPs with a minor allele frequency of <0.05 or that were in linkage disequilibrium with another SNP (pairwise $r^2 > 0.1$) in the 1000 Genomes Project data set or that had a Hardy-Weinberg P value of $<1 \times 10^{-5}$ in the UK10K controls were excluded. After filtering, 7,196 SNPs remained that were spread across all autosomes. The first ten principal components were estimated using the 1000 Genomes Project individuals and were then projected onto the melanoma cohort samples and UK10K controls using EIGENSTRAT[34]. Controls lying greater than 2 s.d. from the mean scores for principal component 1 or 2 (PC1 or PC2, respectively), calculated using only European individuals in the 1000 Genomes Project data set, were removed from subsequent analyses ($n = 20$). This analysis is shown in **Supplementary Figure 4**. An IBD analysis was performed to ensure that members of the UK10K cohort were not related. This analysis was performed using the PLINK toolset[35] and the same set of variants that were used for PCA. For each pair of individuals with an estimated IBD of >0.2, 1 individual was removed at random ($n = 3$). This filtering left 520 exomes for comparison against the melanoma cohort.

Variants in this collection of 520 UK10K control exomes were then filtered as described above (keeping positions with exonic coordinates ±100 bp and removing all variants in Phase 1 of the 1000 Genomes Project October 2011 release[31], the dbSNP 138 release (see URLs) and a collection of 805 control exomes). Because we used an updated version of dbSNP for this step, we also checked that the *POT1* variants found in this study passed this filter. Consequences were then predicted and filtered as described above. From this analysis, we identified 1 individual in 520 control exomes that carried a rare, potentially disruptive variant in *POT1* (a missense variant located outside of the OB domains). We performed a two-tailed Fisher's exact test comparing the 3 out of 104 families with melanoma, excluding a discovery pedigree, to

the 1 individual out of 520 controls carrying rare variants in *POT1*, yielding a *P* value of 0.016.

Genotyping in a population-based case-control series (TaqMan). The variants encoding p.Tyr89Cys, p.Gln94Glu and p.Arg273Leu and the splice-acceptor variant were genotyped in 2,402 control samples belonging to the Leeds Melanoma Case-Control Study. This control set included 499 population-matched control DNA samples, 370 family controls (family members of melanoma cases without a diagnosis of melanoma) and 1,533 DNA samples from the Wellcome Trust Case Control Consortium. All 2,402 samples were wild type for the *POT1* variants. We also genotyped the corresponding positions in 1,739 population-based melanoma cases that were recruited from across Yorkshire, UK, as part of the same study. One case, presenting with MPMs with early onset (48 years old), was found to be a carrier of the variant encoding p.Arg273Leu (**Table 1** and **Supplementary Tables 5** and **6**). This variant was confirmed by PCR sequencing.

Protein alignment, structural modeling and characterization of POT1 variants. The amino acid sequences of POT1 from evolutionarily diverse species were gathered from NCBI and aligned with Clustal Omega[36]. Alignments were displayed using Jalview v2.7 (ref. 37). To estimate the number of substitutions per site in this amino acid alignment, we used the ProtPars routine from PHYLIP[38]. This analysis showed higher conservation for the three altered amino acids (2, 2 and 0 substitutions at positions 89, 94 and 273, respectively) than the average for the OB domains (2.42 substitutions per site) and, in fact, the whole protein (3.49 substitutions per site) across ~450 million years of evolutionary history (since the divergence of the zebrafish and human lineages). If only sequences from eutherian organisms were taken into account, then no substitutions have occurred in any of the three residues, compared to 0.8 substitutions per site in the OB domains and 1.39 substitutions per site in the whole protein. The OB domain regions were defined as amino acids 8–299 in the human sequence, according to Ensembl superfamily domain annotation. The structure of the OB domains of POT1 (3KJO) was obtained from PDB and was rendered with PyMOL v0.99 (see URLs).

Analysis of nucleotide variants coding for the 24 key OB domain residues in close proximity (3.5 Å) to telomeric DNA. Ramsay *et al.*[10] defined a list of 24 residues that lie closer than 3.5 Å to telomeric DNA in the crystal structure of POT1 (PDB 3KJP): residues 31, 33, 36, 39–42, 48, 60, 62, 87, 89, 94, 159, 161, 223, 224, 243, 245, 266, 267, 270, 271 and 273. To assess the statistical significance of finding amino acid substitutions affecting these residues, we searched 6,503 exomes that were part of the National Heart, Lung, and Blood Institute (NHLBI) Grand Opportunity (GO) Exome Sequencing Project (ESP)[14] for substitutions at any of the bases that would cause a change in these amino acids. The genomic positions that encode these 24 residues are shown in **Supplementary Table 7**. In summary, a minimum of 6,498 exomes had all bases covered at a minimum average coverage of 59×. The variant encoding p.Asn224Asp was found at an overall allele frequency of 1 in 13,005. No other amino acid–changing variants were found. We compared the number of variants found in the 24 key OB domain residues in controls (1 in 6,498) to the number of variants found in all analyzed pedigrees (3 in 105), obtaining a *P* value of 1.54×10^{-5} using a two-tailed Fisher's exact test (**Supplementary Table 6**).

***In vitro* translation and G strand binding assays.** We mutated human *POT1* in a T7 expression vector (Origene) by site-directed mutagenesis to generate cDNAs encoding the POT1 Tyr89Cys, Gln94Glu and Arg273Leu variants. Mutant and control T7 expression vectors were used in an *in vitro* translation reaction using the TNT coupled reticulocyte lysate kit (Promega) following the manufacturer's instructions. Briefly, a 50-µl reaction mixture containing 1 µg of plasmid DNA, 2 µl of EasyTag ^{35}S-labeled L-methionine (1,000 Ci/mmol; PerkinElmer) and 25 µl of rabbit reticulocyte lysate was incubated at 30 °C for 90 min. A 5-µl fraction of each reaction was analyzed by SDS-PAGE; proteins were visualized and relative amounts were quantified using the FLA 7000 phosphorimager system (Fujifilm) (**Supplementary Fig. 5**). DNA binding assays were performed as described previously with minor modifications[39]. In 20-µl reaction mixtures, 5 µl of each translation reaction was incubated with 10 nM telomeric oligonucleotide 5′ labeled with ^{32}P (5′-GGTTAGGGTTAGGGTTAGGG-3′) and 1 µg of the nonspecific competitor DNA poly(deoxyinosinic-deoxycytidylic) acid in binding buffer (25 mM HEPES-NaOH, pH 7.5, 100 mM NaCl, 1 mM EDTA and 5% glycerol). Reactions were incubated for 10 min at room temperature, and protein-DNA complexes were analyzed by electrophoresis on a 6% polyacrylamide Tris-borate-EDTA gel run at 80 V for 3 h. Gels were visualized by exposure to a phosphorimager screen.

Analysis of telomere length from next-generation sequencing data. Telomere length was determined essentially as described[15]. The investigator who performed this analysis was blinded to the *POT1* status of the 41 eligible samples (which were all sequenced at the Sanger Institute, as they were the only ones with enough data available; **Supplementary Table 2**).

After calculation of relative telomere length, the 38 samples without germline *POT1* variants were adjusted for age at blood draw and sex using a linear model (**Supplementary Fig. 6**). The corresponding values for *POT1* variant carriers were estimated on the basis of the same adjustment. A Wilcoxon rank-sum test comparing adjusted telomere lengths for non-carrier melanoma cases and the three members of pedigree UF20 (p.Tyr89Cys variant carriers) ($P = 0.00019$) supported the finding of increased telomere lengths for variant carriers. In **Figure 2c**, all values are shown relative to the largest sample measurement.

Analysis of telomere length (PCR). We measured telomere length in melanoma cases recruited from the Leeds Melanoma cohort who did not carry a *POT1* variant, seven *POT1* missense variant carriers (pedigrees UF20, UF31 and UF23 and the carrier individual from the Leeds Melanoma cohort) and two non-carrier family controls (UF23, individual III-1 and UF20, individual III-1). The investigator who performed this analysis was blinded to the *POT1* status of all samples. Relative mean telomere length was ascertained by SYBR Green RT-PCR using a version of the published Q-PCR protocols[40,41] that was modified as described previously[42]. In brief, genomic DNA was extracted from whole blood, and telomere length was ascertained by determining the ratio of detected fluorescence from the amplification of telomere repeat units (TEL) relative to fluorescence for a single-copy reference sequence from the *HBB* (β-globin) gene (CON). Telomere and control reactions were performed separately. For each assay, the PCR cycle at which each reaction crossed a predefined fluorescence threshold was determined (C_t value). The difference in the C_t values, $\Delta C_t = C_t$ TEL $- C_t$ CON, was the measure of telomere length used in the analysis, as in other published data generated using this assay[42,43].

For the analysis, samples with C_t CON < 18, C_t CON > 27 or C_t CON > 2 s.d. away from the mean were removed and considered to represent failed reactions. This filtering left 252 samples from the Leeds Melanoma cohort for further analyses, with no missense variant carriers or non-carrier family controls removed. All samples had between two and eight technical replicates. Mean ΔC_t values for each sample were estimated from all replicates. The estimated mean values of ΔC_t obtained from melanoma cases without germline *POT1* variants were adjusted for age at blood draw and sex using a linear model (**Supplementary Fig. 7**). The corresponding values for *POT1* variant carriers were estimated on the basis of the same adjustment. Adjusted mean ΔC_t values are plotted (**Fig. 2d**), with the histogram showing the non-carrier melanoma cases compared to the missense variant carriers and the non-carrier family controls plotted above. A Wilcoxon rank-sum test comparing the adjusted mean ΔC_t values for the 252 non-carrier melanoma cases with those for the 7 missense variant carriers ($P = 3.62 \times 10^{-5}$) supported the finding of increased telomere lengths for variant carriers.

Analysis of *POT1* mutations in cancer databases. Although mutations in *POT1* have not been found at a high frequency in the cancer studies deposited in COSMIC[18] and IntOGen[19] (which integrates only whole-exome data from ICGC and TCGA as well as other studies), the mutations that have been reported show a tendency to be missense, alter residues that are predicted to interact with DNA and have a high functional impact bias.

To statistically assess the mutational patterns affecting *POT1* in cancer, we compiled a list of all residues closer than 3.5 Å to the telomeric DNA in the crystal structure of POT1 (PDB 3KJP)[10]. We then mined COSMIC database

doi:10.1038/ng.2947

v66 for confirmed somatic mutations absent from the 1000 Genomes Project affecting the ORF of *POT1* across 14 cancer types (breast, central nervous system, endometrium, hematopoietic and lymphoid tissue, kidney, large intestine, liver, lung, ovary, parathyroid, prostate, skin, urinary tract and not specified). This analysis yielded 35 somatic mutations, including 4 that were silent. We also compiled the total frequency of each reference/mutated base pair in the same COSMIC database. Finally, we performed a Monte Carlo simulation with 100,000 groups of 35 mutations at random locations in the *POT1* ORF. The probability of a given mutation from a reference base (for example, A to G) was forced to equal the frequency for that pair in the whole COSMIC database.

Of the 100,000 simulations performed with this method, only 2,971 contained 4 or fewer silent mutations. Therefore, the COSMIC database contains fewer silent mutations affecting *POT1* than expected by chance ($P < 0.03$). To assess the clustering of mutations at sites encoding DNA-binding residues (**Supplementary Table 7**), we only considered missense mutations, as no selection would be expected for nonsense mutations. In the COSMIC database, we found 27 *POT1* missense mutations, 4 of which affected telomere-binding residues. In the Monte Carlo experiment, 9,244 simulations had exactly 27 missense mutations. In only 176 of these simulations were 4 or more residues identified that were classified as disrupting telomere binding. This result suggests that *POT1* missense mutations affect DNA-binding residues at a higher than expected rate in the COSMIC database ($P < 0.02$).

To assess the functional impact bias of somatic mutations in *POT1*, we also looked at all mutations in *POT1* that are present in the IntOGen database[19]. We chose this database because it integrates only samples that have been whole-exome sequenced and thus can provide a valid, non-biased estimate of the functional impact of mutations in *POT1* when they are compared with mutations in the rest of the exome. The frequency of *POT1* mutations in this data set is ~0.01 across 9 cancer sites (all of those available in the database, contained in the 14 sites listed above). *P* values for the three studies for which the gene passed set thresholds defined in the database[19], calculated with Oncodrive-fm[44], were combined to yield a *P* value of 0.021, indicating that this gene is biased toward the accumulation of functional mutations.

26. Aitken, J.F., Green, A.C., MacLennan, R., Youl, P. & Martin, N.G. The Queensland Familial Melanoma Project: study design and characteristics of participants. *Melanoma Res.* **6**, 155–165 (1996).
27. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
28. DePristo, M.A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
29. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
30. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
31. 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
32. McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069–2070 (2010).
33. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
34. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
35. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
36. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
37. Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. & Barton, G.J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
38. Felsenstein, J. PHYLIP—Phylogeny Inference Package (Version 3.2). *Cladistics* **5**, 164–166 (1989).
39. Baumann, P., Podell, E. & Cech, T.R. Human Pot1 (protection of telomeres) protein: cytolocalization, gene structure, and alternative splicing. *Mol. Cell. Biol.* **22**, 8079–8087 (2002).
40. McGrath, M., Wong, J.Y., Michaud, D., Hunter, D.J. & De Vivo, I. Telomere length, cigarette smoking, and bladder cancer risk in men and women. *Cancer Epidemiol. Biomarkers Prev.* **16**, 815–819 (2007).
41. Cawthon, R.M. Telomere measurement by quantitative PCR. *Nucleic Acids Res.* **30**, e47 (2002).
42. Pooley, K.A. *et al.* Telomere length in prospective and retrospective cancer case-control studies. *Cancer Res.* **70**, 3170–3176 (2010).
43. Bojesen, S.E. *et al.* Multiple independent variants at the *TERT* locus are associated with telomere length and risks of breast and ovarian cancer. *Nat. Genet.* **45**, 371–384 (2013).
44. Gonzalez-Perez, A. & Lopez-Bigas, N. Functional impact bias reveals cancer drivers. *Nucleic Acids Res.* **40**, e169 (2012).

Supplementary Material

POT1 loss-of-function variants predispose to familial melanoma

Carla Daniela Robles-Espinoza[1]‡, Mark Harland[2]‡, Andrew J. Ramsay[3]‡, Lauren G. Aoude[4]‡, Víctor Quesada[3], Zhihao Ding[1], Karen A. Pooley[5], Antonia L. Pritchard[4], Jessamy C. Tiffen[1], Mia Petljak[1], Jane M. Palmer[4], Judith Symmons[4], Peter Johansson[4], Mitchell S. Stark[4], Michael G. Gartside[4], Helen Snowden[2], Grant W. Montgomery[6], Nicholas G. Martin[7], Jimmy Z. Liu[8], Jiyeon Choi[9], Matthew Makowski[9], Kevin M. Brown[9], Alison M. Dunning[10], Thomas M. Keane[1], Carlos López-Otín[3], Nelleke A. Gruis[11], Nicholas K. Hayward[4]†, D. Timothy Bishop[2]†, Julia A. Newton-Bishop[2]†, David J. Adams[1]†*

‡ These authors contributed equally to this work

† These authors jointly directed this work

Affiliations:

[1]Experimental Cancer Genetics, Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA. UK.

[2]Section of Epidemiology and Biostatistics, Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, LS9 7TF, UK.

[3]Departamento de Bioquímica y Biología Molecular, Instituto Universitario de Oncología del Principado de Asturias (IUOPA) Universidad de Oviedo, Oviedo, Spain.

[4]Oncogenomics Laboratory, QIMR Berghofer Medical Research Institute, Herston, Brisbane, QLD, Australia.

[5]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.

[6]Molecular Epidemiology Laboratory, QIMR Berghofer Medical Research Institute, Herston, Brisbane, QLD, Australia.

[7]Genetic Epidemiology Laboratory, QIMR Berghofer Medical Research Institute, Herston, Brisbane, QLD, Australia.

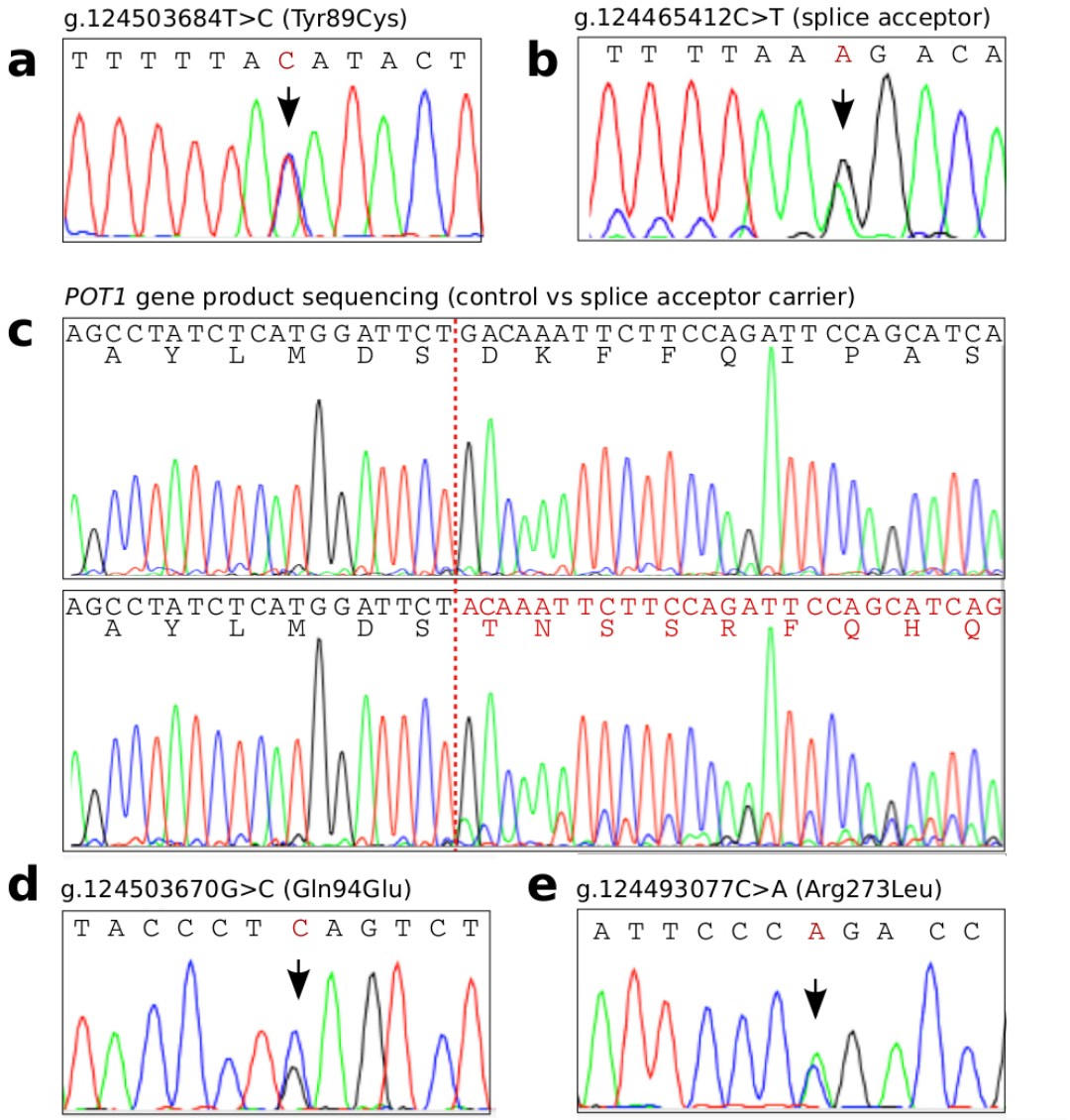
[8]Statistical Genetics, Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA. UK.

[9]Laboratory of Translational Genomics, National Cancer Institute, Bethesda, MD, USA.

[10]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK

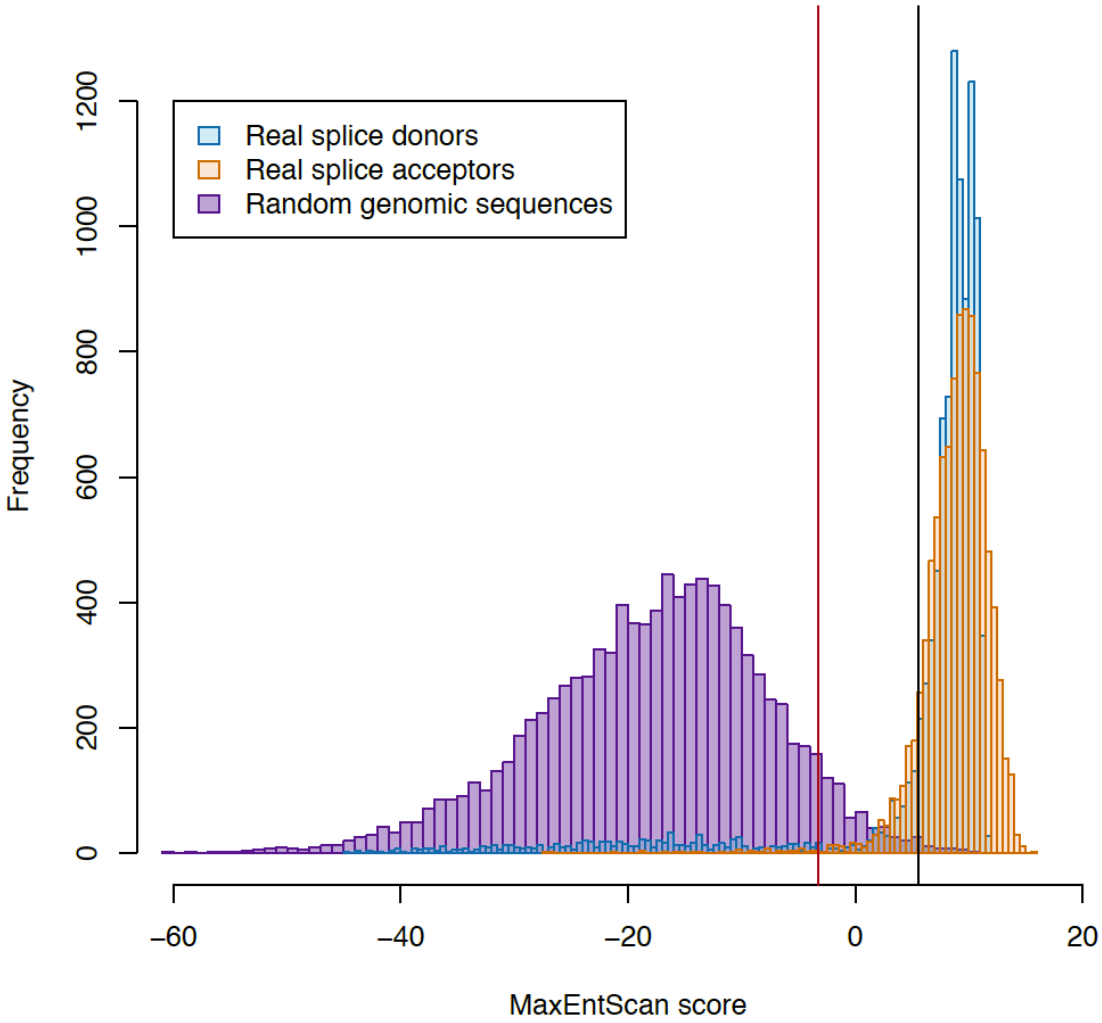
[11]Department of Dermatology, Leiden University Medical Centre, Leiden, The Netherlands.

Nature Genetics: doi:10.1038/ng.2947

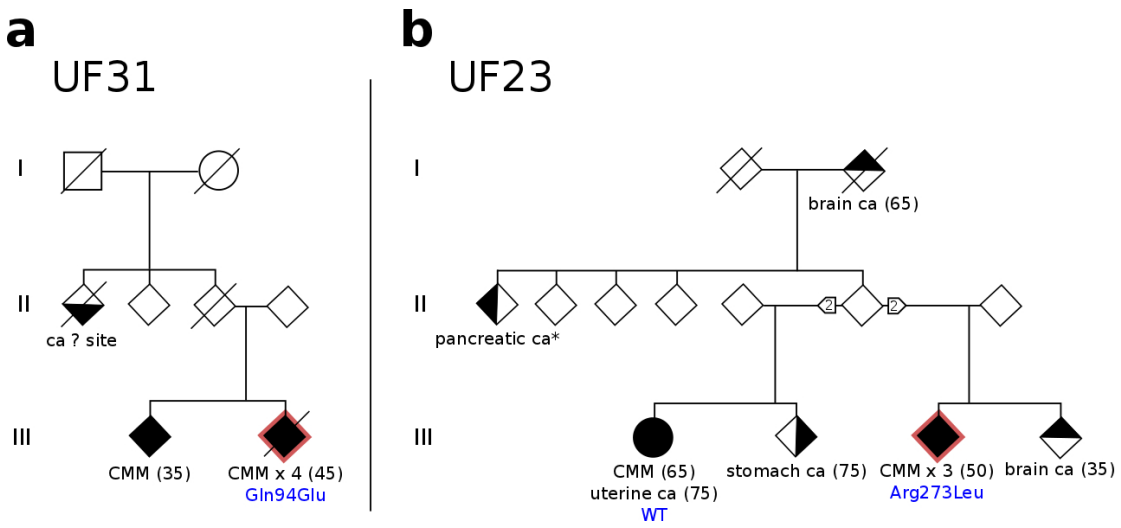


Supplementary Figure 1. Capillary sequencing of *POT1* variants detected by next-generation sequencing. One sample per variant is shown; variants are indicated with an arrow. **(a)** Sequencing showing the presence of the variant in an additional member of pedigree UF20 who was not exome-sequenced (individual III-2, Fig. 1a). **(b)** Sequencing of an affected individual of pedigree AF1; note that, in this case, the complementary strand is shown. **(c)** Sequencing of the *POT1* product in a control with wild-type *POT1* (top) and a carrier of the splice-acceptor variant (pedigree AF1, individual IV-1, Fig. 1a) (bottom). The boundary between exons 17 and 18 is marked with a dotted red line. The wild-type sequence, in nucleotides and in amino acids, is indicated in black, and the sequence the variant results in is indicated in red. The mutant sequence leads to the introduction of a premature stop codon 11 amino acids downstream of the exon 17-exon 18 boundary. **(d)** Sequencing of an affected member of pedigree UF31. **(e)** Sequencing of the carrier member of pedigree UF23.

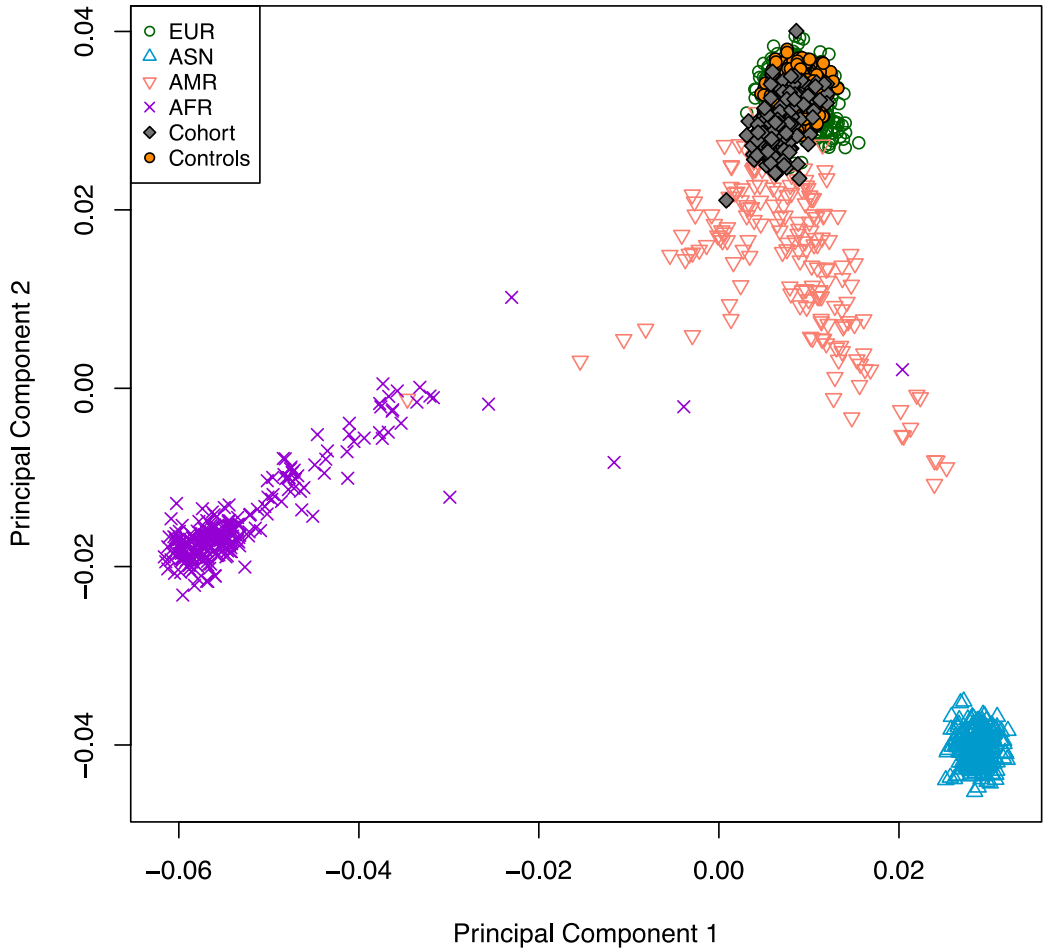
Nature Genetics: doi:10.1038/ng.2947



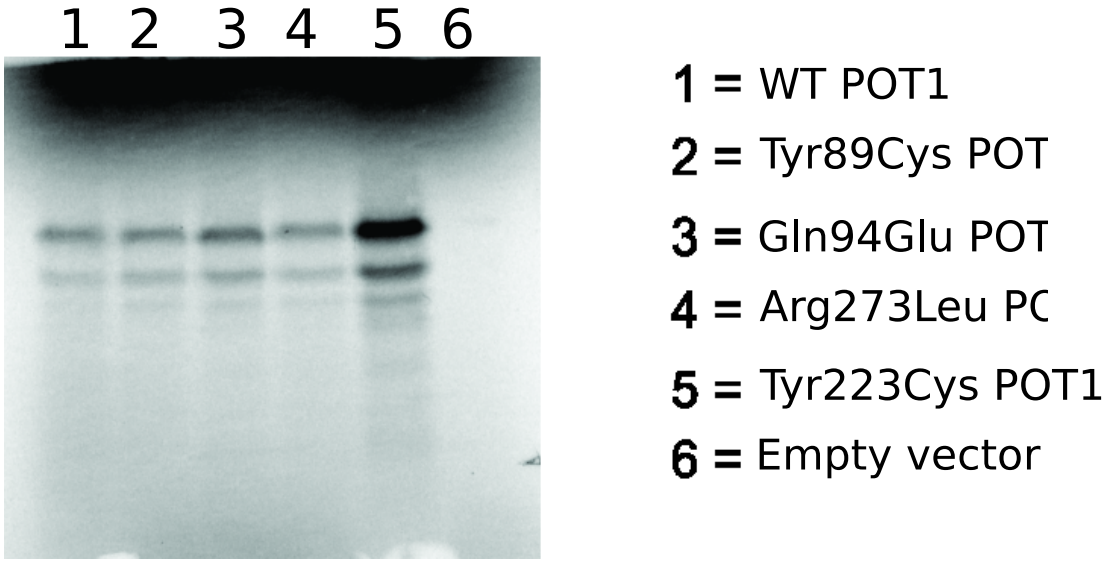
Supplementary Figure 2. MaxEntScan scores for the splice-acceptor variant detected in a familial melanoma pedigree (AF1). The location of the scores for the wild-type (black) and mutated (red) sequences are shown against score distributions for real splice donors, real splice acceptors and random genomic sequences.



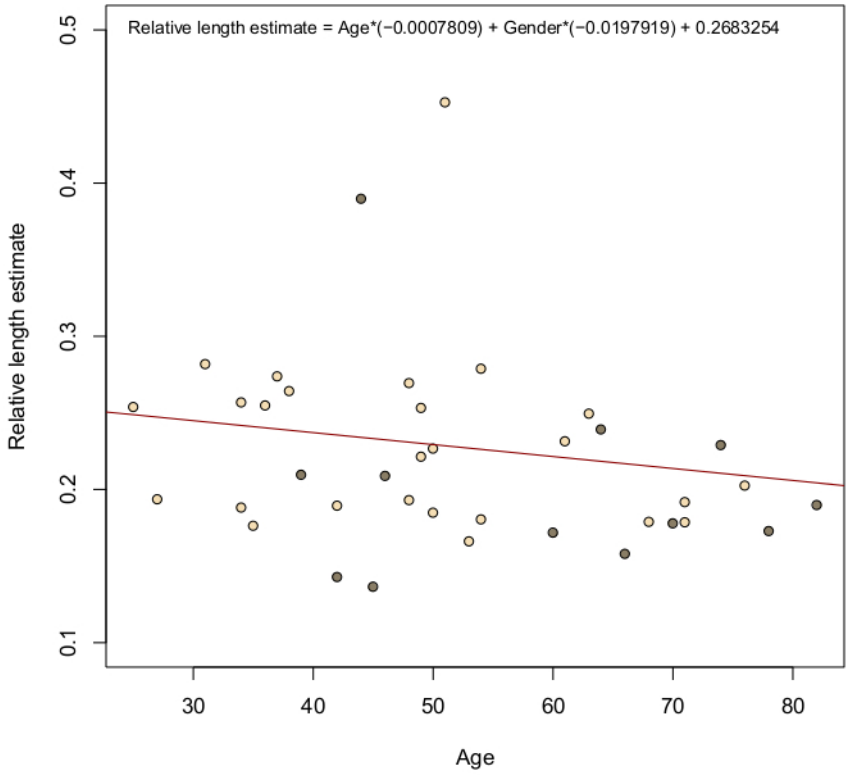
Supplementary Figure 3. Pedigrees with additional *POT1* variants. (a) Pedigree of family UF31 carrying the p.Gln94Glu change. **(b)** Pedigree of family UF23 carrying the p.Arg273Leu variant. Types of cancer are indicated under each symbol, and ages of onset are indicated in parentheses. Genotypes for all samples available for testing are shown in blue. All cancers were confirmed by histological analysis with the exception of one case (indicated by an asterisk). CMM, cutaneous malignant melanoma. Circles represent females; squares represent males; diamonds represent individuals of undisclosed sex. Half-filled symbols represent other cancers. The patients that were sequenced have a red outline.



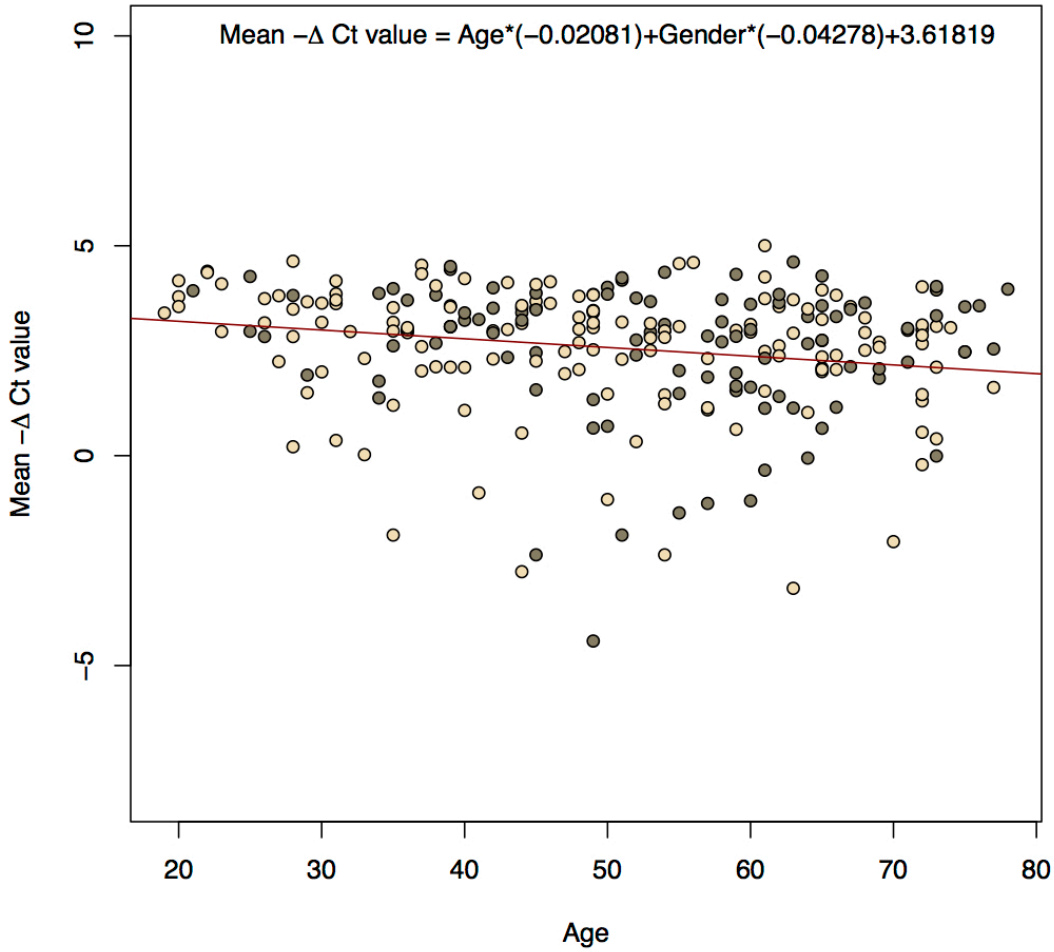
Supplementary Figure 4. Principal-component analysis showing that the melanoma cases and UK10K controls are ancestry matched. Plot showing the first and second principal components (PC1 and PC2, respectively). Ancestry was estimated using the 1000 Genomes individuals[1] and then projected onto the melanoma (gray) and UK10K control (orange) cohorts. Note that controls lying greater than 2 s.d. from the mean PC1 or PC2 scores, calculated using only European individuals in the 1000 Genomes Project data set, are not shown in this plot and were not considered in subsequent analyses. We did not depict three individuals from the QFMP cohort for whom we could not determine the zygosity of the called genotypes.



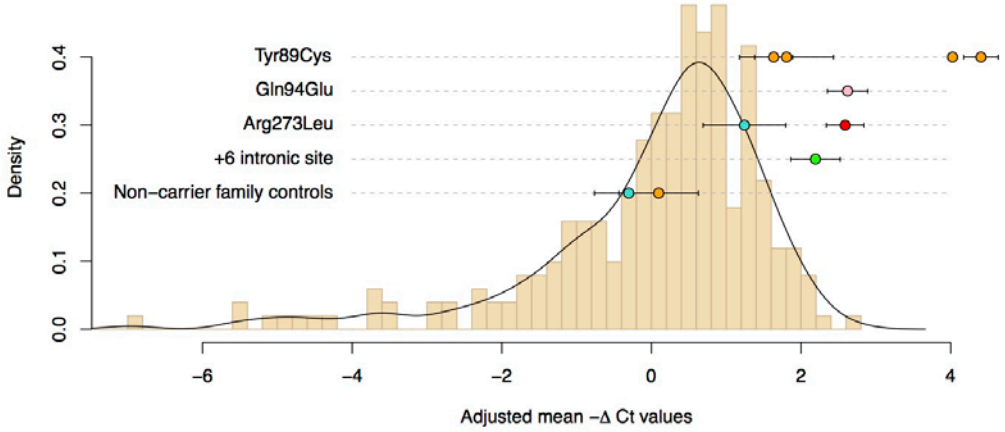
Supplementary Figure 5. ^{35}S gel showing the *in vitro* translation products of wild-type POT1 and OB domain mutants. This gel confirms that each *in vitro* translation reaction successfully produced protein for the electrophoretic mobility shift assay shown in **Figure 2b**. The p.Tyr89Cys, p.Gln94Glu and p.Arg273Leu POT1 variants were identified by exome-sequencing familial melanoma cases. The p.Tyr223Cys variant was somatically acquired in CLL and has previously been shown to be unable to bind to telomeric DNA[2]. The DNA-protein complexes shown in Figure 2b were visualized by ^{32}P labeling of $(TTAGGG)_3$ single-stranded DNA (Online Methods).



Supplementary Figure 6. Linear model used to adjust bioinformatically calculated telomere lengths for age at blood draw and sex. Residuals were used as the adjusted relative telomere lengths. Dark circles represent male samples, and light circles represent females. The sex variable is coded as 0 = female, 1 = male. Note that only two dimensions (relative telomere length and age) are shown.



Supplementary Figure 7. Linear model used to adjust PCR mean ΔC_t values for age at blood draw and sex. Residuals were used as the adjusted mean ΔC_t values. Dark circles represent male samples, and light circles represent females. The sex variable is coded as 0 = female, 1 = male. Note that only two dimensions (mean ΔC_t value and age) are shown.



Supplementary Figure 8. PCR-based estimate of telomere length showing the carrier of a *POT1* intronic splice donor variant. The carrier of the intronic splice donor variant is shown in the fourth row, in green. The rest of the figure is identical to **Figure 2d**.

Supplementary Tables

Supplementary Table 1. The number of pedigrees sequenced by Institute and sequencing centre. The number of pedigrees with *POT1* variants is indicated in parentheses.

Institutes	**Leeds (UK)**		**Leiden (NL)**	**QFMP (Australia)**	**Total**
Sequencing centres	**Sanger**	**BGI**	**Sanger**	**Macrogen**	
Familial pedigrees					
5+ cases	8 (1)	0	1	25 (1)	34 (2)
4 cases	11	1	2	5	19
3 cases	2	12	0	3	17
2 cases	0	18 (2)	0	2	20 (2)
Total	**21 (1)**	**31 (2)**	**3**	**35 (1)**	**90 (4)**
Single cases					
MPM*	0	4	0	0	4
Early age of onset* (<40 yr)	0	10	0	0	10
Different primaries	0	0	0	1[†]	1
Total	**0**	**14**	**0**	**1**	**15**
Total by Institute	**66 (3)**		**3**	**36 (1)**	**105 (4)**

*These samples were selected from the Leeds Melanoma Case-Control study. All cases were chosen for an absence of phenotypic risk markers (self reported sun-sensitivity and/or low mole count).

[†]This case was selected because they presented with three different primary cancers, one of which was melanoma.

Nature Genetics: doi:10.1038/ng.2947

Supplementary Table 2. Pedigrees sequenced as part of this study, with number of cases in pedigree and number of cases sequenced (exome or whole genomes) indicated*

Family ID	Origin (Sequencing centre)	Number of melanoma cases in pedigree	Total number of cases sequenced (exomes + genomes)	Number of whole exomes	Number of whole genomes
Familial melanoma pedigrees					
UF1	Leeds (Sanger)	4	2	2	0
UF2	Leeds (Sanger)	5	1	1	0
UF3	Leeds (Sanger)	4	1	1	0
UF4	Leeds (Sanger)	5	1	1	0
UF5	Leeds (Sanger)	4	1	1	0
UF6	Leeds (Sanger)	4	1	1	0
UF7	Leeds (Sanger)	4	2	2	0
UF8	Leeds (Sanger)	4	1	1	0
UF9	Leeds (Sanger)	5	1	1	0
UF10	Leeds (Sanger)	3	3	3	0
UF11	Leeds (Sanger)	4	1	1	0
UF12	Leeds (Sanger)	4	1	1	0
UF13	Leeds (Sanger)	4	1	1	0
UF14	Leeds (Sanger)	4	2	2	0
UF15	Leeds (Sanger)	8	2	2	0
UF16	Leeds (Sanger)	4	2	2	0
UF17	Leeds (Sanger)	6	1	1	0
UF18	Leeds (Sanger)	5	1	1	0
UF19	Leeds (Sanger)	6	2	2	0
UF20	Leeds (Sanger)	5	3	3	0
UF21	Leeds (Sanger)	3	2	2	0
UF22	Leeds (BGI)	2	1	1	0
UF23	Leeds (BGI)	2	1	1	0
UF24	Leeds (BGI)	4	1	1	0
UF25	Leeds (BGI)	3	1	1	0
UF26	Leeds (BGI)	3	1	1	0
UF27	Leeds (BGI)	3	1	1	0
UF28	Leeds (BGI)	3	1	1	0
UF29	Leeds (BGI)	3	1	1	0
UF30	Leeds (BGI)	3	1	1	0
UF31	Leeds (BGI)	2	1	1	0
UF32	Leeds (BGI)	2	1	1	0

UF33	Leeds (BGI)	3	1	1	0
UF34	Leeds (BGI)	2	1	1	0
UF35	Leeds (BGI)	3	1	1	0
UF36	Leeds (BGI)	2	1	1	0
UF37	Leeds (BGI)	3	1	1	0
UF38	Leeds (BGI)	3	1	1	0
UF39	Leeds (BGI)	2	1	1	0
UF40	Leeds (BGI)	3	1	1	0
UF41	Leeds (BGI)	2	1	1	0
UF42	Leeds (BGI)	2	1	1	0
UF43	Leeds (BGI)	2	1	1	0
UF44	Leeds (BGI)	2	1	1	0
UF45	Leeds (BGI)	3	1	1	0
UF46	Leeds (BGI)	2	1	1	0
UF47	Leeds (BGI)	2	1	1	0
UF48	Leeds (BGI)	2	1	1	0
UF49	Leeds (BGI)	2	1	1	0
UF50	Leeds (BGI)	2	1	1	0
UF51	Leeds (BGI)	2	1	1	0
UF52	Leeds (BGI)	2	1	1	0
NF1	Leiden (Sanger)	4	3	3	0
NF2	Leiden (Sanger)	4	2	2	0
NF3	Leiden (Sanger)	5	4	4	0
AF1	QFMP (Macrogen)	6	3	3	0
AF2	QFMP (Macrogen)	2	1	0	1
AF3	QFMP (Macrogen)	5	3	2	1
AF4	QFMP (Macrogen)	8	3	2	1
AF5	QFMP (Macrogen)	8	3	2	1
AF6	QFMP (Macrogen)	5	3	2	1
AF7	QFMP (Macrogen)	3	3	2	1
AF8	QFMP (Macrogen)	8	3	2	1
AF9	QFMP (Macrogen)	4	2	1	1
AF10	QFMP (Macrogen)	5	3	2	1
AF11	QFMP (Macrogen)	6	4	3	1
AF12	QFMP (Macrogen)	4	2	1	1
AF13	QFMP (Macrogen)	11	2	1	1
AF14	QFMP (Macrogen)	3	2	1	1
AF15	QFMP (Macrogen)	5	3	2	1
AF16	QFMP (Macrogen)	2	2	1	1
AF17	QFMP (Macrogen)	9	3	3	0
AF18	QFMP (Macrogen)	5	3	3	0
AF19	QFMP (Macrogen)	5	3	3	0

AF20	QFMP (Macrogen)	5	3	3	0
AF21	QFMP (Macrogen)	8	3	3	0
AF22	QFMP (Macrogen)	6	3	3	0
AF23	QFMP (Macrogen)	6	4	4	0
AF24	QFMP (Macrogen)	4	1	1	0
AF25	QFMP (Macrogen)	3	3	3	0
AF26	QFMP (Macrogen)	9	4	4	0
AF27	QFMP (Macrogen)	5	2	2	0
AF28	QFMP (Macrogen)	5	3	3	0
AF29	QFMP (Macrogen)	6	2	2	0
AF30	QFMP (Macrogen)	4	2	2	0
AF31	QFMP (Macrogen)	6	4	4	0
AF32	QFMP (Macrogen)	6	4	4	0
AF33	QFMP (Macrogen)	6	4	4	0
AF34	QFMP (Macrogen)	4	1	1	0
AF35	QFMP (Macrogen)	8	3	3	0
Total			***169***	***154***	***15***

Single cases

UN1	Leeds (BGI)	1	1	1	0
UN2	Leeds (BGI)	1	1	1	0
UN3	Leeds (BGI)	1	1	1	0
UN4	Leeds (BGI)	1	1	1	0
UN5	Leeds (BGI)	1	1	1	0
UN6	Leeds (BGI)	1	1	1	0
UN7	Leeds (BGI)	1	1	1	0
UN8	Leeds (BGI)	1	1	1	0
UN9	Leeds (BGI)	1	1	1	0
UN10	Leeds (BGI)	1	1	1	0
UN11	Leeds (BGI)	1	1	1	0
UN12	Leeds (BGI)	1	1	1	0
UN13	Leeds (BGI)	1	1	1	0
UN14	Leeds (BGI)	1	1	1	0
AN1	QFMP (Macrogen)	1	1	0	1
Total			***15***	***14***	***1***
Grand total			**184**	**168**	**16**

*Whole genomes were only sequenced as part of the QFMP cohort.

Supplementary Table 4. Primers used for validating *POT1* variants.

Gene	Variant Location	Forward Primer	Reverse Primer	Amplicon Size
POT1	7:124493077	TCTGGCAAGACCCTGATTCC	GCCTTCATACCAAACTTCAATCAA	112
POT1	7:124503684	GTGCGAGGTATGATAGGGGC	GCCCTCTCTTAATTATGCTGGC	216
POT1	7:124503670	GTGCGAGGTATGATAGGGGC	GCCCTCTCTTAATTATGCTGGC	216
POT1	7:124465412	TGGAGTTGAGACCAGCATTC	ATCCCTGCCACTCTCTTCAT	436

Supplementary Table 5: Details for pedigrees with *POT1* variants identified in this study

Pedigree	Source	Variant	Number of melanoma cases in pedigree	Study cohort	Identification technique
UF20	Sanger	Tyr89Cys	5	UK Familial Melanoma Study	Exome Sequencing
AF1	Macrogen	Splice acceptor	6	QFMP/Australia	Exome Sequencing
UF31	BGI	Gln94Glu	2	UK Familial Melanoma Study	Exome Sequencing
UF23*	BGI	Arg273Leu	2	UK Familial Melanoma Study	Exome Sequencing
CT1663	Leeds	Arg273Leu	1	Leeds Melanoma Case-Control study	Candidate variant†

*One patient with early onset disease and multiple primaries was sequenced and found to carry the Arg273Leu variant. A second case had a different clinical presentation (solitary melanoma *in situ*) in the 6th decade and did not carry the Arg273Leu variant.

†Assessment of POT1 variants in the Leeds Melanoma Case-Control Study (details above). A TaqMan assay was used for this analysis. A case with MPM and an early age of onset (48 years old) was identified. This variant was validated by PCR-resequencing.

Nature Genetics: doi:10.1038/ng.2947

Supplementary Table 6. Controls used in this study

Cohort	Description	No. samples	Phase
UK10K	This project aims to sequence 10,000 individuals with well-defined phenotypes (www.uk10k.org). All individuals from three cohorts with neurological phenotypes unrelated to cancer were chosen.	520	Frequency of *POT1* variants in an ancestry, technically matched population control
Leeds Melanoma Case-Control Series / WTCCC	Controls in this cohort are either part of the Leeds Melanoma Case Control Series (869), of which 499 are population-matched controls and 370 are unaffected individuals related to cases, or part of the Wellcome Trust Case Control Consortium (1,533).	2,402	Genotyping of the four variants found in familial melanoma families
NHLBI GO ESP	The NHLBI Grand Opportunity Exome Sequencing Project has sequenced and released 6,503 exomes from multiple cohorts with heart, lung and blood disorders (http://evs.gs.washington.edu/EVS/).	6,498*	Frequency of variants in residues in close proximity to DNA in a technically matched control

*This was the minimum number of exomes with coverage in each of the 24 amino acids.

Supplementary Table 7. Genomic location of bases encoding the 24 key OB domain residues (GRCh37)

Amino Acid	Genomic position
31	g.124532351-124532353
33	g.124532345-124532347
36	g.124532336-124532338
39	g.124532327-124532329
40	g.124532324-124532326
41	g.124532321-124532323
42	g.124532320 and g.124511094-124511095
48	g.124511076-124511078
60	g.124511040-124511042
62	g.124511034-124511036
87	g.124503689-124503691
89	g.124503683-124503685
94	g.124503668-124503670
159	g.124503473-124503475
161	g.124503467-124503469
223	g.124499044-124499046
224	g.124499041-124499043
243	g.124493166-124493168
245	g.124493160-124493162
266	g.124493097-124493099
267	g.124493094-124493096
270	g.124493085-124493087
271	g.124493082-124493084
273	g.124493076-124493078

Supplementary Table 8. Variants in *MC1R* in *POT1* pedigrees

Pedigree	Variants in known genes
UF20	One member carries chr. 16: g.89985844G>T (rs1805005, Val60Leu); two members carry g.89985940G>A (rs2228479, Val92Met).
UF31	The sole member sequenced carries g.89985844G>T (rs1805005, Val60Leu)
UF23	The individual that was exome-sequenced carries g.89985940G>A (rs2228479, Val92Met)
AF1	All three individuals carry g.89986117C>T (rs1805007, Arg151Cys)

Supplementary Note

Analysis of a variant at the +6 intronic splice donor of *POT1* in a familial melanoma pedigree (g.124467262A>C)

Here we focused on the identification of familial melanoma cases with missense variants in the OB domains of POT1 and also deleterious variants throughout the gene. As part of our analysis we did, however, identify a variant at +6 in the consensus splice donor sequence of intron 17-18 in one of the families with one member sequenced. Transcriptome sequencing of lymphocyte RNA from this patient (deposited at the European Genome-phenome Archive, accession number EGAS00001000571) failed to reveal evidence of aberrant splice forms, although expression of *POT1* appeared to be low, and MaxEntScan analysis of the splice site sequence (as described above) suggested that the variant was unlikely to be deleterious. We did note, however, that the telomeres of the case carrying this +6 variant appeared longer than the telomeres of controls, suggesting some effect on telomere regulation (**Supplementary Fig. 8**).

References

1. 1000 Genomes Project Consortium *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).
2. Ramsay, A.J. *et al.* POT1 mutations cause telomere dysfunction in chronic lymphocytic leukemia. *Nat Genet* (2013).

Chapter 5

Nonsense mutations in the shelterin complex genes *ACD* and *TERF2IP* in familial melanoma

Manuscript 4

Lauren G. Aoude*, Antonia L. Pritchard*, Carla Daniela Robles-Espinoza*, Karin Wadt*, Mark Harland*, Jiyeon Choi, Michael Gartside, Víctor Quesada, Peter Johansson, Jane M. Palmer, Andrew J. Ramsay, Xijun Zhang, Kristine Jones, Judith Symmons, Elizabeth A. Holland, Helen Schmid, Vanessa Bonazzi, Susan Woods, Ken Dutton-Regester, Mitchell S. Stark, Helen Snowden, Remco van Doorn, Grant W. Montgomery, Nicholas G. Martin, Thomas M. Keane, Carlos López-Otín, Anne-Marie Gerdes, Håkan Olsson, Christian Ingvar, Åke Borg, Nelleke A. Gruis, Jeffrey M. Trent, Göran Jönsson, D. Timothy Bishop, Graham J. Mann, Julia A. Newton-Bishop†, Kevin M. Brown†, David J. Adams†, Nicholas K. Hayward†. Nonsense mutations in the shelterin complex genes *ACD* and *TERF2IP* in familial melanoma. *Journal of the National Cancer Institute* (in press).

*joint first author; †joint last author

5.1 Relevance to thesis aims

Following the finding of inactivating germline mutations in *POT1* in familial melanoma (Chapter 4) we searched for mutations in the other five shelterin components (*ACD*, *TERF1*, *TERF2*, *TERF2IP* and *TINF2*). We interrogated whole-genome, exome and targeted sequencing data for novel germline variants in 510 melanoma families from the QFMP and GenoMEL. We found novel nonsense mutations in *ACD* and *TERF2IP* which segregated with melanoma in a seven-case Queensland family and a three-case British family respectively. Additionally, novel missense mutations were found in these two genes in other probands, so that overall 10 families, from Australia, the UK, Sweden and Denmark, carried potentially deleterious *ACD* and *TERF2IP* mutations. Interestingly, these cases come from pedigrees enriched with other malignancies, suggesting these mutations may represent a general mechanism of tumour susceptibility. This study addressed Aims 2 to 4 of my thesis and adds to the growing support for telomere dysregulation as a key process associated with melanoma susceptibility.

The shelterin complex is responsible for all normal telomere functions, including the protection of telomeres from degradation, illegitimate recombination events and from being inappropriately processed by the DNA-repair pathway [132]. This is an important finding as it implicates other members of the shelterin protein complex in melanoma predisposition. Along with the findings from Chapter 4, we see three members of this complex implicated in melanoma predisposition in families from Australia, the UK, Sweden and Denmark. Importantly, mutations in *ACD, TERF2IP* and *POT1* collectively account for approximately 9% of high-density melanoma families that lack mutations in the known familial melanoma genes (*CDKN2A, CDK4, BAP1, MITF* and the *TERT* promoter), making the shelterin complex the second most mutated pathway in familial melanoma.

5.2 Contribution of candidate

As with the *POT1* study described in the previous chapter, this was a large collaborative effort between international research groups and therefore there are elements contained within this manuscript that were outside the scope of this thesis. I performed the analysis of the whole-genome and exome data from Australia, Sweden and Denmark as well as the validation of variants using Sanger sequencing methods. I performed the co-segregation analysis of mutations in the Australian, Swedish and Danish families. I was also responsible for the sample preparation for the Australian families including the extraction of DNA and RNA from whole blood and the co-segregation analysis. I wrote and edited the manuscript with Nicholas Hayward and Antonia Pritchard. I also compiled the majority of the tables and figures for the body of the manuscript and the supporting information.

5.3 Contribution of other authors

Nicholas K. Hayward, David J. Adams, Kevin M. Brown, Julia A. Newton Bishop, D. Timothy Bishop and Mark Harland designed the study. Samples were provided by Grant W. Montgomery, Nicholas G. Martin, Nicholas Hayward, Helen Snowden, Julia A. Newton Bishop, D. Timothy Bishop, Graham J. Mann, Anne-Marie Gerdes, Håkan Olsson, Christian Ingvar, Åke Borg, Remco van Doorn and Nelleke A. Gruis. I performed the whole-genome and exome data analysis and experiments for Australian samples in conjunction with Antonia L. Pritchard, Mitchell S. Stark Vanessa Bonazzi, Susan Woods, Ken Dutton-Regester and Michael G. Gartside. Carla Daniela Espinoza-Robles performed the NGS data-analysis for the UK families and the control samples. Xijun Zhang and Kristine Jones performed the Ion Torrent targeted sequencing while Jiyeon Choi

and Kevin M. Brown analysed the Ion Torrent data. Peter Johansson, Zhihao Ding and Thomas M. Keane provided bioinformatics support. Protein modelling was performed by Víctor Quesada, Andrew J. Ramsay and Carlos López-Otín. Karin Wadt recruited the Danish families and collected samples and clinical information. Jane M. Palmer, Judith Symmons, Elizabeth A. Holland and Helen Schmid contacted the families and collected the patient data and samples for the Australian cases. Nicholas Hayward and Antonia helped write the manuscript with the input of Carla Daniela Robles-Espinoza and David J. Adams. All authors read and approved the manuscript.

Nonsense mutations in the shelterin complex genes *ACD* and *TERF2IP* in familial melanoma

Lauren G. Aoude[1*], Antonia L. Pritchard[1*], Carla Daniela Robles-Espinoza[2*], Karin Wadt[3*], Mark Harland[4*], Jiyeon Choi[5], Michael Gartside[1], Víctor Quesada[6], Peter Johansson[1], Jane M. Palmer[1], Andrew J. Ramsay[6], Xijun Zhang[7], Kristine Jones[7], Judith Symmons[1], Elizabeth A. Holland[11,12], Helen Schmid[11,12], Vanessa Bonazzi[1], Susan Woods[1], Ken Dutton-Regester[1], Mitchell S. Stark[1], Helen Snowden[4], Remco van Doorn[8], Grant W. Montgomery[1], Nicholas G. Martin[1], Thomas M. Keane[2], Carlos López-Otín[6], Anne-Marie Gerdes[3], Håkan Olsson[9], Christian Ingvar[9], Åke Borg[9], Nelleke A. Gruis[8], Jeffrey M. Trent[10], Göran Jönsson[9], D. Timothy Bishop[4], Graham J. Mann[11,12], Julia A. Newton-Bishop[4†], Kevin M. Brown[5†], David J. Adams[2†], Nicholas K. Hayward[1†]

[1]QIMR Berghofer Medical Research Institute, Brisbane, QLD 4006, Australia

[2]The Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK

[3]Department of Clinical Genetics, Rigshospitalet, Copenhagen, Denmark

[4]University of Leeds, Leeds, LS9 7TF, UK

[5]Laboratory of Translational Genomics, National Cancer Institute, Bethesda, MD, USA

[6]Departamento de Bioquímica y Biología Molecular, Instituto Universitario de Oncología del Principado de Asturias (IUOPA) Universidad de Oviedo, Oviedo, Spain

[7]Cancer Genomics Research Laboratory, NCI Frederick, MD, USA

[8]Department of Dermatology, Leiden University Medical Centre, Leiden, The Netherlands

[9]Department of Clinical Sciences, Lund University, Lund, Sweden

[10]Translational Genomics Institute, Phoenix, AZ, USA

[11]University of Sydney at Westmead Millennium Institute, Sydney, NSW, Australia

[12]Melanoma Institute Australia, North Sydney, NSW, Australia

*These authors contributed equally

†These authors contributed equally

Corresponding author: Lauren Aoude, QIMR Berghofer Medical Research Institute, 300 Herston Road, Herston, Brisbane, QLD 4006, Australia, phone +61 7 3362 0323, fax +61 7 3845 3508

Abstract

Background: The shelterin complex protects chromosomal ends by regulating how the telomerase complex interacts with telomeres. Following the recent finding in familial melanoma of inactivating germline mutations in POT1, encoding a member of the shelterin complex, we searched for mutations in the other five components of the shelterin complex in melanoma families.

Methods: Next generation sequencing techniques were used to screen 510 melanoma families (with unknown genetic aetiology) and control cohorts, for mutations in shelterin complex encoding genes: ACD, TERF2IP, TERF1, TERF2 and TINF2.

Results: Six families had mutations in ACD and four families carried TERF2IP variants, which included nonsense mutations in both genes (p.Q320X and p.R364X, respectively) and point mutations that co-segregated with melanoma. Of five distinct mutations in ACD, four clustered in the POT1 binding domain, including p.Q320X. This clustering of novel mutations in the POT1 binding domain of ACD was statistically higher ($p = 0.005$) in melanoma probands compared to population controls ($n = 6785$), as were all novel and rare variants in both ACD ($p = 0.040$) and TERF2IP ($p = 0.022$). Families carrying ACD and TERF2IP mutations were also enriched with other cancer types, suggesting that these variants also predispose to a broader spectrum of cancers than just melanoma. Novel mutations were also observed in TERF1, TERF2 and TINF2, but these were not convincingly associated with melanoma.

Conclusions: Our findings add to the growing support for telomere dysregulation as a key process associated with melanoma susceptibility.

Introduction

Germline mutations in the high penetrance melanoma susceptibility genes *CDKN2A* and *CDK4* account for cutaneous malignant melanoma (CMM) development in ~40% of multi-case families [1]. Additionally, rare mutations in *BAP1* have been associated with uveal and cutaneous melanoma predisposition [2]. More recently, germline mutations in the promoter of *TERT* (telomerase reverse transcriptase) [3], as well as inactivating mutations in the shelterin component *POT1* (protection of telomeres 1) [4, 5], implicate telomere dysregulation as a novel pathway underlying familial melanoma.

Shelterin is a telomere-specific protein complex that protects the ends of chromosomes by mediating the interaction of telomerase with telomeres. It is made up of six family members, encoded by the genes *POT1*, *ACD*, (adrenocortical dysplasia protein homolog; also known as *TPP1*, *TINT1*, *PIP1* and *PTOP*), *TERF1* (telomeric repeat binding factor 1, also known as *TRF1*), *TERF2* (telomeric repeat binding factor 2, also known as *TRF2*), *TERF2IP* (telomeric repeat binding factor 2 interacting protein; also known as *RAP1* and *DRIP5*) and *TINF2* (TERF1-interacting nuclear factor 2; also known as *TIN2*) (**Supplementary Figure 1**). The shelterin components are collectively necessary for all telomere functions, which include the protection of telomeres from degradation, aberrant recombination, from being inappropriately processed by the DNA-repair pathway, and also the facilitation of chromosome capping to mediate telomerase activity [6]. Thus, protein-altering variants located within this complex have recently been shown to have a great impact on diseases related to cellular lifespan, particularly cancer [4, 5, 7].

Here, following on from recent reports of germline *POT1* mutations in familial melanoma [4, 5], we establish a key role for other components of the shelterin complex in susceptibility to CMM.

Methods

We screened for germline ACD, TERF2IP, TERF1, TERF2 and TINF2 variants in exome, whole-genome, or targeted pulldown sequence data from 601 individuals belonging to 510 families with CMM.

Samples used for whole-genome, exome or targeted sequence analysis

All cases gave written informed consent for participation. Each was wild type for CDKN2A, BAP1, POT1, BRCA2, CDK4 and the TERT promoter. Samples were ascertained through the Queensland Familial Melanoma Project (QFMP) [8, 9], the Sydney Genetic Epidemiology of Melanoma study [10], the UK Familial Melanoma Study, the Leiden University Medical Centre; the Danish Project of Hereditary Malignant Melanoma and the Oncogenetic Clinic at Skåne University Hospital.

Ethics approval was granted by the Committee of Biomedical Research Ethics of the Capital Region of Denmark, the UK Multicentre Research Ethics Committee, and the Human Research Ethics Committees of: the QIMR Berghofer Medical Research Institute, the Lund University, the University of Sydney, and the Leiden University Medical Centre.

Next Generation Sequencing

Whole-genome or exome sequencing was performed on 113 CMM families from Australia, UK, The Netherlands, Denmark and Sweden. Between 1 and 5 cases were sequenced from each family, totalling 204 individuals. Supplementary Table 1 details samples by centre. Supplementary Table 2 summarises ages of CMM onset and other cancers in individuals sequenced. Sequencing was performed on the Illumina Hiseq 2000 platform with Agilent SureSelect Human All Exon V4+UTRs enrichment kits. Paired-end reads of 75-100 bp were generated, with mean coverage of 60-96X. The Burrows-Wheeler Aligner was used to map reads to the UCSC hg19 [11]. UK and Leiden samples, were filtered for duplicate reads using Picard [12], recalibrated, and aligned using Genome Analysis Tool Kit [13]. SNPs and indels were annotated using bcftools and SAMtools mpileup with disabled BAQ computation [14]. Data were filtered using: quality score > 40, alternate reads > 2, and alternate reads $\geq$ 20% of total reads. Variants in dbSNP135, or the 1000 Genomes Project (April 2012) were removed.

Targeted sequencing of the shelterin genes was carried out in 397 QFMP probands (Supplementary Table 3) using two Ampliseq panels. The first panel (105 amplicons; 10202 bp) included the coding regions and had 100% coverage except TERF1 (96%). The second panel (68 amplicons; 7723 bp) comprised of UTRs, promoters, and alternative exons. Libraries were barcoded and run on an Ion PGM using 318 chips, with minimum 30X coverage. Sequence reads were processed by the Ion

Torrent Suite (Life Technologies, CA, USA) with alignment and variants called as above. Variants were filtered to exclude those listed in dbSNP, the 1000 Genomes Project, Kaviar and synonymous changes. Variants had to have a quality score ≥ 30 and ≥ 10 alternate reads.

Sanger sequencing

Sanger sequencing was used to confirm variants found by next-generation sequencing. Primers are listed in Supplementary Table 4.

Sequenom iPLEX

Seven variants (ACD: p.N249S, p.A200T, p.Q320X, p.I322F; TERF2IP: p.M5I, p.D10H, p.R364X) were genotyped in an Australian case-control sample. Cases (n = 1669) were derived from the QFMP [8, 15] and the controls (n = 1590) were parents of twins ascertained in the Brisbane Twin Naevus Study (BTNS) [16]. The controls self-reported their melanoma history and had not developed CMM at the time of sample collection. The Sequenom iPLEX gold system (Sequenom Bioscience, CA, USA) was used to genotype the variants; primers were designed using Assay Design Suite (Supplementary Table 4).

Analysis of novel variants in shelterin genes in control exomes

Publicly available exome data from 1965 Danish individuals [17], or a subset of exomes from the UK10K sequencing project (http://www.uk10k.org), or the European-American samples lodged in the National Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project (ESP) database (http://evs.gs.washington.edu/EVS/) were also used as controls.

Statistical Analyses

Maximum likelihood analysis

The age of the founder mutation was estimated using a statistical model described by Neuhausen and colleagues [18]. Assuming 1 cM equals 1 Mb and a de novo mutation rate of 1.2 x 10-8 per generation [19], a joint likelihood of the genotype data was calculated taking into account ancestral haplotype, number of generations, G, since the ancestor, and allele frequencies of the SNPs among the European population. An estimate of G was calculated that maximized the likelihood. A confidence interval was calculated by finding the range of values of G that yielded a likelihood of at least one tenth of the maximum likelihood.

LOD score analysis

An autosomal dominant model was used to generate the LOD score for families carrying the ACD p.N249S variant. Parametric linkage analysis was done using the Genehunter MOD score algorithm. Penetrance was specified as 5%, 95% and 95% for the three genotype classes, with allele frequency at the disease and marker locus specified as 0.001.

Mutation clustering

Chi-squared tests were used to determine whether mutations were significantly enriched in melanoma cases versus controls and whether they clustered more often in the POT1 binding domain of ACD. Fisher's Exact test was used where any value was < 5. A p-value of < 0.05 was considered statistically significant (one-tailed with Yates' correction).

Results

Novel germline ACD and TERF2IP mutations in melanoma

Segregating nonsense mutations in ACD and TERF2IP were found in a five-case and a four-case family, respectively. The p.Q320X mutation in ACD, in family AUS1, was present in all four cases available for genotyping (Figure 1) and results in a truncated protein, disrupting the POT1 binding domain and eliminating the TINF2 binding domain downstream (Figure 2A). Three CMM cases in family UK1 were carriers of a p.R364X nonsense mutation in TERF2IP (Figure 1). This results in truncation of the protein 36 amino acids from the C-terminus, disrupting the TERF2 binding domain (Figure 2A).

In addition, we found novel (not in dbSNP135, or the 1000 Genomes Project data at the time of analysis) missense mutations in ACD and TERF2IP that co-segregated in all available invasive CMM cases in four other families (Table 1, Figure 1). This included a p.N249S mutation in ACD in two separate families, which like the p.Q320X mutation, also occurs in the POT1 binding domain (Figure 2A). This mutation segregated in all seven available cases of AUS2, with eight confirmed and four unconfirmed cases of CMM. Danish family, DK1, which harboured the same mutation (Figure 1), presented with five confirmed cases and one unconfirmed case of melanoma. Of the individuals available for analysis, three invasive CMM cases were found to harbour this mutation; a person diagnosed with a melanoma in situ was not a carrier. A second novel mutation, which co-segregated with all three melanoma cases in family AUS3, was identified in the POT1 binding domain of ACD (p.V272M). Unfortunately, as the crystal structure of the POT1 binding domain of ACD, containing the p.N249 and the p.V272 residues, has not been resolved, there is insufficient information for in silico modelling of the effects of these variants. Another completely co-segregating mutation we discovered was in the MyB domain of TERF2IP (p.Q191R) in a two-case melanoma family (Table 1, Figure 1).

Additionally, four novel variants were observed in ACD or TERF2IP, which did not fully segregate with all melanoma cases in the respective families (Supplementary Figure 2A). These included a p.A200T substitution within the OB domain of ACD, a p.I322F substitution within the POT1 binding domain of ACD, as well as p.M5I and p.D10H substitutions, both of which lie in the BRCT domain of TERF2IP (Figure 1 and Supplementary Figure 2A). All of these co-segregating missense mutations in ACD and TERF2IP occur at highly evolutionarily conserved sites across species (Figure 2, B and C).

Prevalence, linkage and haplotype analysis of the ACD p.N249S mutation

The ACD p.N249S variant was the only novel mutation at that time of initial analysis of our exome data that was subsequently observed in the ESP database, albeit exceedingly rarely (3/8600 European-American chromosomes). We therefore additionally screened an Australian population-based case-control panel for this variant, and did not see it in any of 1669 cases or 1590 controls; it has also not been seen in the 1000 Genomes Project data, or in the exomes from ~1000 Danish diabetes cases and ~1000 metabolically healthy controls [17].

Linkage analysis of the two families segregating the ACD p.N249S variant gave a combined LOD score of 1.14, which equates to a p-value of 0.011. Analysis of a possible common founder in these families was carried out using whole-genome SNP arrays. Each of the carriers tested shared an allele for 112 SNPs, stretching from rs12918121 (chr16:67187795) to rs16957597 (chr16:67946356), a region 758561 bp long, spanning ACD (Supplementary Table 5). The data are thus consistent with all affected individuals sharing the same haplotype. Estimates of when the mutation arose gave a maximum likelihood for 129 generations ago (with a 90% confidence interval of 28-362 generations).

ACD and TERF2IP variants in controls

To further assess the association between ACD and TERF2IP variants and familial melanoma susceptibility, we screened 6785 publicly available control exomes for rare (variant allele frequency (VAF) < 0.001) variants in these genes. Controls included the ESP cohort (n = 4300), Danish controls (n = 1965) [17] and the UK10K sequencing project (http://www.uk10k.org; n = 520). ACD and TERF2IP variants in these control cohorts are listed in Supplementary Tables 6 – 8 and the relative positions with respect to protein domains given in Supplementary Figure 3. Taking all ACD variants with a VAF < 0.001, we found significant enrichment in melanoma cases versus controls (6/510 melanoma probands versus 33/6785 controls; p = 0.040). Moreover, if only novel ACD variants were considered the association with melanoma was stronger (6/510 melanoma probands versus 8/2485 controls (note ESP controls were not included since by definition variants in the ESP were not considered novel); p = 0.013). The most significant association was found when we only considered variants with a VAF < 0.001 that occurred in the POT1 binding domain; 5/510 CMM probands versus 16/6785 controls (p = 0.005).

No novel or rare variants were observed in TERF2IP in the 520 UK10K samples or the 1965 Danish controls. There were 8 variants (12 individuals) listed amongst the 4300 European-

American ESP controls. Thus there was significant enrichment ($p = 0.022$) of novel plus rare (VAF < 0.001) TERF2IP variants in melanoma probands (4/510) versus controls (12/6785).

Lastly, to assess the potential contribution of ACD and TERF2IP to sporadic melanoma we conducted a case-control analysis for a total of six variants (other than ACD p.N249S reported above) in these two genes (ACD: p.A200T, p.Q320X, p.I322F; TERF2IP: p.M5I, p.D10H, p.R364X) and did not observe any other individual that carried one of these variants amongst 1669 QFMP cases and 1590 BTNS controls.

Germline variants in other shelterin genes

Novel variants were also observed in TERF1, TERF2 and TINF2 (Supplementary Table 9), but these were not convincingly associated with melanoma (see Supplementary Results). Supplementary Table 10 lists incidences of all cancers in families carrying shelterin mutations.

Discussion

The aim of this study was to screen individuals with strong personal or family histories of CMM that do not carry a mutation in one of the known high penetrance melanoma risk genes, for germline mutations in *ACD, TERF1, TERF2, TERF2IP* and *TINF2*. We identified ten mutations in *ACD*/*TERF2IP*, including two fully segregating nonsense mutations and four novel co-segregating missense mutations.

ACD, in a subunit with POT1, mediates the interaction between shelterin and TERT through its oligonucleotide/oligosaccharide-binding folds (OB-folds); **Figure 2A** [20]. When the ACD/POT1 subunit is inhibited, the telomerase complex increases telomere length, indicating that this sub-complex is required to inhibit the elongation of chromosome ends [21-25]. Additionally, within the shelterin complex, ACD links POT1 to other members (**Supplementary Figure 1**) and increases the affinity of POT1 for telomeric single stranded DNA [26]. We identified a nonsense mutation in ACD, p.Q320X, which disrupts the POT1 binding domain and abolishes the TINF2 binding domain, so would therefore be predicted to result in an unformed shelterin complex. Individuals in this family presented with early onset melanoma at: 23, 25, 29, 39 and 56 years (**Figure 1**). The p.N249S mutation in ACD was identified in 2 families, from Australia and Denmark, which shared a large founder haplotype across the ACD locus. A co-segregating ACD p.V272M mutation was identified in a lower density melanoma family (**Figure 1**). Two *ACD* mutations were identified that did not fully-segregate with all CMM cases in the family; p.A200T was found in four of seven cases that could be tested from an eight -case family and p.I322F was found in three of four available cases from a six-case family (**Supplementary Figure 2A**). Missense mutations p.A200T, p.V272M and p.I322F were predicted to be damaging or possibly damaging by the Polyphen2 and SIFT prediction programs (**Table 1**). Overall, of the five distinct mutations in ACD, four clustered in highly conserved residues in the POT1 binding domain (**Figure 2A**) and are enriched for occurrence in melanoma cases compared to controls ($p = 0.005$), indicating this domain plays an important role in melanoma susceptibility.

TERF2IP, associates with the shelterin complex via its C-terminus to a central region of TERF2, forming a stable 1:1 complex. TERF2IP, as part of the shelterin complex, is vital for the repression of homology-directed repair of double strand chromosomal break at the telomere [27]. *TERF2IP* p.R364X results in premature truncation of the protein 36 amino acids from the C-terminus, resulting in disruption of the TERF2 binding domain and is therefore predicted to result in a loss of binding to the shelterin complex. Of the three novel missense variants observed in *TERF2IP*, p.Q191R was found in the two cases in family AUS6, who both developed CMM at an early age

(15 and 24 years) and p.D10H was predicted to be damaging/probably damaging by Polyphen2 and SIFT (**Table 1**); all three missense mutations occur at highly conserved amino acid residues (**Figure 2, B and C**).

Many families harboring mutations in *ACD*/*TERF2IP* included members with multiple primary melanomas (MPM) and other cancer types (**Figure 1**; **Supplementary Figure 2A**). The *TERF2IP* nonsense mutation, in family UK1, was found in an individual without CMM, but who developed breast cancer at age 85. Family AUS2, harboring the ACD p.N249S mutation, included six individuals with MPM and/or early onset melanoma (15, 26 and 35 years), four CMM cases also developed other cancers (three lung and one breast) and a mutation carrier without CMM developed breast cancer aged 50. In family DK1, harboring ACD p.N249S, two mutation carriers developed MPM, one of whom also developed B-cell lymphoma aged 82. The ACD p.V272M mutation occurred in a family (AUS3) with three cases of CMM, all of whom developed other cancers; two mutation carriers developed three different primary cancers: CMM, colon and lung; or CMM, bowel and leukemia. The two TERF2IP p.Q191R carriers in family AUS6 had both CMM and cervical cancer. The p.M5I missense mutation occurred in a sporadic melanoma case, with bilateral ovarian cancer at 77 years, and meningioma aged 78. Two ACD p.A200T carriers had CMM and prostate cancer and two other carriers had MPM. Finally, a carrier of the ACD p.I322F mutation in family AUS5 had MPM. Taken together, these data are strongly suggestive that mutations in *ACD* and *TERF2IP* are associated with early onset CMM, and MPM, and may predispose to a broader spectrum of cancers than just melanoma. Limitations of this study are that numbers of these other tumour types are too low to determine whether they are robustly associated with germline mutations in *ACD* and *TERF2IP* and that tumour blocks were not available on these cancers, or melanomas, to determine whether loss of heterozygosity is required for tumorigenesis.

In summary, the loss-of-function mutations we report here in *ACD* and *TERF2IP*, along with those previously published in *POT1,* suggest that multiple components of the shelterin complex play a role in melanoma predisposition. Collectively, mutations in *ACD, TERF2IP* and *POT1* account for ~9% (12/132) of high-density melanoma families (≥ 3 CMM cases) lacking mutations in *CDKN2A, CDK4*, *TERT* and *BAP1*. Given that the shelterin complex directly interacts with the product of *TERT*, a recently reported melanoma predisposition gene, the evidence we document here indicates that dysregulated telomere maintenance is a key pathway controlling melanoma development.

URLs

1000 Genomes project (April 2012 release), http://browser.1000genomes.org/
dbSNP135, http://www.ncbi.nlm.nih.gov/projects/SNP/
EGA, https://www.ebi.ac.uk/ega/
Multiple SeqDoC, http://research.imb.uq.edu.au/seqdoc/multi.html
National Heart, Lung, and Blood Institute Exome Sequencing Project,
http://evs.gs.washington.edu/EVS/
UK10K Sequencing Project, http://www.uk10k.org/
VCFtools, http://bioinformatics.oxfordjournals.org/content/27/15/2156

Accession codes

Sequence data have been deposited at the European Genome-phenome Archive (EGA) hosted by the European Bioinformatics Institute under accession EGAS (to be advised).

Acknowledgements

L.G.A. and K.W. are supported by PhD scholarships from the Australia and New Zealand Banking Group Limited Trustees and Rigshospitalet, University Hospital of Copenhagen, respectively. A.L.P. is supported by Cure Cancer Australia. N.K.H. and G.W.M. are supported by fellowships from the National Health and Medical Research Council of Australia (NHMRC). D.J.A, C.D.R-E, J.C.T, M.P and T.M.K were supported by Cancer Research UK and the Wellcome Trust. C.D.R-E was also supported by Consejo Nacional de Ciencia y Tecnología of Mexico. M.H. J.A.N-B, D.T.B were supported by Cancer Research UK. N.A.G. and RvD were supported by a grant from the Dutch Cancer Society (UL 2012-5489). G.J, Å.B, H.O and C.I were supported by the Swedish Research Council, the Swedish Cancer Society and the Nordic Cancer Union. H.O was also supported by ERC. K.M.B. and J.C. are supported by the Division of Cancer Epidemiology and Genetics, National Cancer Institute. C.L-O. is an Investigator of the Botín Foundation and work at his laboratory is supported by grants from the Spanish Ministry of Economy and Competitiveness and Red Temática de Investigación del Cáncer (RTICC). V.Q. is a Ramón y Cajal fellow supported by the Consolider-Ingenio RNAREG Consortium. The work was funded in part by the NHMRC, Cancer Institute NSW, Cancer Council Queensland, the Melanoma Research Alliance, and the Aase

and Ejnar Danielsens fund. We are grateful to Dr Stuart MacGregor for help with the LOD score analysis and to Lotte Krogh, Department of Clinical Genetics, University hospital of Odense, Denmark, for her valued input. We are indebted to the melanoma patients and their families for their participation in this study.

Author contributions

Study concept and design: D.J.A., K.M.B., J.N.B., D.T.B., J.M.T., G.J. and N.K.H.
Acquisition of data: L.G.A., A.L.P., C.D.R-E., K.W., M.H., J.C., M.G., J.M.P., X.Z., K.J., J.S., E.A.H., H.Sc., V.B., S.W., K.D-R., M.S.S., and R.vD.
Sample contribution: J.N.B., D.T.B., G.J.M., N.G.M., G.W.M., N.A.G., A-M.G., H.O., C.I., A.B. and N.K.H.
Data analysis: J.C., P.J., V.Q., A.R., C.L-O., and N.K.H.
Bioinformatics: P.J., C.D.R-E., H.Sn., and T.M.K.
Drafting of the manuscript: L.G.A., A.L.P., C.D.R-E., D.J.A., K.M.B., and N.K.H.

Competing financial interests

The authors declare there were no competing financial interests.

References

1. Goldstein AM, Chan M, Harland M, et al. High-risk melanoma susceptibility genes and pancreatic cancer, neural system tumors, and uveal melanoma across GenoMEL. Cancer Res 2006;66(20):9818-28.
2. Wiesner T, Obenauf AC, Murali R, et al. Germline mutations in BAP1 predispose to melanocytic tumors. Nat Genet 2011;43(10):1018-21.
3. Horn S, Figl A, Rachakonda PS, et al. TERT promoter mutations in familial and sporadic melanoma. Science 2013;339(6122):959-61.
4. Robles-Espinoza CD, Harland M, Ramsay AJ, et al. POT1 loss-of-function variants predispose to familial melanoma. Nat Genet 2014; 10.1038/ng.2947.
5. Shi J, Yang XR, Ballew B, et al. Rare missense variants in POT1 predispose to familial cutaneous malignant melanoma. Nat Genet 2014; 10.1038/ng.2941.
6. de Lange T. Shelterin: the protein complex that shapes and safeguards human telomeres. Genes Dev 2005;19(18):2100-10.
7. Ramsay AJ, Quesada V, Foronda M, et al. POT1 mutations cause telomere dysfunction in chronic lymphocytic leukemia. Nat Genet 2013;45(5):526-30.
8. Baxter AJ, Hughes MC, Kvaskoff M, et al. The Queensland Study of Melanoma: environmental and genetic associations (Q-MEGA); study design, baseline characteristics, and repeatability of phenotype and sun exposure measures. Twin Res Hum Genet 2008;11(2):183-96.
9. Aitken J, Welch J, Duffy D, et al. CDKN2A variants in a population-based sample of Queensland families with melanoma. J Natl Cancer Inst 1999;91(5):446-52.
10. Kasparian NA, Meiser B, Butow PN, et al. Anticipated uptake of genetic testing for familial melanoma in an Australian sample: An exploratory study. Psychooncology 2007;16(1):69-78.
11. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25(14):1754-60.
12. Picard Sequence Analysis Suite. [http://picard.sourceforge.net].
13. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20(9):1297-303.
14. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25(16):2078-9.
15. Aitken JF, Green AC, MacLennan R, et al. The Queensland Familial Melanoma Project: study design and characteristics of participants. Melanoma Res 1996;6(2):155-65.

16. McGregor B, Pfitzner J, Zhu G, et al. Genetic and environmental contributions to size, color, shape, and other characteristics of melanocytic naevi in a sample of adolescent twins. Genet Epidemiol 1999;16(1):40-53.
17. Lohmueller KE, Sparso T, Li Q, et al. Whole-exome sequencing of 2,000 Danish individuals and the role of rare coding variants in type 2 diabetes. Am J Hum Genet 2013;93(6):1072-86.
18. Neuhausen SL, Mazoyer S, Friedman L, et al. Haplotype and phenotype analysis of six recurrent BRCA1 mutations in 61 families: results of an international study. Am J Hum Genet 1996;58(2):271-80.
19. Kong A, Frigge ML, Masson G, et al. Rate of de novo mutations and the importance of father's age to disease risk. Nature 2012;488(7412):471-5.
20. Xin H, Liu D, Wan M, et al. TPP1 is a homologue of ciliate TEBP-beta and interacts with POT1 to recruit telomerase. Nature 2007;445(7127):559-62.
21. Loayza D, De Lange T. POT1 as a terminal transducer of TRF1 telomere length control. Nature 2003;423(6943):1013-8.
22. Ye JZ, Hockemeyer D, Krutchinsky AN, et al. POT1-interacting protein PIP1: a telomere length regulator that recruits POT1 to the TIN2/TRF1 complex. Genes Dev 2004;18(14):1649-54.
23. Ye JZ, de Lange T. TIN2 is a tankyrase 1 PARP modulator in the TRF1 telomere length control complex. Nat Genet 2004;36(6):618-23.
24. Liu D, Safari A, O'Connor MS, et al. PTOP interacts with POT1 and regulates its localization to telomeres. Nat Cell Biol 2004;6(7):673-80.
25. Zhong FL, Batista LF, Freund A, et al. TPP1 OB-fold domain controls telomere maintenance by recruiting telomerase to chromosome ends. Cell 2012;150(3):481-94.
26. Wang F, Podell ER, Zaug AJ, et al. The POT1-TPP1 telomere complex is a telomerase processivity factor. Nature 2007;445(7127):506-10.
27. Sfeir A, Kabir S, van Overbeek M, et al. Loss of Rap1 induces telomere recombination in the absence of NHEJ or a DNA damage signal. Science 2010;327(5973):1657-61.

Table 1: Germline mutations in the *ACD* and *TERF2IP* genes identified in melanoma families by exome sequencing

Family ID	Chr	Start§	Gene	cDNA change	mRNA refSeq	Protein change	cases in family	Genotyped cases who are carriers^	VAF ESP6500~	Polyphen2 prediction†	SIFT prediction†	GERP++
AUS1	16	67692665	*ACD*	c.958C>T	NM_001082486	p.Q320X	5	4/4	0	(truncating)	(truncating)	2.28
AUS2	16	67693137	*ACD*	c.746A>G	NM_001082486	p.N249S	5	7/7	3/8597	benign	tolerated	-9.69
DK1	16	67693137	*ACD*	c.746A>G	NM_001082486	p.N249S	4	3/4	3/8597	benign	tolerated	-9.69
AUS3	16	67692920	*ACD*	c.814G>A	NM_001082486	p.V272M	3	2/2	0	probably damaging	damaging	3.44
AUS4	16	67693513	*ACD*	c.598G>A	NM_001082486	p.A200T	6	4/6	0	probably damaging	damaging	3.4
AUS5	16	67692659	*ACD*	c.964A>T	NM_001082486	p.I322F	5	3/4	0	possibly damaging	damaging	0.87
UK1	16	75690399	*TERF2IP*	c.1090C>T	NM_018975	p.R364X	4	3/3	0	(truncating)	(truncating)	4.88
AUS6	16	75682352	*TERF2IP*	c.572A>G	NM_018975	p.Q191R	2	2/2	0	benign	tolerated	4.62
SE1	16	75681808	*TERF2IP*	c.28G>C	NM_018975	p.D10H	5	1/3	0	probably damaging	damaging	5.75
AUS7	16	75681795	*TERF2IP*	c.15G>A	NM_018975	p.M5I	1	1/1	0	benign	damaging	2.55

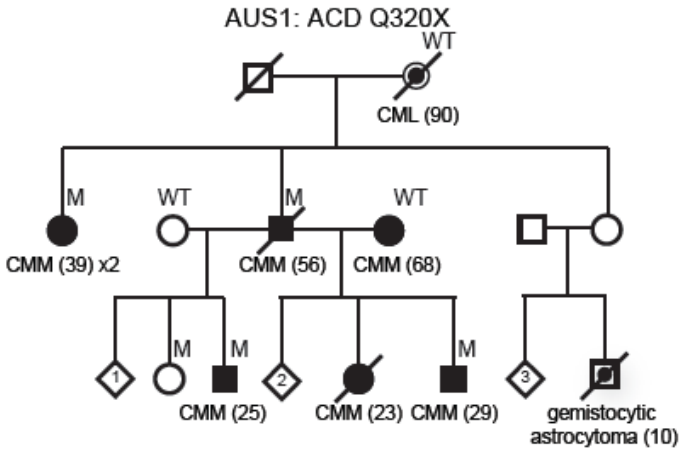
§ genome build GRCh37, hg19

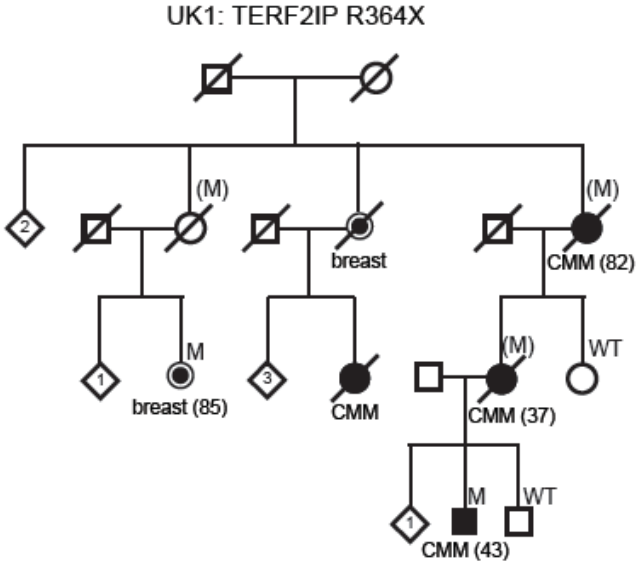
^ includes obligate carriers

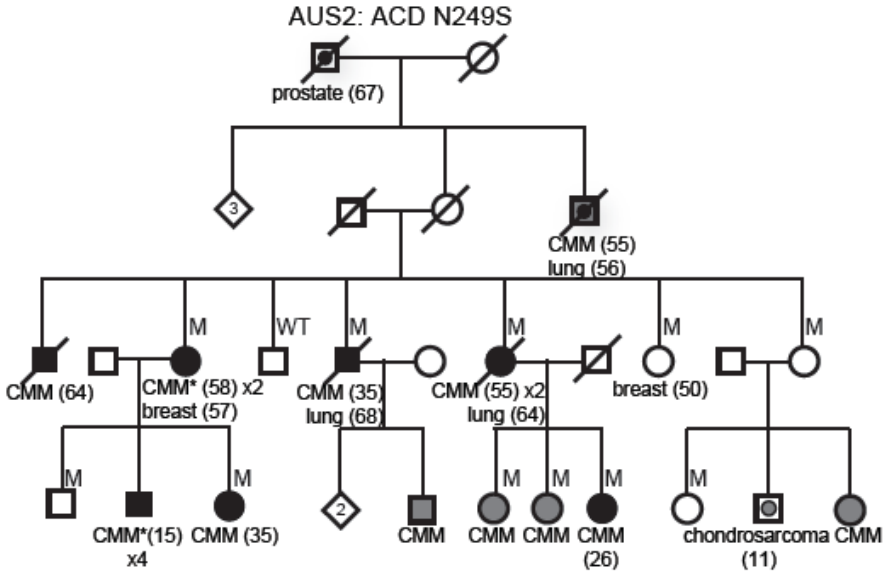
~European American population

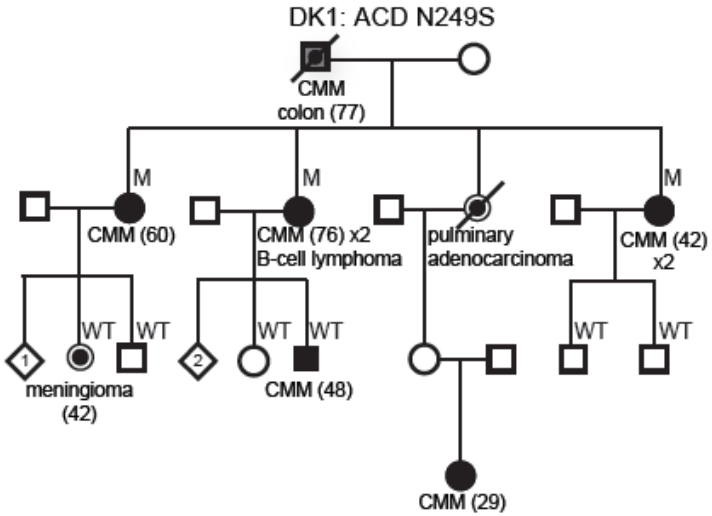
†effect of nonsense mutations not predicted by either Polyphen2 or SIFT are annotated as truncating

GERP++ is an estimate of the constrained elements in the human genome


AUS1: ACD Q320X
WT
CML (90)
M
WT
M
WT
CMM (39) x2
CMM (56)
CMM (68)
M
M
M
CMM (25)
CMM (23)
CMM (29)
gemistocytic
astrocytoma (10)


UK1: TERF2IP R364X
(M)
(M)
breast
CMM (82)
M
breast (85)
CMM
(M)
WT
CMM (37)
M
WT
CMM (43)


AUS2: ACD N249S
prostate (67)
CMM (55)
lung (56)
M
WT
M
M
M
M
CMM (64)
CMM* (58) x2
breast (57)
CMM (35)
lung (68)
CMM (55) x2
lung (64)
breast (50)
M
M
M
M
M
M
M
M
CMM*(15)
x4
CMM (35)
CMM
CMM
CMM
CMM
(26)
chondrosarcoma
(11)
CMM


DK1: ACD N249S
CMM
colon (77)
M
M
M
M
CMM (60)
CMM (76) x2
B-cell lymphoma
pulminary
adenocarcinoma
CMM (42)
x2
WT
WT
WT
WT
WT
WT
meningioma
(42)
CMM (48)
CMM (29)

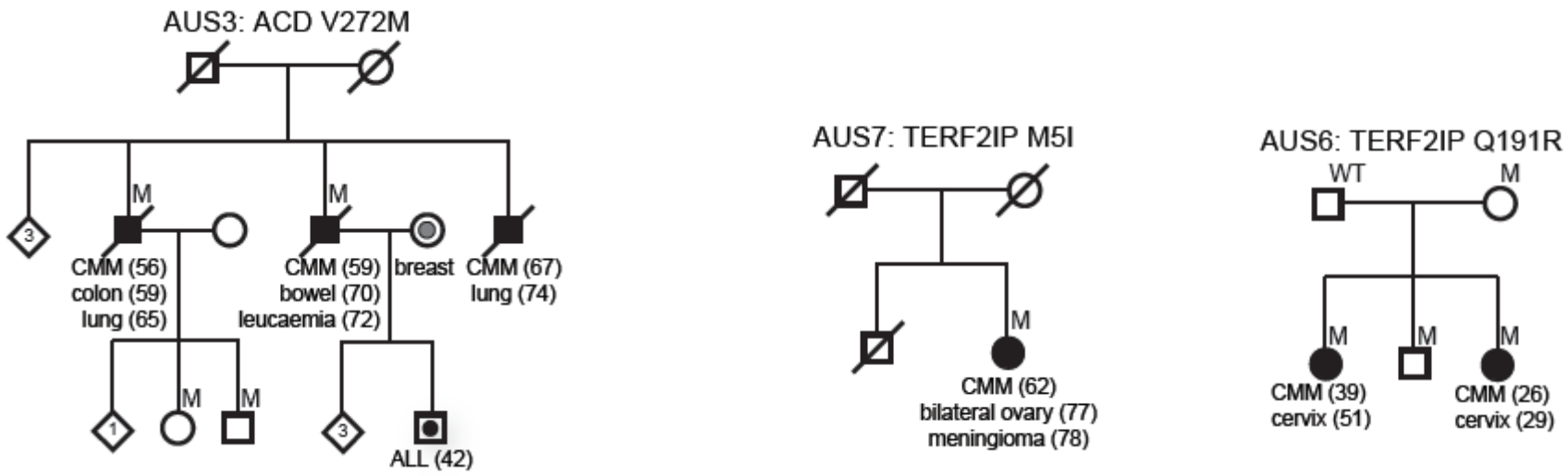


Figure 1: Mutations in ACD and TERF2IP that segregate with melanoma. The age at first diagnosis of cutaneous malignant melanoma (CMM) is indicated in brackets. If the individual has had more than one primary melanoma the first age at onset is annotated and the total number of CMM is given. A line through a symbol indicates the individual is deceased. Individuals carrying a mutation are indicated by 'M', while family members that are wild-type for the indicated variant are annotated 'WT'. '(M)' indicates as individual is an obligate carrier. Square symbols indicate males and circles females. Black symbols represent confirmed CMM cases and grey symbols unconfirmed cases. Symbols with a central black circle represent individuals with a confirmed cancer other than CMM. Those family members with other unconfirmed cancers are indicated by symbols containing a central grey circle. Unaffected siblings are indicated by a diamond with the number of siblings shown in the centre of the symbol.

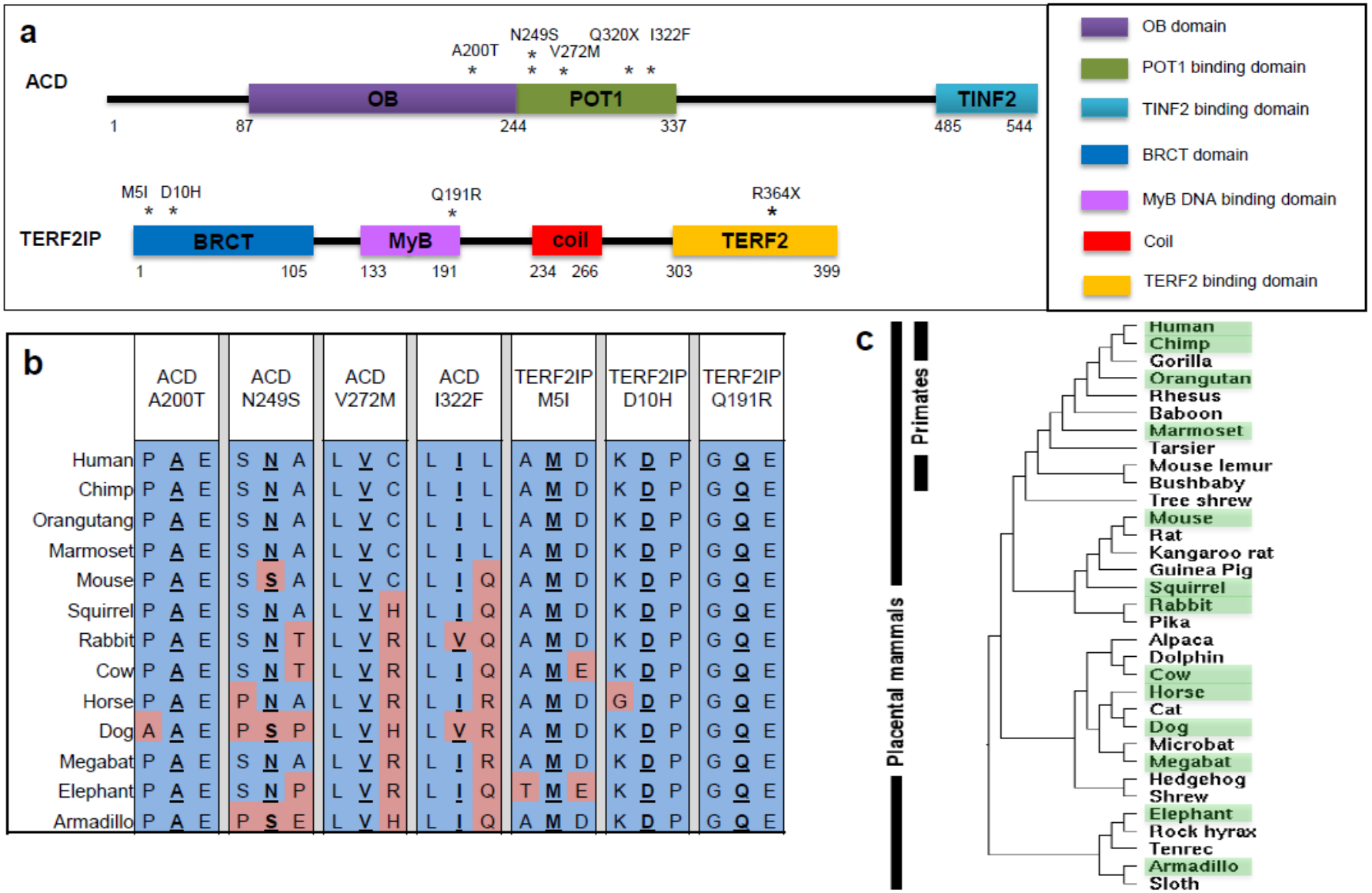


Figure 2: Relative location of germline variants in melanoma cases with respect to ACD and TERF2IP protein domains (a) and conservation of ACD and TERF2IP variants in melanoma families across placental mammals (b) and spanning species clades (c).

Supplementary results

Novel mutations were also observed in *TERF1*, *TERF2* and *TINF2*, encoding the three other members of the shelterin complex studied here (Supplementary Table 8). The main role of TERF1, telomeric repeat-binding factor 1 (also known as TRF1), and TERF2, telomeric repeat-binding factor 2 (also known as TRF2), is to bind telomeric double-stranded DNA (dsDNA) and thus mediate recruitment of different factors to telomeres[1,2]. These subunits directly recognise the telomeric repeat motif and bind using a DNA-binding Myb domain[3-5]. Each of these proteins also contains a TRFH domain (Supplementary Fig. 4) that allows homodimerization[6]. A TERF1 variant (p.A105P) was found in the only individual available for testing from a three-case family; this individual presented with MPM and thyroid cancer (Supplementary Fig. 2b). Two mutations were discovered in TERF2. A p.S119C substitution occurred in the TRFH domain in both affected members of a two-case family (Supplementary Fig. 2b). A p.S379C variant was observed in three of five confirmed cases in an Australian family. One individual had MPM and a second CMM case had lung and unconfirmed cervical cancer (Supplementary Fig. 2b). This mutation in TERF2 occurs at a highly conserved amino acid residue across species (Supplementary Fig. 5a,b) and is predicted to be damaging/probably damaging by Polyphen2 and SIFT analysis (Supplementary Table 8). Two variants could be mapped on a known protein structure: TERF1 p.A105P and TERF2 p.S119C (Supplementary Fig. 6a). Both mutations are predicted to destabilize the protein structure with putative disruption of binding sites. Specifically, p.A105P occurs in an alpha helix, the variant residue being unfavoured for this structural motif[7]. Importantly, alpha helix 2 (p.A105P) is one of the three critical alpha helices that constitute the binding interface for homodimerisation[6]. Therefore, this model suggests that the variant impairs the homodimerization of the protein. Furthermore the variant p.S119C in TERF2 affects the putative binding interface of this molecule with TINF2[8]. As shown in Supplementary Fig. 6b, this residue interacts with two tightly packed water molecules which are important in the structure of the TINF2-binding loop. A mutation of this residue to cysteine, while being conservative, is expected to disrupt this arrangement for sterical reasons. This geometrical constrainment suggests that the binding interface would be disrupted.

Finally, a single novel variant was found in TINF2 (p.V22L) in one member of a two-case family (Supplementary Fig. 2b). TINF2, TERF1-interacting nuclear factor 2 (also known as TIN2), interacts with the TERF1-TERF2 dsDNA binding complex. TINF2 is at the centre of the complex, tethering the ACD-POT1 partnership to TERF1 and TERF2. It is crucial for the assembly of the shelterin complex as it is the only mechanism available for TERF1 to interact with TERF2[9-13].

To further assess the association between *TERF1*, *TERF2* and *TINF2* gene variants and familial melanoma susceptibility we searched for variants in these genes in exome data from a similar population of control individuals in the UK10K sequencing project (http://www.uk10k.org). After filtering for such variants in data from 520 exomes in a similar manner to that used for the CMM families (see Online Methods) we identified a total of six non-synonymous variants in these three genes (Supplementary Table 6). Two novel variants occurred each in *TERF1* and *TERF2*, one of which (TERF2 p.S119C) was identical to a variant found in a CMM family (Supplementary Table 8). Additionally, two rare variants were found in TINF2: p.G25A and p.S208P, which have been seen in 1/4133 and 15/4144 individuals of European descent in the ESP database, respectively.

Similar analysis of these three genes in sequence data from 2000 Danish control exomes revealed a total of 17 rare (N=7) or novel (N=10) variants (summarized in Supplementary Table 7).

References

1. Broccoli, D., Smogorzewska, A., Chong, L. & de Lange, T. Human telomeres contain two distinctMyb-related proteins, TRF1 and TRF2. *Nat Genet* 17, 231-5 (1997).

2. Smogorzewska, A. *et al.* Control of human telomere length by TRF1 and TRF2. *Mol Cell Biol* 20, 1659-68 (2000).

3. Nishikawa, T. *et al.* Solution structure of a telomeric DNA complex of human TRF1. *Structure* 9, 1237-51 (2001).

4. Court, R., Chapman, L., Fairall, L. & Rhodes, D. How the human telomeric proteins TRF1 and TRF2 recognize telomeric DNA: a view from high-resolution crystal structures. *EMBO Rep* 6, 39-45 (2005).

5. Hanaoka, S., Nagadoi, A. & Nishimura, Y. Comparison between TRF2 and TRF1 of their telomeric DNA-bound structures and DNA-binding activities. *Protein Sci* 14, 119-30 (2005).

6. Fairall, L., Chapman, L., Moss, H., de Lange, T. & Rhodes, D. Structure of the TRFH dimerization domain of the human telomeric proteins TRF1 and TRF2. *Mol Cell* 8, 351-61 (2001).

7. Pace, C.N. & Scholtz, J.M. A helix propensity scale based on experimental studies of peptides and proteins. *Biophys J* 75, 422-7 (1998).

8. Chen, Y. *et al.* A shared docking motif in TRF1 and TRF2 used for differential recruitment of telomeric proteins. *Science* 319, 1092-6 (2008).

9. Ye, J.Z. *et al.* TIN2 binds TRF1 and TRF2 simultaneously and stabilizes the TRF2 complex on telomeres. *J Biol Chem* 279, 47264-71 (2004).

10. Kim, S.H., Kaminker, P. & Campisi, J. TIN2, a new regulator of telomere length in human cells. *Nat Genet* 23, 405-12 (1999).

11. Houghtaling, B.R., Cuttonaro, L., Chang, W. & Smith, S. A dynamic molecular link between the telomere length regulator TRF1 and the chromosome end protector TRF2. *Curr Biol* 14, 1621-31 (2004).

12. Kim, S.H. *et al.* TIN2 mediates functions of TRF2 at human telomeres. *J Biol Chem* 279, 43799-804 (2004).

13. Ye, J.Z. *et al.* POT1-interacting protein PIP1: a telomere length regulator that recruits POT1 to the TIN2/TRF1 complex. *Genes Dev* 18, 1649-54 (2004).

Supplementary Table 1: Total number of families exome or genome sequenced, broken down by research centre and the number of cases sequenced in each centre.

Origin of family	Number of families sequenced	Number of melanoma cases sequenced
Brisbane	33	96
Copenhagen	7	13
Leeds	62	70
Leiden	3	9
Lund	4	9
Sydney	4	7
Total:	**113**	**204**

Supplementary Table 2: Summary information for individual's exome or genome sequenced.

Family ID	CMM cases in family #	N cases sequenced in family	Sample ID	Exome or genome	Ages of onset of CMM (years)	Other cancers (age in years)
AUS7	1	1	001	genome	62	bilateral ovary (77), meningioma (78)
AUS12	6	3	007	exome	29, 29	
AUS12	6	3	009	genome	53, 63, 64, 64, 66, 66, 68	
AUS12	6	3	010	exome	30, 32, 34, 38, 38	
AUS13	9	3	002	exome	54, 54, 60	
AUS13	9	3	007	exome	55, 65, 67, 73, 73, 73	prostate (79), bowel (83)
AUS13	9	3	058	exome	68, 77, 83, 83	
AUS14	5	3	006	exome	30	
AUS14	5	3	008	exome	22	
AUS14	5	3	010	exome	47	
AUS15	8	3	001	exome	55, 56, 62	
AUS15	8	3	003	genome*	61, 64, 69, 76, 76, 76, 76	
AUS15	8	3	004	exome	22	
AUS16	8	3	006	genome*	23, 45	
AUS16	8	3	011	exome	30, 33	
AUS16	8	3	015	exome	44	
AUS1	5	3	002	exome	39, 57	
AUS1	5	3	005	exome	29	
AUS1	5	3	006	exome	25	
AUS10	5	3	002	exome	78, 79	
AUS10	5	3	007	exome	65	
AUS10	5	3	010	exome	53	lung (60)
AUS17	8	4	002	exome	57, 75	
AUS17	8	4	007	exome	59, 60	
AUS17	8	4	009	exome	34	
AUS18	6	3	001	exome	47, 59, 61, 64, 74, 77	
AUS18	6	3	003	exome	34	
AUS18	6	3	004	exome	20	
AUS4	6	4	001	exome	34	
AUS4	6	4	004	exome	14	
AUS4	6	4	005	exome	74, 74	
AUS4	6	4	016	exome	59	prostate (69)
AUS19	4	1	007	exome	27, 27, 27, 27, 28, 28, 30	
AUS20	5	3	001	exome	50, 52, 58, 66, 68, 74, 76, 77, 77	prostate (71)
AUS20	5	3	003	genome*	55, 55	bladder (58), leukaemia (76), lung (77)
AUS20	5	3	004	exome	45, 45, 60, 71	prostate (69)
AUS21	8	3	001	exome	47	

AUS21	8	3	009	exome	41, 42, 44, 47, 50, 50, 53, 54, 58, 61, 63, 65, 66, 66, 66, 66, 67, 67, 68, 68, 68	
AUS21	8	3	015	exome	42, 42, 57, 62, 62, 62, 64, 64, 65, 65, 65	
AUS22	3	3	002	exome	59	
AUS22	3	3	004	exome	51, 51, 55, 60, 67, 69, 71, 72	
AUS22	3	3	005	exome	18, 47, 71, 71, 71, 71, 71, 71, 76, 76, 76	
AUS23	3	3	002	exome	53	
AUS23	3	3	003	genome	61	lymphoma (70), prostate (71), bowel (81)
AUS23	3	3	012	exome	37	
AUS24	9	4	001	exome	41, 69	
AUS24	9	4	003	exome	29	
AUS24	9	4	007	exome	14	
AUS24	9	4	016	exome	58	
AUS25	5	2	002	exome	43	
AUS25	5	2	007	exome	46, 69	
AUS26	5	3	001	exome	27, 31	
AUS26	5	3	003	exome	62, 74	
AUS26	5	3	005	exome	14, 14	
AUS27	6	2	002	exome	16	
AUS27	6	2	004	exome	60	
AUS28	8	3	002	exome	37, 55	
AUS28	8	3	003	genome	46, 36	
AUS28	8	3	020	exome	38, 38	
AUS29	4	2	005	exome	73	kidney, bladder
AUS29	4	2	006	exome	76	colon (71)
AUS30	6	4	004	exome	32	
AUS30	6	4	007	exome	55	
AUS30	6	4	008	exome	48	cervix (46)
AUS30	6	4	016	exome	17	
AUS31	6	4	001	exome	65	
AUS31	6	4	004	exome	79, 79	
AUS31	6	4	008	exome	49	
AUS31	6	4	009	exome	41	
AUS32	4	2	001	genome	54, 60, 71, 73, 73, 73, 74 (+22 others before compulsory records)	prostate (74)
AUS32	4	2	004	exome	84	
AUS2	5	3	001	genome	58, 59	
AUS2	5	3	005	exome	35	lung (68)
AUS2	5	3	009	exome	26	
AUS33	6	4	001	exome	47	breast (46)

AUS33	6	4	004	exome	57	
AUS33	6	4	006	exome	54	
AUS33	6	4	013	exome	43	
AUS34	6	4	001	genome	31, 31	breast (58)
AUS34	6	4	002	exome	19, 33, 37, 47, 48, 53	bilateral breast (51)
AUS34	6	4	003	exome	36	
AUS34	6	4	004	exome	19, 26, 26, 44, 44	
AUS35	4	2	001	genome	61	breast (79), lung (79)
AUS35	4	2	002	exome	58	
AUS36	11	2	003	exome	26, 27	
AUS36	11	2	014	genome	79, 79	kidney (70), bladder (80)
AUS37	3	2	002	exome	70	lung (75)
AUS37	3	2	003	genome	69	prostate (79), oesophagus (89)
AUS5	5	3	001	genome	57, 59, 61, 77	
AUS5	5	3	002	exome	30	
AUS5	5	3	013	exome	46	
AUS38	17	5	001	exome	23, 33, 48, 52, 52	
AUS38	17	5	007	exome	24	
AUS38	17	5	008	exome	34	
AUS38	17	5	011	exome	23	
AUS38	17	5	022	exome	38	
SE1	5	3	564	exome	77	
SE1	5	3	615	exome	52	
SE1	5	3	648	exome	31	
SE2	3	2	396	exome	47	
SE2	3	2	641	exome	29	
SE3	3	2	638	exome	78	
SE3	3	2	700	exome	72 (multiple)	
SE4	3	2	56	exome	<35	
SE4	3	2	64	exome	<30	
DK2	3	2	110736	exome	53	cervical CIS (38), uterus (75)
DK2	3	2	210949	exome	40	
DK3	3	1	200564	exome	28	
DK4	3	2	070650	exome	56	
DK4	3	2	090146	exome	63	
DK5	3	2	020946	exome	60	
DK5	3	2	160655	exome	53	
DK6	4	2	270550	exome	53, 55	
DK6	4	2	280944	exome	55	
DK1	3	2	020830	exome	76,80	large diffuse B cell lymphoma (82)
DK1	3	2	221143	exome	42,52,59	
DK7	4	2	130948	exome	55	
DK7	4	2	200450	exome	54	
AUS39	8	1	1067	genome*	58, 79, 88, 89	Wilms' tumour (53)

AUS40	3	1	2835	exome	20	
AUS41	8	3	2669	genome*	57	
AUS41	8	3	2673	exome	25	
AUS41	8	3	2680	exome	41	
AUS42	5	3	1762	exome	24	
AUS42	5	3	1764	genome*	28	
NL1	4	3	3	exome	41	
NL1	4	3	10	exome	44	
NL1	4	3	8	exome	24	
NL2	4	2	1	exome	61	
NL2	4	2	8	exome	45	
NL3	5	4	1	exome	71	breast (66)
NL3	5	4	3	exome	56	brain (53)
NL3	5	4	5	exome	71	
NL3	5	4	12	exome	24	
UK2	3	3	2024F	exome	68	
UK2	3	3	4353F	exome	62, 62, 63	
UK2	3	3	4354F	exome	35	
UK3	4	1	2033F	exome	18	
UK4	4	1	2114F	exome	44	cervix (26)
UK5	4	1	2141F	exome	42	
UK6	4	2	2177F	exome	36	
UK6	4	2	2176F	exome	36	
UK7	8	2	18646F	exome	54	
UK7	8	2	2213F	exome	21	
UK8	4	2	2453F	exome	72	
UK8	4	2	2388F	exome	49	
UK9	6	1	973F	exome	47, 48	
UK10	5	1	324F	exome	21	
UK11	6	2	1778F	exome	31	bladder (46)
UK11	6	2	1771F	exome	27	
UK12	5	1	2506F	exome	57	
UK13	3	2	1795F	exome	80, 81	
UK13	3	2	1792F	exome	41	
UK14	2	1	1326F	exome**	28, 32	
UK15	4	1	1814F	exome**	39, 52	
UK16	3	1	1952F	exome**	30	
UK17	3	1	2140F	exome**	35, 44	
UK18	3	1	2172F	exome**	9, 41	
UK19	3	1	2191F	exome**	41	
UK20	3	1	2143F	exome**	45, 45	
UK1	4	1	65F	exome	43	
UK21	3	1	2314F	exome**	37, 38, 38	
UK22	2	1	2289F	exome**	24, 34, 34	

UK23	3	1	2304F	exome**	17	
UK24	2	1	2394F	exome**	59, 69, 72, 74	
UK25	3	1	2417F	exome**	32, 54, 54	
UK26	2	1	2402F	exome**	41, 42, 50, 51, 53	
UK27	3	1	2406F	exome**	50	
UK28	3	1	2507F	exome**	47	
UK29	2	1	2330F	exome**	45, 68, 68, 68	breast (71)
UK30	5	1	1485F	exome	34, 34, 34, 38	
UK31	3	1	99F	exome**	55, 56, 56, 70	
UK32	2	1	1290F	exome**	49, 58, 58, 58	
UK33	2	1	788F	exome**	45, 45, 45	
UK34	2	1	986F	exome**	23, 30, 44	
UK35	2	1	1476F	exome**	44, 50, 51, 64, 64, 65	
UK36	3	1	1561F	exome**	59, 66	prostate (67)
UK37	2	1	1687F	exome**	39, 46, 50	
UK38	4	1	1819F	exome	50, 54	breast (52)
UK39	4	1	1823F	exome	42	
UK40	4	2	1892F	exome	36	
UK40	4	2	1893F	exome	38	
UK41	4	1	1894F	exome	34	
UK42	5	1	2029F	exome	57	
UK43	1	1	CT0024	exome**	39	
UK44	1	1	CT0177	exome**	40	
UK45	1	1	CT1241	exome**	75, 80	
UK46	1	1	CT1406	exome**	31	
UK47	1	1	CT1746	exome**	38	
UK48	1	1	CT1912	exome**	40	
UK49	1	1	CT3775	exome**	65, 67	
UK50	1	1	CT3761	exome**	63, 74	
UK51	1	1	CT0155	exome**	38	
UK52	1	1	CT4128	exome**	58, 58	
UK53	1	1	CT0673	exome**	38	
UK54	1	1	CT0861	exome**	33	
UK55	1	1	CT1036	exome**	40	
UK56	2	1	CT1643	exome**	28	
UK57	2	1	CT1163	exome**	57	
UK58	2	1	CT1204	exome**	68	
UK59	2	1	CT1927	exome**	45	
UK60	2	1	CT2996	exome**	67	
UK61	2	1	CT0598	exome**	34	
UK62	1	1	CT1687	exome**	33	

#confirmed cases only; *done by Complete Genomics; **done by Beijing Genomics Institute

Supplementary Table 3: Summary information for targeted pull down samples.

Proband ID	CMM cases in family #	Ages of onset of CMM (years)
000045	2	39
000085	2	57
000121	2	36
000122	2	56
000147	2	36
000178	2	46, 62
000246	1	50
000336	2	38
000371	2	57
000461	2	65, 67, 67, 75
000578	2	29
000602	2	20
000786	2	53
000906	2	21
000970	2	59
000971	2	53, 70
001032	2	26, 28, 41
001077	2	62
001093	2	41
001142	2	56
001285	2	58
001375	2	68, 76
001399	2	36
001458	2	63
001599	2	48
001669	2	55
001746	2	52
001872	2	55
001891	2	44
001918	2	46
001968	3	36
001971	2	33
002012	2	48
002135	2	46
002145	2	32
002214	2	66, 71
002274	2	37
002295	3	41
002458	2	41
002527	2	68
002591	2	68, 73
002676	2	58
002703	2	66, 72

002771	2	35
002810	2	40
002832	3	63
002860	2	57, 59, 61, 77
002869	2	38, 53
002918	2	66
002948	2	65
002967	2	61, 70, 83, 84
003055	2	26
003057	2	64, 70, 79, 81
003091	2	55
003153	2	53, 74
003167	2	17
003173	2	61
003190	2	55
003254	2	58, 76, 82
003267	2	28
003402	2	63
003412	2	59
003451	2	26, 38
003611	2	44, 56
003655	2	51
003723	3	54
003741	2	40
003754	2	39
003771	2	31
004169	2	29, 32
004278	2	48
004279	2	47, 48, 57
004288	2	65
004307	3	52
004359	2	33
004497	2	45, 48
004604	2	64, 68
004874	2	54
004963	2	30
004966	2	19
004970	2	75, 85, 85, 85, 89
004974	2	54, 59, 63, 67, 69, 70
005036	2	56
005048	2	56, 57
005055	2	35
005116	2	25
005177	2	69, 87
005223	2	34
005247	2	50

005429	2	28
005510	2	71, 78, 87
005577	2	52
005638	2	69
005651	2	64
005711	2	24
005814	2	73, 73
005851	2	55
005899	2	35, 54
005910	2	33
005921	2	54
005929	2	62
005939	2	46
006096	2	61
006120	2	51, 55
006122	2	33, 37
006126	2	61, 65, 65
006128	2	71
006180	2	36
006190	2	64
006202	2	31
006270	2	35
006281	2	24
006349	1	27
006416	3	56
006487	3	62, 76
006648	2	66
006649	3	50, 63, 66
006686	2	69
006713	2	48
006865	2	50
006915	2	75
006945	2	57
007005	2	34
007099	2	56, 66
007137	2	51
007178	2	74
007303	2	40
007311	2	23
007319	2	40, 41
007321	1	63
007325	2	63
007346	2	56
007356	2	41
007373	2	36
007376	2	49

007415	2	52
007444	2	64
007579	2	59
008047	2	75
008136	2	50, 53
008191	2	38
008224	2	61
008245	2	42
008290	2	49
008654	2	48, 55, 64
008779	4	38
008856	2	19
008876	2	24
008891	2	64
008972	2	73
009005	2	53, 62, 63
009161	2	61
009171	3	14
009198	2	41
009237	2	65
009254	2	14
009341	2	37
009386	2	50
009494	2	50
009522	2	35
009537	2	18, 32
009547	2	48
009626	2	57
009644	2	67
009650	2	29
009760	2	28
009782	2	52
009787	2	63
009811	3	57
009815	2	74, 75, 87, 87
009818	2	27
009828	2	62
009864	2	24
009909	2	64
009912	3	38
009965	1	43
010069	2	62, 78
010129	1	18
010133	4	12
010156	2	45, 45, 59
010187	2	38

010213	2	58
010274	2	66
010284	2	12
010335	3	48, 58
010357	3	65, 65, 83
010373	2	55, 72
010408	2	60
010441	2	27
010448	2	54
010471	3	29, 42
010482	3	54
010587	2	51, 58
010709	2	53
010867	2	82, 82, 87
011093	2	67
011115	2	46, 47
011132	2	35
011321	2	25
011351	2	50
011428	2	76
011437	1	44
011453	2	36
011488	2	32
011571	2	46
011630	2	76
011661	2	19, 30
011692	2	49, 54, 55, 56, 65
011734	2	41, 48
011746	2	34, 46
011751	2	63
011756	2	69, 78, 80, 83
011802	2	59
011809	2	34
011848	2	46
011880	2	67
011893	2	66
011905	2	27
011954	2	36
011971	2	61
012081	2	34
012086	2	32
012150	2	62
012152	2	32
012182	2	50
012283	2	58
012319	2	56

012324	2	40
012592	2	68
012617	2	53
012681	2	46
012698	2	60
012811	2	42
012812	2	45, 49
012828	2	50
012878	2	43
012883	2	44
012982	2	60
013014	2	51
013023	2	76
013049	2	44
013148	2	43, 44
013304	2	61
013341	2	70, 70
013362	2	39
013398	2	42
013434	2	39
013438	3	37
013447	2	67, 78
013477	2	31
013538	2	21
013546	2	44
013650	2	36, 40
013674	2	26
013685	2	40
013698	2	52
013809	2	28
013811	2	53, 57, 61, 61, 62
013829	2	53, 55
013868	2	46, 46
013900	2	48
013910	2	33
013917	2	63
013921	2	55, 65
013999	2	58
014008	2	45
014254	2	39
014296	2	40
014354	2	53, 53, 60, 60
014359	2	65, 66
014392	2	46
014401	2	43
014438	2	29

014465	2	34
027982	3	53
030713	3	66
032327	3	72
033751	2	19
050193	2	41
050251	2	56
050261	2	52
050280	2	58
050314	2	57
050339	2	24
050383	2	39
050400	2	61, 63
050413	2	51
050427	2	46
050443	2	46
050492	2	64
050497	2	41
050519	2	43
050520	2	65
050611	2	64
050657	2	58
050695	2	71, 75, 79, 83
050731	2	45, 46
050744	2	66
050764	2	57
050795	3	51
050826	2	43
050837	2	68
050872	2	71
050885	2	34
050904	2	64
050971	2	63, 64
051009	2	39
051016	3	50
051047	2	65
051061	2	35
051101	2	31, 36
051121	2	65
051146	2	66
051177	2	49
051226	2	33
051269	2	20
051362	2	26, 26
051412	2	40
051457	2	53, 58, 63, 67

051484	2	62
051502	2	71
051506	2	49
051522	2	54
051538	2	66
051550	2	34
051594	2	60
051600	2	66
051656	2	64
051662	2	53
051680	2	62
051802	2	55
051820	2	45
051824	2	40
051836	2	68, 77, 80, 81, 82
051839	2	34
052037	2	68, 70, 79, 81, 82, 83
052094	2	44
052101	2	46, 46
052104	2	45
052116	2	50
052193	2	30
052242	2	26
052260	2	38
052280	2	47
052313	2	69
052364	2	41
052447	2	70
052448	2	46
052465	2	44
052499	2	67
052507	2	62
052708	3	16
052730	2	59
052749	2	52, 63
052793	2	46
052809	2	41
052810	2	42
052833	2	72
AUS43-006	2	51
AUS44-001	2	60
AUS45-001	2	52
AUS46-002	1	57
AUS47-002	4	48
AUS48-001	1	20
AUS49-001	2	71

AUS50-001	2	76
AUS51-001	2	62
AUS52-001	1	45
AUS53-001	1	61
AUS54-008	3	29
AUS55-001	2	43
AUS56-001	2	67
AUS57-001	1	47
AUS58-001	2	57
AUS59-002	2	68
AUS60-001	2	43
AUS70-001	2	54
AUS71-001	1	44, 48, 48, 49, 49, 50, 50, 51, 52, 54, 54, 56
AUS72-001	1	23, 25, 25, 25
AUS73-001	2	61
AUS74-001	2	53, 53, 60, 60, 60, 60
AUS75-001	5	66, 69, 70, 70, 81
AUS76-005	4	18, 47
AUS77-001	6	51, 51
AUS78-002	2	49
AUS79-702	3	59, 59, 59
AUS80-002	3	43, 45, 59
AUS81-001	4	25
AUS82-001	1	55
AUS83-005	1	53
AUS84-001	2	44
AUS85-003	5	40
AUS86-001	3	43
AUS87-001	5	46, 47, 50, 52, 55, 55, 57, 58, 58, 64
AUS88-001	4	44, 52, 67, 67
AUS89-001	4	36
AUS90-003	4	73, 74, 74

#confirmed cases only

Supplementary Table 4: Primers used in the Sanger sequencing and Sequenom MassArray genotyping.

	Variant	Forward primer	Reverse primer	Extension primer
Sanger sequencing	ACD A200T	gtaaaacgacggccagtcttgttgcagggaggagaag	attggacgaggtggactctg	
	ACD N249S	gtaaaacgacggccagtgcccccattaactaccctta	ccaactcctcaccctgacat	
	ACD I322F	gtaaaacgacggccagtatgtcagggtgaggagttgg	tgaggaactgggtgaggaag	
	ACD V272M	gtaaaacgacggccagtgcccccattaactaccctta	ccaactcctcaccctgacat	
	ACD Q320X	gtaaaacgacggccagtatgtcagggtgaggagttgg	tgaggaactgggtgaggaag	
	TERF1 S119C	gtaaaacgacggccagttgtcagtttcttgccactcg	caatctctccccattgctgt	
	TERF2 S379C	gtaaaacgacggccagtccagccctcaaaaacaagag	ctttccattcccactcatgc	
	TERF2IP M5I	gtaaaacgacggccagtcgacccaaaagtaaggagga	gggcattttccttcacgtag	
	TERF2IP D10H	gtaaaacgacggccagtcgacccaaaagtaaggagga	gggcattttccttcacgtag	
	TERF2IP Q191R	gtaaaacgacggccagtggacgacgtagccatcctta	cagctacttgggaggctgaa	
	TERF2IP R364X	acttggatctatcaacagttacaca	ttgaccaatgcctctctggt	
	TINF2 V22L	gtaaaacgacggccagtttaaagctgagcgacccagt	tgattcagggctttcaggac	
MassArray	ACD N249S	acgttggatgtctcttcctgtagggagcac	acgttggatggatcactggtcaagctctac	acctttcagagtccacctcgtcca
LOH	ACD A200T	acgttggatgctcttcgtgacgattctgag	acgttggatgaagcggtccacctggagata	ccgaactcctccagccc
	ACD N249S	acgttggatggttaggatcactggtcaagc	acgttggatgtctcttcctgtagggagcac	agagtccacctcgtcca
	ACD I322F	acgttggatgctcaatgctgtgcatctctg	acgttggatgaacctctccagtcccccctta	gcctagagagctcagaa
	ACD Q320X	acgttggatgctcaatgctgtgcatctctg	acgttggatgaacctctccagtcccccctta	tgcatctctgagaatgac
	TERF2 S379C	acgttggatgaagcagattggtcttggagg	acgttggatgatggtggcgctgaagcggc	cagcgcaggcctcaact
	TERF2IP M5I	acgttggatgttctagtagtgctcggcgtc	acgttggatgacgaacagagtcgaggaatg	gacatggcggaggcgat

Supplementary Table 5: *ACD* haplotype analysis. The table shows 112 markers detected in the region of *ACD*. The SNPs listed are consistent with a shared haplotype.

rsID	Position	41202-001	41202-007	41202-009	41202-021	221143	20830	100132	Shared haplotype	Reference allele freq.
rs12918121	67187795	GG	GG	GG	GG	GG	GG	GG	G	0.980
rs13312727	67188443	AA	AA	AA	AA	AA	AA	AC	A	0.970
rs13312723	67191696	GG	GG	GG	GG	GG	GG	GG	G	0.970
rs11642409	67203777	CC	CC	CC	CC	CC	CC	CC	C	0.990
rs2233456	67207962	GG	GG	GG	GG	GG	GG	--	G	0.990
rs8057598	67209255	AA	AA	AA	AA	AA	AA	AG	A	0.940
rs13339140	67213923	GG	GG	GG	GG	GG	GG	GG	G	0.940
rs868213	67220457	AA	AA	AA	AA	AA	AA	AG	A	0.940
rs2346676	67223803	AA	AA	AA	AA	AA	AA	AG	A	0.940
rs3729639	67225501	GG	GG	GG	GG	GG	GG	AG	G	0.940
rs3730393	67226405	GG	GG	GG	GG	GG	GG	GG	G	0.970
rs3730397	67228128	GG	GG	GG	GG	GG	GG	GG	G	0.980
rs3730403	67229486	AA	AA	AA	AA	AA	AA	AA	A	0.980
rs3730406	67230422	AG	GG	GG	GG	GG	GG	GG	G	0.930
rs11700	67232684	AA	AA	AA	AA	AA	AA	AG	A	0.940
rs12923138	67233266	AC	AA	AC	AA	AA	AA	AC	A	0.550
rs8058861	67234134	AA	AA	AA	AA	AA	AA	AG	A	0.940
rs12051247	67241282	GG	GG	GG	GG	GG	GG	AG	G	0.940
rs13338846	67247444	AA	AA	AA	AA	AA	AA	AA	A	0.980
rs13338688	67248831	GG	GG	GG	GG	GG	GG	AG	G	0.940
rs8049470	67250226	GG	GG	GG	GG	GG	GG	GG	G	0.990
rs7193713	67254419	AA	AA	AA	AA	AA	AA	AG	A	0.940
rs7196793	67259567	CC	CC	CC	CC	CC	CC	AC	C	0.960
rs6499118	67265360	CC	CC	CC	CC	CC	CC	AC	C	0.940
rs28730536	67266617	GG	GG	GG	GG	GG	GG	AG	G	0.960
rs13330273	67271106	AA	AA	AA	AA	AA	AA	AG	A	0.960

rs16957265	67272377	AA	AA	AA	AA	AA	AA	AA	A	0.970
rs12922483	67282234	AA	AA	AA	AA	AA	AA	AA	A	0.960
rs12447443	67295020	GG	GG	GG	GG	GG	GG	GG	G	0.990
rs1061356	67304915	AA	AA	AA	AA	AA	AA	AG	A	0.960
rs7186310	67305513	AA	AA	AA	AA	AA	AA	AC	A	0.890
rs8050745	67311767	GG	GG	GG	GG	GG	GG	AG	G	0.930
rs11860295	67316234	GG	GG	GG	GG	GG	GG	AG	G	0.930
rs7200919	67316600	AG	GG	AG	GG	GG	GG	AG	G	0.560
rs8044843	67318242	AA	AA	AA	AA	AA	AA	AG	A	0.900
rs3868142	67320223	GG	GG	GG	GG	GG	GG	AG	G	0.930
rs785029	67320920	GG	GG	GG	GG	GG	GG	GG	G	0.940
rs9922130	67323664	AA	AA	AA	AA	AA	AA	AG	A	0.960
rs16957304	67334969	AA	AA	AA	AA	AA	AA	AA	A	0.960
rs10852437	67338497	AA	AA	AA	AA	AA	AA	AG	A	0.960
rs4783754	67358816	GG	GG	GG	GG	GG	GG	GG	G	0.910
rs16942883	67387817	AA	AA	AA	AA	AA	AA	AA	A	0.960
rs11859352	67390448	AA	AA	AA	AA	AA	AA	AG	A	0.900
rs10500541	67392027	AA	AA	AA	AA	AA	AA	AG	A	0.960
rs16957358	67394541	AA	AA	AA	AA	AA	AA	AA	A	0.960
rs8052655	67409180	GG	GG	GG	GG	GG	GG	GG	G	0.960
rs6499127	67417003	AA	AA	AA	AA	AA	AA	AA	A	0.960
rs16957407	67418598	AA	AA	AA	AA	AA	AA	AA	A	0.960
rs8058996	67418860	GG	GG	GG	GG	GG	GG	GG	G	0.999
rs16957415	67418957	AA	AA	AA	AA	AA	AA	AA	A	0.960
rs8059226	67419025	CC	CC	CC	CC	CC	CC	CC	C	0.999
rs12920590	67420603	AG	AG	AG	AA	AA	AA	AA	A	0.610
rs2242140	67421798	AA	AA	AA	AA	AA	AA	AA	A	0.970
rs1053612	67423939	AA	AA	AA	AA	AA	AA	AG	A	0.920
rs11552319	67424122	GG	GG	GG	GG	GG	GG	GG	G	0.960
rs3785095	67426869	GG	GG	GG	GG	GG	GG	GG	G	0.997

rs5479	67469733	CC	CC	CC	CC	CC	CC	CC	C	0.960
rs2279023	67478251	GG	GG	GG	GG	GG	GG	GG	G	0.970
rs9922624	67478976	GG	GG	GG	GG	GG	GG	GG	G	0.930
rs3892816	67499714	AA	AA	AA	AA	AA	AA	AA	A	0.960
rs8044652	67506247	GG	GG	GG	GG	GG	GG	GG	G	0.950
rs34702968	67509400	GG	GG	GG	GG	GG	GG	GG	G	0.990
rs13338993	67515312	AA	AA	AA	AA	AA	AA	AA	A	0.930
rs35079605	67517538	GG	GG	GG	GG	GG	GG	GG	G	0.996
rs13334182	67524024	GG	GG	GG	GG	GG	GG	GG	G	0.950
rs7196853	67560613	AA	AA	AA	AA	AA	AA	AG	A	0.900
rs1210978	67578396	GG	GG	GG	GG	GG	GG	GG	G	0.960
rs8054369	67587281	GG	GG	GG	GG	GG	--	GG	G	0.999
rs9940960	67598274	AA	AA	AA	AA	AA	AA	AA	A	0.999
rs34523198	67607987	AA	AA	AA	AA	AA	AA	AA	A	0.960
rs28711261	67617186	AA	AG	AA	AA	AA	AA	AG	A	0.870
rs6499135	67625872	AC	AC	AC	AC	AA	--	AC	A	0.520
rs17686899	67627635	CC	AC	CC	CC	CC	CC	CC	C	0.950
rs16957489	67651879	GG	AG	GG	GG	GG	GG	AG	G	0.870
rs7191281	67655133	AA	AG	AA	AA	AA	AA	AG	A	0.870
rs6499137	67671804	AA	AC	AA	AA	AA	AA	AC	A	0.910
rs34012739	67675969	AA	AG	AA	AA	AA	AA	AG	A	0.840
rs9972635	67682580	AA	AA	AA	AA	AA	AA	AA	A	0.950
rs14920	67691515	CC	CC	CC	AC	CC	CC	CC	C	0.940
ACD p.N249S	**67693137**									
rs35356834	67696365	GG	GG	GG	GG	GG	GG	GG	G	0.960
rs9927609	67699257	AA	AA	AA	AA	AA	AA	AA	A	0.996
rs7185559	67699734	GG	GG	GG	GG	GG	GG	GG	G	0.996
rs7187476	67699948	AA	AG	AA	AA	AA	AA	AG	A	0.840
rs12449157	67708897	AA	AG	AA	AA	AA	AA	AA	A	0.840
rs34895002	67709262	AG	GG	GG	GG	GG	GG	GG	G	0.970

rs12927959	67727069	AG	AA	AG	AG	AA	AA	AA	A	0.660
rs4474673	67758778	GG	AG	GG	GG	GG	GG	GG	G	0.880
rs8051587	67796315	AA	AC	AA	AA	AA	AA	AA	A	0.840
rs6499141	67797134	AA	AA	AA	AA	AA	AA	AA	A	0.960
rs13335252	67808212	GG	GG	GG	GG	GG	GG	GG	G	0.996
rs8057184	67827247	GG	GG	GG	GG	AG	AG	GG	G	0.820
rs17619927	67834585	GG	GG	GG	GG	GG	GG	GG	G	0.960
rs12325430	67853425	AG	AG	AG	AG	AA	AA	AA	A	0.510
rs34693072	67861501	AA	AA	AA	AA	AA	AA	AA	A	0.950
rs3743733	67863451	AG	AG	AG	AG	AA	AA	AA	A	0.510
rs11558533	67865937	AA	AA	AA	AA	AA	AA	AA	A	0.990
rs11558534	67867739	CC	CC	CC	CC	CC	CC	CC	C	0.960
rs6499143	67874264	AA	AG	AA	AA	AA	AA	AA	A	0.830
rs3809630	67879400	GG	AG	GG	GG	GG	GG	GG	G	0.870
rs1113232	67879827	CC	CC	CC	CC	CC	CC	CC	C	0.960
rs28679372	67882932	AA	AA	AA	AA	AA	AA	AA	A	0.960
rs12930280	67891074	AA	AA	AA	AA	AA	AA	AA	A	0.970
rs1124324	67897487	GG	AG	GG	GG	GG	GG	GG	G	0.870
rs2271293	67902070	GG	AG	GG	GG	GG	GG	GG	G	0.870
rs34132524	67909150	GG	GG	GG	GG	GG	GG	GG	G	0.920
rs8060686	67911517	AA	AG	AA	AA	AA	AA	AA	A	0.820
rs12447425	67919535	GG	GG	GG	GG	GG	GG	GG	G	0.990
rs10468274	67922342	GG	AG	GG	GG	GG	GG	GG	G	0.830
rs7196789	67927124	GG	AG	GG	GG	GG	GG	GG	G	0.830
rs34298659	67927697	GG	GG	GG	GG	GG	GG	GG	G	0.970
rs16942887	67928042	GG	AG	GG	GG	GG	GG	GG	G	0.870
rs16957597	67946356	CC	CC	CC	CC	CC	CC	CC	C	0.997

Supplementary Table 6: Rare (VAF < 0.01) germline variants in the shelterin complex identified in UK10K control exomes.

Sample	Chr	Position	Gene	cDNA change	mRNA refSeq	Protein change	VAF ESP6500~	dbSNP 138	Polyphen2	SIFT	GERP
MUIR 1523795	16	67694167	ACD	c.215G>T	NM_001082486	p.A72E	0	0	probably damaging	damaging	1.04
MUIR 1523826	16	69391449	TERF2	c.1316T>C	NM_005652*	p.D439G	0	0	benign	tolerated	5.29
COLLIER 5120586	16	69418607	TERF2	c.356G>C	NM_005652*	p.S119C	0	0	probably damaging	damaging	5.52
ABDN 5050486	8	73921375	TERF1	c.254C>G	NM_017489	p.S85C	0	0	probably damaging	tolerated	4.08
COLLIER 5120685	8	73951369	TERF1	c.1058A>T	NM_017489	p.E353V	0	0	benign	tolerated	-0.77
MUIR 1523793	14	24710064	TINF2	c.622A>G	NM_001099274.1	p.S208P	1/8267	rs377436580	probably damaging	damaging	4.02
MUIR 1523853	14	24711465	TINF2	c.74C>G	NM_001099274.1	p.G25A	15/8303	rs202093758	benign	tolerated	2.64

*refers to canonical protein isoform Q15554-1

~ European American population

Supplementary Table 7: Rare (VAF < 0.01) germline variants in the shelterin complex identified in 2000 Danish control exomes.

Chr	Position	Gene	cDNA change	mRNA refSeq	Protein change	dbsnp137	VAF ESP6500 ~	VAF 1000 Genomes	VAF Danish	PolyPhen2	Sift	GERP
16	67691750	ACD	c.1471C>A	NM_001082486	p.P491T	rs201441120	2/8598		0.001018	damaging	damaging	3.27
16	67691947	ACD	c.1406G>A	NM_001082486	p.R469Q	rs200293827	1/8598		0.000509	benign	tolerated	-7.84
16	67692017	ACD	c.1336G>A	NM_001082486	p.A446T				0.000509	probably damaging	tolerated	1.6
16	67692121	ACD	c.1232C>A	NM_001082486	p.T411N				0.001272	benign	tolerated	1.99
16	67692631	ACD	c.992G>A	NM_001082486	p.R331K				0.000255	benign	damaging	1.15
16	67692863	ACD	c.871A>G	NM_001082486	p.T291A	rs139438549	11/8589		0.000763	damaging	damaging	4.61
16	67693939	ACD	c.367G>A	NM_001082486	p.D123N	rs142662151	6/8554		0.000255	probably damaging	tolerated	4.27
16	67694044	ACD	c.338G>A	NM_001082486	p.R113Q	rs142507451	14/8574	0.0005	0.001527	probably damaging	tolerated	-3.55
16	67694102	ACD	c.280G>A	NM_001082486	p.V94I	rs149365469	6/8584	0.0014	0.000255	probably damaging	tolerated	2.64
16	67694254	ACD	c.128G>A	NM_001082486	p.R43Q				0.000509	probably damaging	damaging	-2.44
16	67694333	ACD	c.49G>A	NM_001082486	p.A17T				0.001018	benign	damaging	1.57
16	67692855	ACD	c.879G>A	NM_001082486	p.W293X				0.000255	na	na	4.61
16	67693131	ACD	c.751+1G>T	NM_001082486					0.000255	na	na	
8	73934512	TERF1	c.599T>C	NM_017489	p.I200T				0.000255	probably damaging	damaging	4.29
8	73944309	TERF1	c.980A>G	NM_017489	p.Q327R				0.000255	benign	tolerated	1.78
8	73951417	TERF1	c.1106C>T	NM_017489	p.P369L	rs148824563	6/8592		0.000255	damaging	tolerated	-1.76
16	69390938	TERF2	c.1492G>A	NM_005652*	p.E498K	rs150757154	18/8582	0.0005	0.00229	probably damaging	tolerated	5.26
16	69395328	TERF2	c.1405G>A	NM_005652*	p.D469N				0.000255	probably damaging	tolerated	4.1
16	69400902	TERF2	c.1148C>T	NM_005652*	p.P383L				0.000255	probably damaging	tolerated	4.53
16	69401069	TERF2	c.981G>A	NM_005652*	p.M327I	rs141046852	1/8599		0.000255	benign	tolerated	-0.479
16	69404453	TERF2	c.773C>A	NM_005652*	p.T258N				0.000255	probably damaging	tolerated	0.697

14	24709052	TINF2	c.1307C>T	NM_001099274	p.A436V	rs369249473	1/8247		0.000255	benign	tolerated	-0.37
14	24709502	TINF2	c.1096C>A	NM_001099274	p.L366I				0.000255	probably damaging	tolerated	-0.796
14	24709724	TINF2	c.962G>C	NM_001099274	p.R321T				0.000509	damaging	damaging	-0.995
14	24709952	TINF2	c.734C>A	NM_001099274	p.S245Y	rs142777869	4/8304	0.0005	0.000255	benign	damaging	0.481
14	24710007	TINF2	c.679C>T	NM_001099274	p.L227F				0.000255	benign	tolerated	-0.456
14	24710064	TINF2	c.622T>C	NM_001099274	p.S208P	rs377436580	1/8267		0.000255	damaging	damaging	3.94
14	24711177	TINF2	c.216G>T	NM_001099274	p.Q72H				0.000255	damaging	damaging	-2.44
14	24711465	TINF2	c.74G>C	NM_001099274	p.G25A	rs202093758	15/8303		0.003053	probably damaging	tolerated	0.443
14	24711179	TINF2	c.214C>T	NM_001099274	p.Q72X				0.000255	na	na	2.46

*refers to canonical protein isoform Q15554-1

~ European American population

GERP++ is an estimate of the constrained elements in the human genome

Supplementary Table 8: Germline mutations in the *TERF, TERF2* and *TINF2* genes identified in melanoma families by exome sequencing.

Family ID	Chr	Start§	Gene	cDNA change	mRNA refSeq	Protein change	Cases in family	Genotyped cases who are carriers^	VAF ESP6500 ~	Polyphen2 †	SIFT †	GERP++
AUS8	8	73921434	TERF1	c.313G>C	NM_017489	p.A105P	3	1/1	0	probably damaging	damaging	4.85
AUS9	16	69418607	TERF2	c.356G>C	NM_005652*	p.S119C	2	2/2	0	probably damaging	damaging	5.52
AUS10	16	69400788	TERF2	c.1136C>G	NM_005652*	p.S379C	5	3/5	0	probably damaging	damaging	5.09
AUS11	14	24711475	TINF2	c.64C>G	NM_001099274	p.V22L	2	1/2	0	possibly damaging	damaging	5.41

§ genome build GRCh37, hg19

* refers to canonical protein isoform Q15554-1

^ includes obligate carriers

~European American population

†effect of nonsense mutations not predicted by either Polyphen2 or SIFT are annotated as truncating

GERP++ is an estimate of the constrained elements in the human genome

Supplementary Table 9: Histology of confirmed cancers in shelterin mutation-positive families

Family ID	ID	carrier	Year of birth	Tumour verification	Year of diagnosis	CMM type	CMM site	Clark level	Breslow Depth (mm)	Other cancer	Other cancer site
AUS1	001	yes	1904	cancer registry	1994					myeloid leukemia, NOS	bone marrow
AUS1	002	yes	1926	histology of primary	1983	SSM	right scapular	I			
AUS1	002	yes	1926	histology of primary	1966	NOS	upper arm	II			
AUS1	003	yes	1928	histology of primary	1984	SSM	face	II	0.25		
AUS1	005	yes	1959	histology of primary	1989	NOS	upper right arm	II	0.51		
AUS1	006	yes	1969	histology of primary	1994	SSM	upper right arm	II	0.38		
AUS1	011	na	1955	histology of metastatic disease	1982	NOS	right axillary lymph node	na			
AUS1	703	na	1960	histology of primary	1972					gemistocytic astrocytoma	frontal lobe
AUS2	001	yes	1942	histology of primary	1999					intraductal carcinoma noninfiltrating, NOS	right breast
AUS2	001	yes	1942	histology of primary	2000	NOS	left lower leg	I	0.3		
AUS2	005	yes	1940	histology of primary	2008					small cell carcinoma, NOS	right lung
AUS2	005	yes	1940	histology of metastatic disease	1975	unknown primary	NOS	na			
AUS2	007	yes	1938	histology of primary	2002					non-small cell carcinoma	left lung
AUS2	007	yes	1938	histology of primary	2000	in situ	back	I			
AUS2	007	yes	1938	histology of primary	1993	NOS	back	II	0.3		
AUS2	009	yes	1959	histology of primary	1985	SSM	leg	III	0.56		
AUS2	010	na	1966	histology of primary	1997	NOS	forehead	II	0.5		
AUS2	010	na	1966	cancer registry	1987	NOS	left scapular	na			

AUS2	010	na	1966	cancer registry	1987	NOS	leg	na			
AUS2	015	yes	1943	histology of primary	1994					mucin-producing adenocarcinoma	breast
AUS2	021	yes	1971	histology of primary	2007	nodular	right elbow	II	0.5		
AUS2	804	na	1923	death certificate	1923					neoplasm	lung
AUS2	901	na	1894	death certificate	1961					malignant epithelial tumour	prostate
AUS3	001	yes	1925	histology of primary	1984	SSM	back	II	0.5		
AUS3	001	yes	1925	histology of primary	1995					adenocarcinoma	colon
AUS3	001	yes	1925	death certificate	1997					leukemia	
AUS3	002	yes	1923	clinical record	1979	SSM	left forearm				
AUS3	002	yes	1923	histology of primary	1982					adenocarcinoma	colon
AUS3	701	na	1948	death certificate	1990					acute lymphoid leukemia	
AUS3	804	na	1927	histology of primary	1994	SSM	right preauricular	IIII	4		
AUS4	001	yes	1931	histology of primary	2003					adenocarcinoma,NOS	prostate
AUS4	001	yes	1931	histology of primary	1966	SSM	right shoulder	na			
AUS4	003	yes	1932	histology of primary	1968	NOS	upper left arm	III			
AUS4	003	yes	1932	histology of primary	2002	SSM	upper left arm	I			
AUS4	004	no	1967	histology of primary	1982	SSM	left calf	I			
AUS4	005	yes	1906	histology of primary	1994	spindle cell	right ankle	IV	2.8		
AUS4	005	yes	1906	histology of primary	1991	NOS	right calf	III	1.65		
AUS4	006	no	1937	histology of primary	1988	SSM	right scapular	I			
AUS4	012	yes	1956	histology of primary	2012					B-cell lymphoma	lymph node
AUS4	016	yes	1932	histology of primary	2001					adenocarcinoma,NOS	prostate

AUS4	016	yes	1932	histology of primary	1991	NOS	left calf	II	0.54		
AUS4	039	no	1935	histology of primary	2010	SSM	upper back of left arm	II	0.55		
AUS4	804	na	1903	death certificate	1980					neoplasm	colon
AUS5	001	yes	1926	histology of primary	2004	SSM	forehead	II	0.3		
AUS5	001	yes	1926	histology of primary	1986	SSM	upper right arm	II	0.4		
AUS5	001	yes	1926	histology of primary	1984	SSM	left thigh	II	0.74		
AUS5	001	yes	1926	histology of primary	1987	SSM	right knee	II	0.29		
AUS5	002	yes	1951	clinical record	1981	SSM	NOS				
AUS5	007	yes	1939	histology of primary	1995	SSM	right scapular	I			
AUS5	010	na	1929	histology of primary	1996					fibroxanthoma	ear
AUS5	013	no	1957	histology of primary	2003	SSM	left cheek	III	1		
AUS5	020	na	1938	histology of primary	1982					adenocarcinoma	rectum
AUS5	020	na	1938	histology of primary	2012					adenocarcinoma	prostate
AUS5	601	na	1949	death certificate	2006					Burkitt lymphoma, NOS	lymph node
AUS5	602	na	1969	death certificate	2001					acute myeloid leukemia	
AUS5	701	na	1919	histology of metastatic disease	1992	SSM	unknown				
AUS6	006281	yes	1961	histology of primary	1985	SSM	NOS	III	0.2		
AUS6	006281	yes	1961	histology of primary	2012					adenocarcinoma, in situ	cervix
AUS6	005682	yes	1969	histology of primary	1984	SSM	NOS	III	0.3		
AUS6	005682	yes	1969	histology of primary	1999					squamous intraepithelial neoplasia grade III	cervix
AUS7	001	yes	1924	histology of primary	2005					transitional meningioma	meninges, arachnoid
AUS7	001	yes	1924	histology of primary	2001					serous surface papillary carcinoma	ovary

AUS7	001	yes	1924	histology of primary	1987	SSM	left thigh	III	0.35		
AUS8	006126	yes	1924	histology of primary	1985	nodular	face	III	1.6		
AUS8	006126	yes	1924	histology of primary	1989	LMM	face	II	0.42		
AUS8	006126	yes	1924	histology of primary	1989	in situ	back	I	0		
AUS8	006126	yes	1924	histology of primary	2003	LMM	face	II	0.9		
AUS8	006126	yes	1924	histology of primary	2004					papillary adenocarcinoma	thyroid
AUS8	023818	no	1942	histology of primary	2010	in situ	anterior chest	I			
AUS8	023818	no	1942	histology of primary	1997					carcinoma, NOS	left breast
AUS8	023819	na	1944	histology of primary	2008	LMM	neck	II	0.26		
AUS8	052712	na	1948	histology of primary	1980	SSM	lower limb	II	0.25		
AUS8	052712	na	1948	histology of primary	1980	SSM	lower limb	IIII	2.6		
AUS9	009262	yes	1918	histology of primary	1987	SSM	right ankle	IIII	0.7		
AUS9	011115	yes	1942	histology of primary	1988	SSM		III	1.8		
AUS9	029671	yes	1944	histology of primary	2004					intraepithelial neoplasm	prostate
AUS10	001	yes	1900	histology of primary	1984	nodular	left foot	V	6.4		
AUS10	002	yes	1922	histology of primary	2001	in situ	nose	I			
AUS10	002	yes	1922	histology of primary	2001	SSM	left leg	I			
AUS10	003	yes	1927	histology of primary	2001					infiltrating duct carcinoma, NOS	right breast
AUS10	007	no	1937	histology of primary	2003	LMM	left shoulder	II	0.73		
AUS10	010	yes	1951	histology of primary	2004	SSM	left shoulder	I			

AUS10	010	yes	1951	histology of primary	2013					adenocarcinoma,NOS	lung
AUS10	701	no	1916	histology of primary	1982	SSM	right calf	IV	1		
AUS10	704	na	1936	histology of primary	1993					adenocarcinoma,NOS	stomach
AUS10	704	na	1936	histology of primary	1991					infiltrating duct and lobular carcinoma	left breast
AUS10	704	na	1936	histology of primary	1987					intraductal carcinoma noninfiltrating	left breast
AUS10	808	na	1911	death certificate	1993					carcinoma, metastaic, NOS	pancreas
AUS11	001	yes	1953	histology of primary	1984	SSM	back	III	0.55		
AUS11	001	yes	1953	histology of primary	2006					adenocarcinoma,NOS	large bowel
AUS11	002	no	1944	histology of primary	1989	nodular	left forearm	II	0.74		
UK1	65F	yes	1947	statistics office	1990	SSM	leg	na	na		
UK1	193F	yes	1915	statistics office/ death certificate	1952	NOS	na	na	na		
SE1	564	no	1920	histology of primary	1997						
SE1	614	no	1945	histology of primary	1985						
SE1	648	Yes	1954	histology of primary	1997						
DK1	1	yes	1943	histology of primary	1986	SSM	left shoulder	II			
DK1	1	yes	1943	histology of primary	2003	SSM	knee	II	0.34		
DK1	2	yes	1932	histology of primary	1992	SSM	arm	I			
DK1	3	yes	1930	histology of primary	2006	SSM	knee				
DK1	3	yes	1930	histology of primary	2010	SSM	calf	II	0.33		
DK1	3	yes	1930	histology	2012					B-cell lymphoma, diffuse	lumbal bone marrow
DK1	4	no	1964	histology of primary	2013	in situ	anterior chest	I			

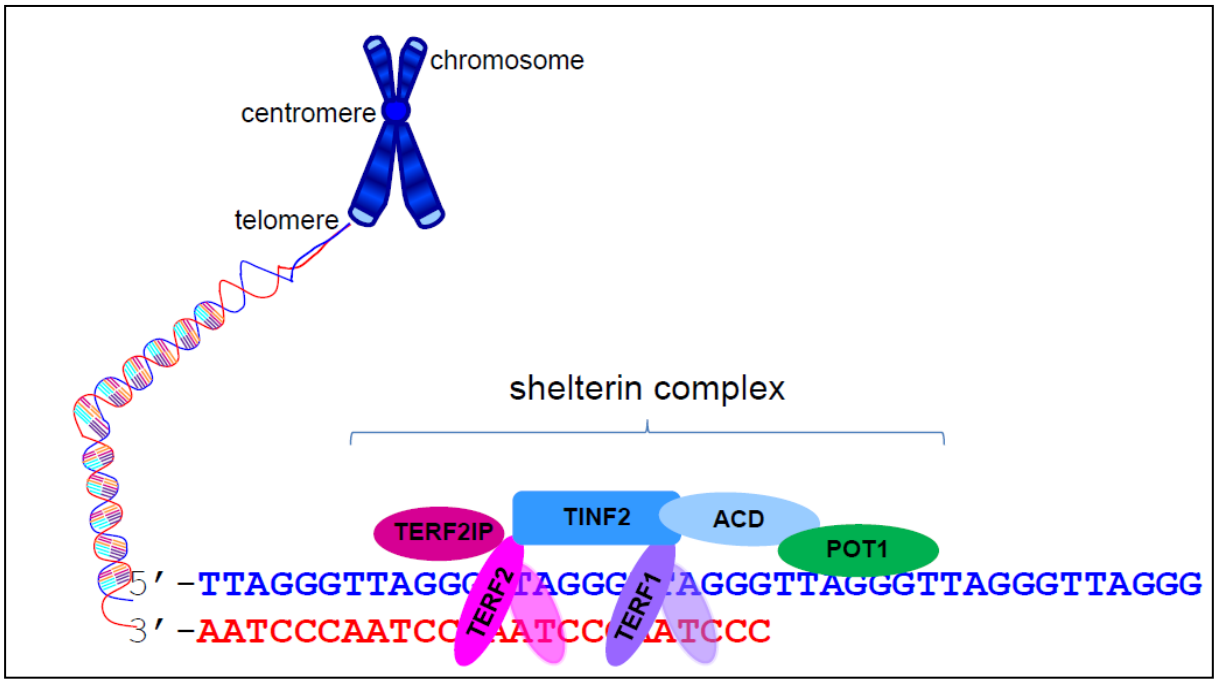
DK1	5	na	1903	death certificate	1980					adenocarcinoma	colon
DK1	6	na	1938	histology of primary	2009					adenocarcinoma and squamous carcinoma	lung

na is not available

NOS is not otherwise specified

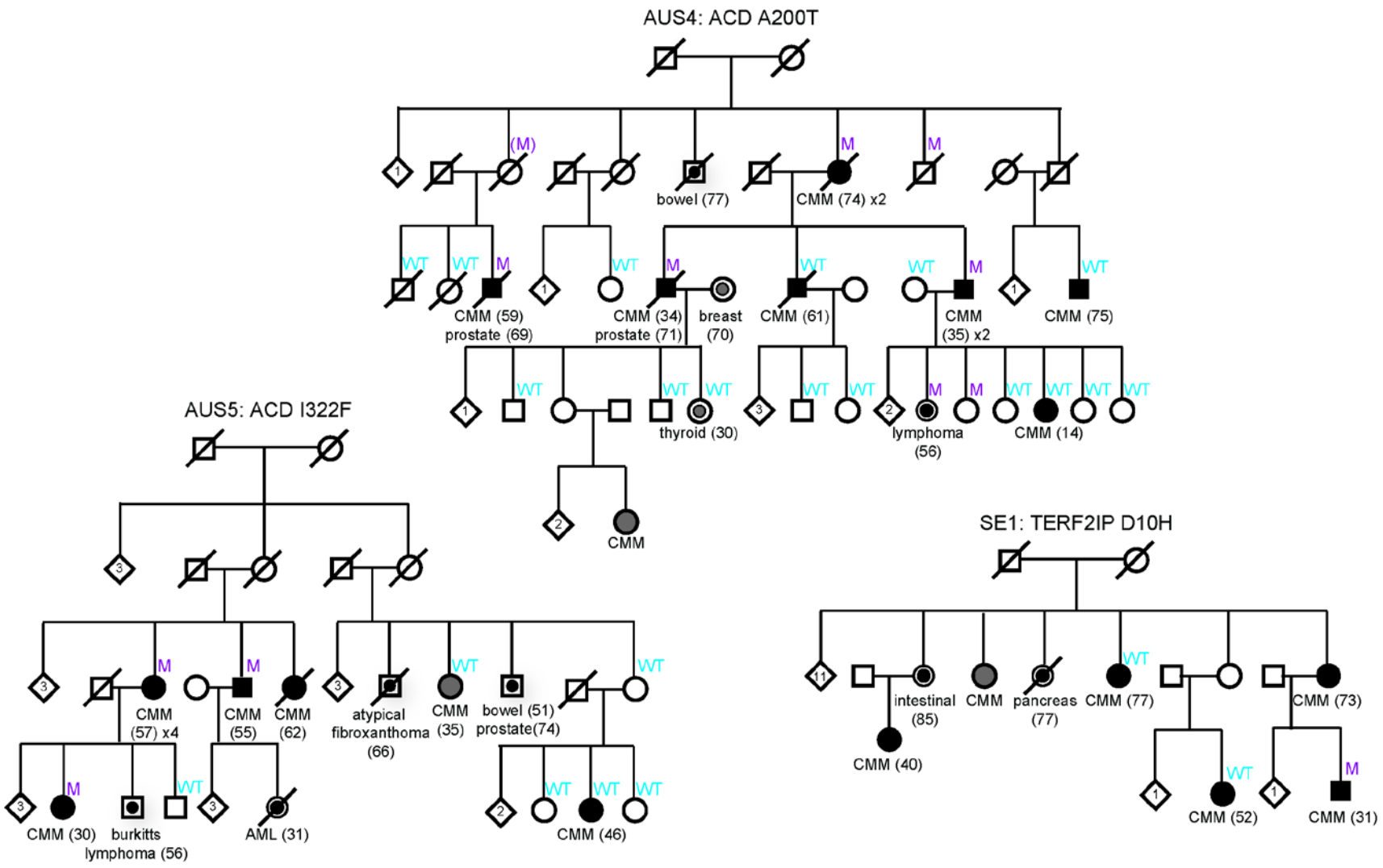
SSM is superficial spreading melanoma

LMM is lentigo maligna melanoma

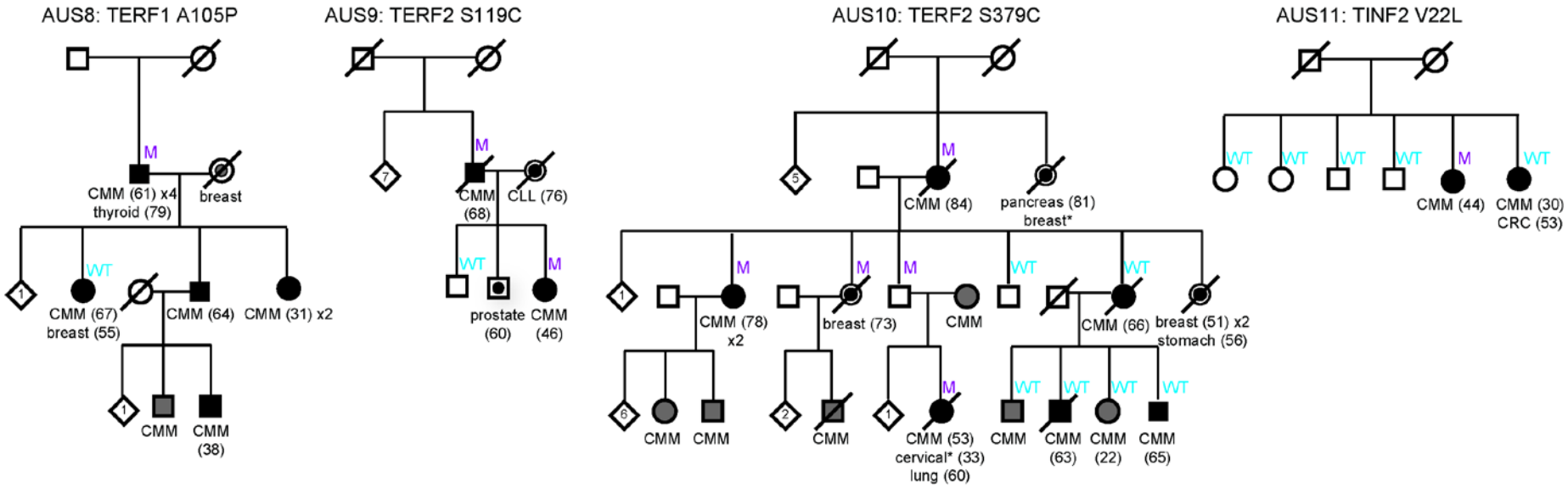


Supplementary Figure 1: The shelterin complex. Telomeres consist of variable numbers of TTAGGG repeats with a single stranded telomeric 3'-tail. This 3' tail has two key functions: 1) binding of the shelterin complex telomere specific proteins, which cap chromosome ends, promoting genome stability; 2) providing the primer for telomere extension by the telomerase complex. The shelterin complex consists of six proteins: TERF1 and TERF2, which form homodimers and anchor the complex to the double stranded DNA (dsDNA) telomere. POT1 contains a single DNA binding domain, which is highly specific to the single stranded telomeric 3'-tail and has the recognition sequence 5'-(T)TAGGGTTAG-3'. In the presence of the telomerase complex POT1 does not bind to the single stranded DNA overhang, resulting in telomere lengthening. TERF2IP binds exclusively to TERF2, augmenting its dsDNA binding capacity. TINF2 has a critical role in the shelterin complex, binding to dsDNA-interacting TERF1 and TERF2 homodimers and recruiting ACD to the complex. ACD binds to POT1.

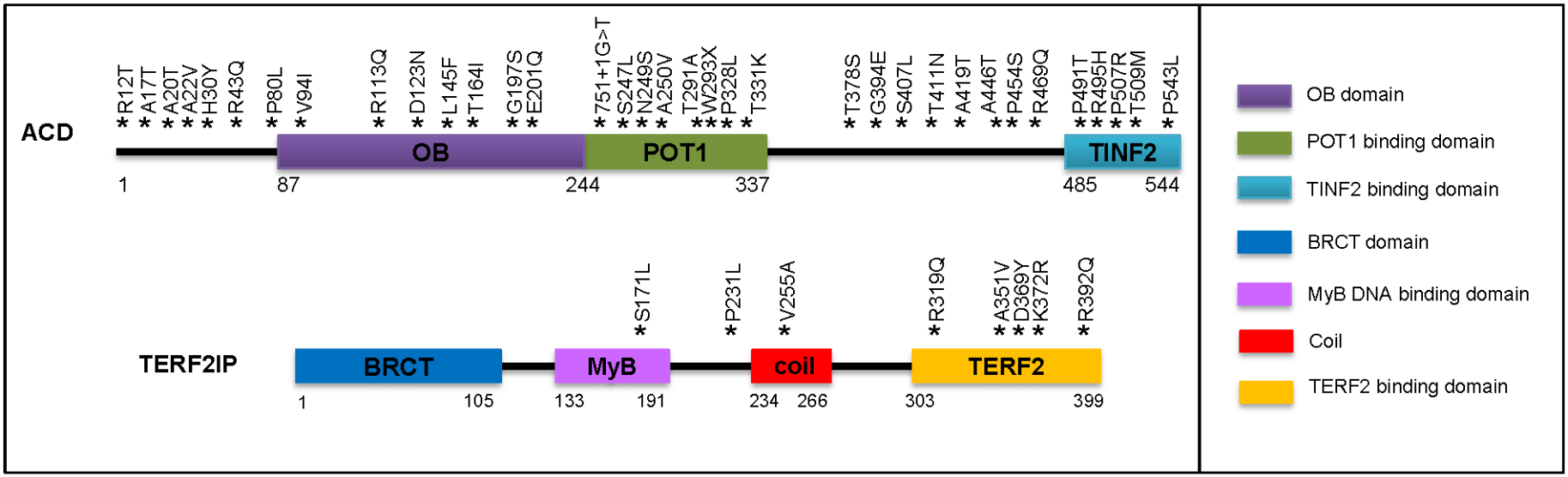
A


AUS4: ACD A200T
bowel (77)
CMM (74) x2
CMM (59) prostate (69)
CMM (34) prostate (71)
breast (70)
CMM (61)
CMM (35) x2
CMM (75)
thyroid (30)
lymphoma (56)
CMM (14)
CMM
AUS5: ACD I322F
CMM (57) x4
CMM (55)
CMM (62)
atypical fibroxanthoma (66)
CMM (35)
bowel (51) prostate(74)
CMM (30)
burkitts lymphoma (56)
AML (31)
CMM (46)
SE1: TERF2IP D10H
intestinal (85)
CMM
pancreas (77)
CMM (77)
CMM (73)
CMM (40)
CMM (52)
CMM (31)

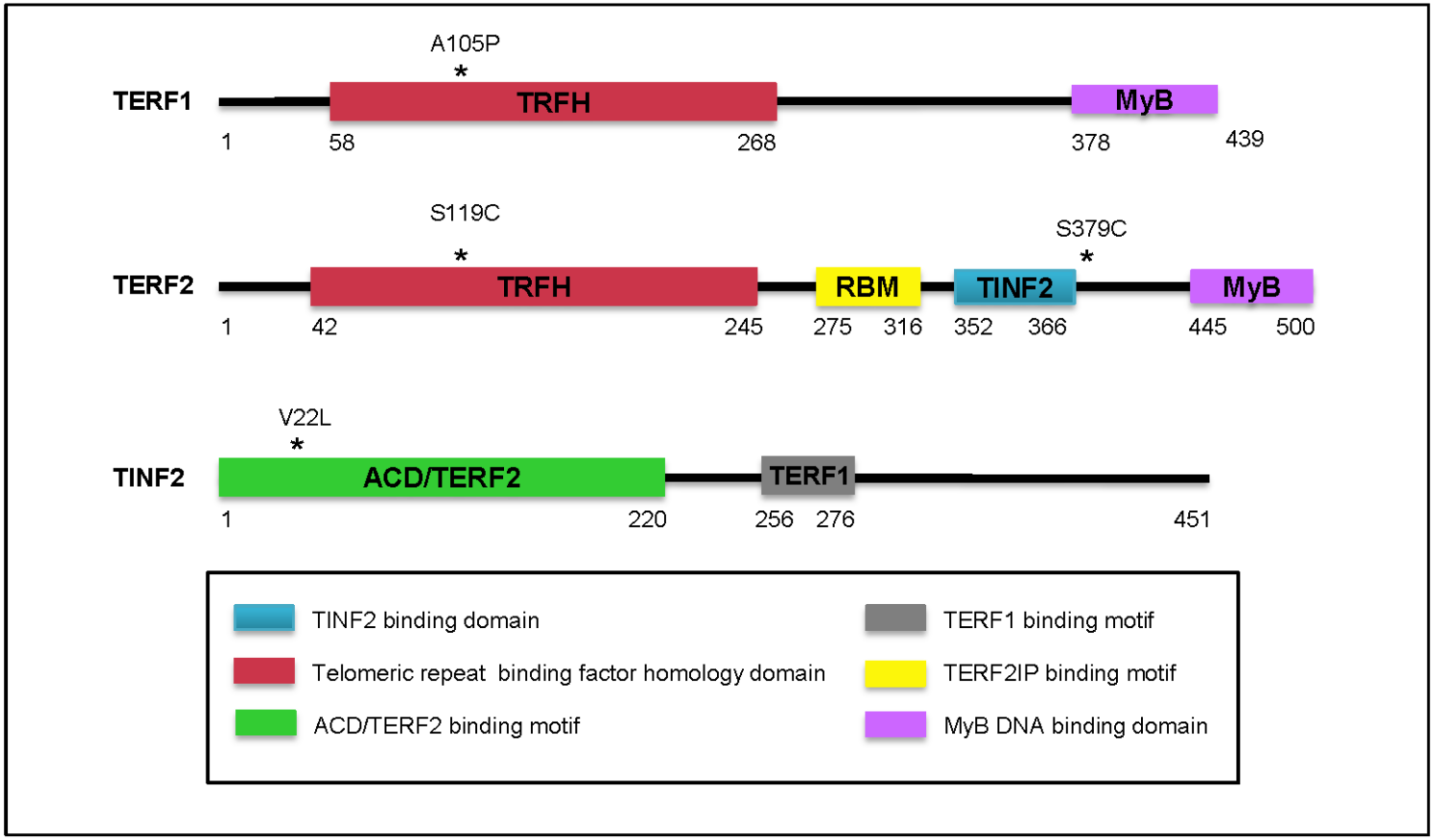
B



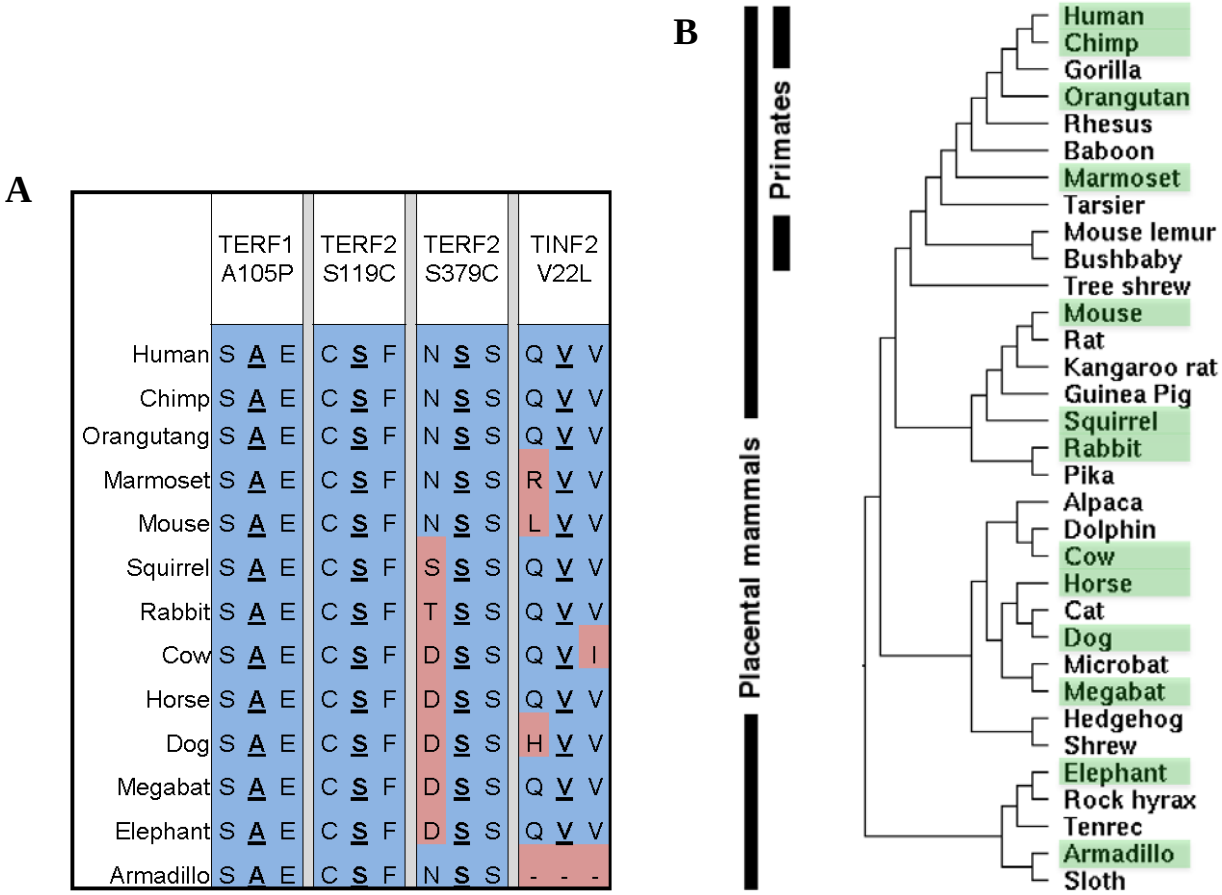
Supplementary Figure 2: Variants identified in *ACD* and *TERF2IP* that incompletely segregate with melanoma (A), and segregation analysis of variants identified in TINF2, TERF1 and TERF2 (B). Legend as for Figure 1.



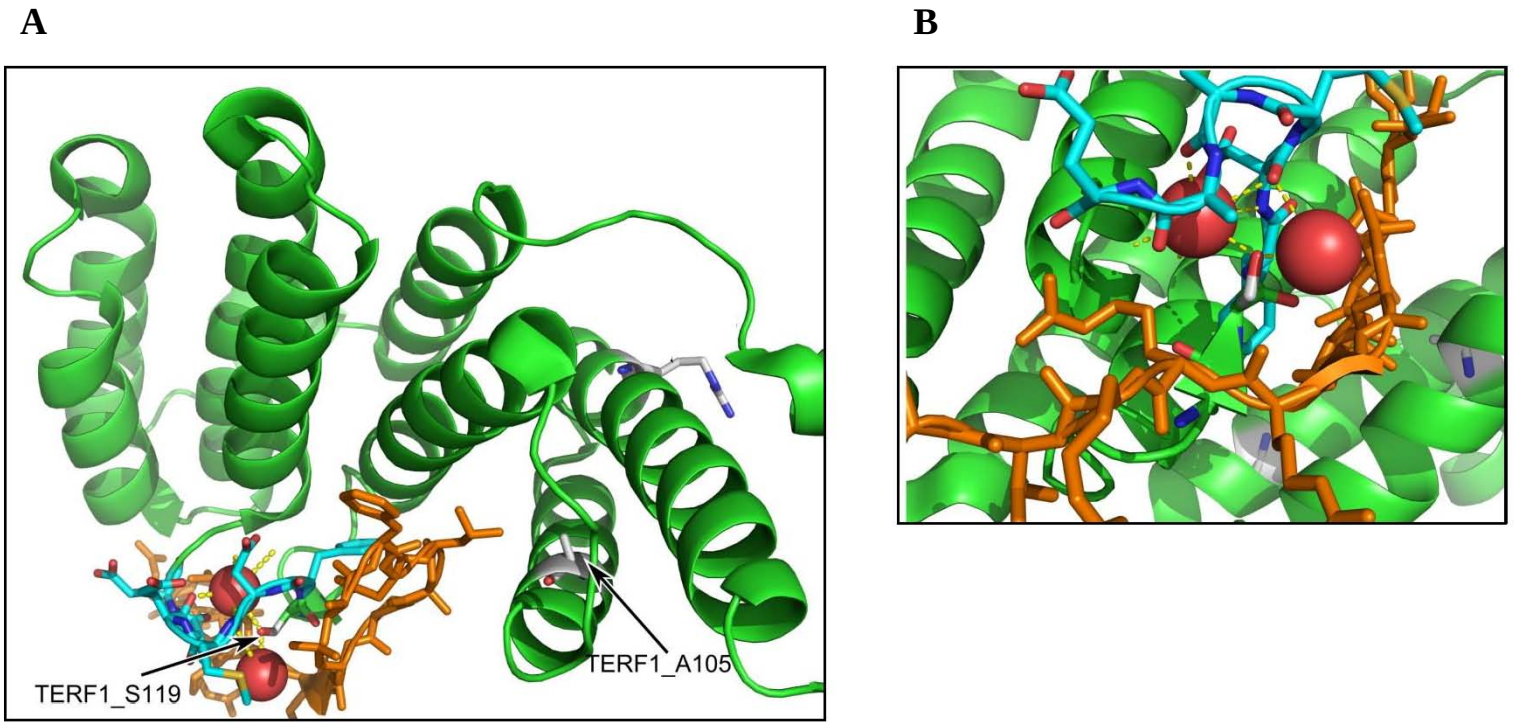
Supplementary Figure 3: Relative location of germline variants in NHLBI Exome Sequencing Project, UK10K and Danish controls with respect to ACD and TERF2IP protein domains.



Supplementary Figure 4: Relative location of germline variants in melanoma families with respect to TERF1, TERF2 and TINF2 protein domains.



Supplementary Figure 5: Conservation of TERF1, TERF2 and TINF2 variants in melanoma families across placental mammals (A) and spanning species clades (B).



Supplementary Figure 6: Protein structural modelling. (**a**) Model structure from TERF1 bound to TINF2 (PDB id 3BU8) showing the predicted location of new variants. The model structure is shown in green, except for the TINF2-interacting loop, in blue. Variant residues are shown in grey. TINF2 is shown in orange. Two water molecules located in the interface between both proteins are displayed as red spheres. (**b**) Closer view of the interaction between TERF1 and TINF2 showing the Ser119 residue.

Chapter 6

Prevalence of germline *BAP1* mutations in a population-based sample of uveal melanoma cases

Manuscript 5

Lauren G. Aoude, Claire M. Vajdic, Anne Kricker, Bruce Armstrong and Nicholas K. Hayward. Prevalence of germline *BAP1* mutation in a population-based sample of uveal melanoma cases. *Pigment Cell & Melanoma Research*. 2013 Mar; 26(2):278-9.

6.1 Relevance to thesis aims

This chapter addresses Aim 1 of this thesis by detailing the prevalence of *BAP1* mutation in an Australian population-based sample of uveal melanoma cases. To date, *BAP1* is the only known UMM predisposition gene. The aim of the study was to determine the contribution of germline *BAP1* mutation to an unselected sample of individuals with UMM. A group of 66 blood samples from UMM cases was genotyped for *BAP1* mutation using Sanger sequencing methods. The project revealed that 2 out of 66 individuals harboured novel germline mutations in *BAP1*. The findings were published as a letter to the editor in *Pigment Cell & Melanoma Research*. This was the first published study to look at *BAP1* in a sporadic sample of UMM cases to estimate its contribution to disease development. Germline *BAP1* mutations are estimated to occur in 3% of population-based UMM cases.

6.2 Contribution of candidate

I was involved in the initial concept and study design which was carried out in conjunction with Nicholas Hayward. I performed all laboratory experiments including the primer design for the 17 exons of *BAP1*, PCR, Sanger sequencing and analysis of chromatograms. I wrote the manuscript under the supervision of Nicholas Hayward.

6.3 Contribution of other authors

The samples for this study were collected by Claire M. Vajdic, Anne Kricker, and Bruce Armstrong. These samples were originally collected as part of a study to better characterise UMM occurrence in Australia [132]. Nicholas Hayward contributed to the study concept and design and helped write and edit the manuscript. All authors read and approved of the manuscript.

Prevalence of germline *BAP1* mutation in a population-based sample of uveal melanoma cases

Lauren G. Aoude[1], Claire M. Vajdic[2], Anne Kricker[3], Bruce Armstrong[3,4] and Nicholas K. Hayward[1]

[1]Oncogenomics Laboratory, Queensland Institute of Medical Research, Brisbane, QLD, Australia
[2]Adult Cancer Program, Lowy Cancer Research Program, Prince of Wales Clinical School, Medicine, University of New South Wales, Sydney, NSW, Australia
[3]Sydney School of Public Health, University of Sydney, Sydney, NSW, Australia
[4]Sax Institute, Sydney, NSW, Australia
Correspondence Lauren Aoude, e-mail: lauren.aoude@qimr.edu.au

doi: 10.1111/pcmr.12046

Dear Editor,

In a seminal study by Harbour et al. (2010), somatic mutations of *BAP1* (BRCA-1 associated protein 1) were found to occur in 84% of metastasizing uveal melanomas. Additionally, this study reported one germline mutation. Subsequent studies have shown that germline mutations of BAP1 predispose to uveal and cutaneous melanoma as well as mesothelioma and a range of other tumour types (Abdel-Rahman et al., 2011; Testa et al., 2011; Wadt et al., 2012). To date, there have been no studies assessing the contribution of *BAP1* mutation to a population-based sample of uveal melanoma (UMM) cases, although Tsao and co-workers have shown that *BAP1* contributes to a small proportion of selected UMM cases from a clinic-based sample (Njauw et al., 2012). Here, we sought to determine the prevalence of germline *BAP1* mutations in a population-based sample of Australian UMM cases to gain a better understanding of the relative contribution of *BAP1* mutation to UMM susceptibility.

We examined 66 UMM cases unselected for family history of UMM sampled from New South Wales, Australia. Individuals were eligible for inclusion in the study if they were diagnosed with primary UMM at the age of 50 or younger or if they were diagnosed with bilateral UMM. This sample comprises 51 individuals with choroidal melanoma, five with melanoma of the ciliary body, seven that presented with melanoma of the iris and three samples of unknown site. UMM that presented on the conjunctiva were excluded from the current study. Previous published work showed that none of these individuals carries a deleterious mutation of *CDKN2A* (Vajdic et al., 2003).

Sanger sequencing was used to interrogate the 17 exons of *BAP1* for disease causing mutations. For information detailing DNA extraction, PCR and Sanger sequencing methods please refer to Supporting Information Appendix S1 and Table S1. Sequencing results showed that 2/66 sporadic UMM DNA samples harboured protein altering *BAP1* mutations (see Figure S1). One resulted in a frameshift in exon 13, p.D494fs, caused by a two base pair deletion at c.1480_1481delGA. This occurred in an individual with a primary tumour of the ciliary body. The second mutation occurred in exon 14 at c.1806g>c and resulted in a p.E602D substitution. The individual harbouring this mutation had a tumour of the choroid. SNPs have been reported in dbSNP occurring at the same location as this variant. This however is a novel protein change that has not been previously reported. Although this is a conservative amino acid change and not likely to truncate the protein through a cryptic exonic splice site, it is located in the BRCA1 binding domain, which may be functionally relevant and deleterious to protein activity despite the fact that the amino acid substitution seems conservative. A synonymous variant, L334L (rs28997577), was found in the individual carrying the E602D mutation. This occurs at low frequency in the general population and is unlikely to be associated with UMM risk in this person. Aside from variants found in the coding region, there were also several other known SNPs that were identified in intronic regions of the gene as well as other unreported variants in intronic regions (see Table S2). These were excluded from further consideration as the reported SNPs occur in the general population, and the other mutations are unlikely to be disease causing.

The overall outcome of our screening revealed that 2/66 (3%) unselected UMM cases had novel mutations in *BAP1* that may have contributed to their disease risk. The findings of our study are consistent with the data published by Njauw et al. (2012), who found germline *BAP1* mutations in approximately 4% of the sporadic UMM cohort screened in their study. These two studies show remarkably consistent rates of *BAP1* mutation in sporadic UMM cases but also suggest that other genes must be involved in the development of UMM because *BAP1* mutation accounts for susceptibility in only a very small proportion of UMM cases. Further research into identifying genes responsible for UMM predisposition is needed.

Acknowledgements

Funding for this project was provided by the National Health and medical Research Council of Australia. Infrastructure support was

provided by the Cancer Council, New South Wales. The authors would also like to thank the participants of the study for making this research possible.

References

Abdel-Rahman, M.H., Pilarski, R., Cebulla, C.M., Massengill, J.B., Christopher, B.N., Boru, G., Hovland, P., and Davidorf, F.H. (2011). Germline BAP1 mutation predisposes to uveal melanoma, lung adenocarcinoma, meningioma, and other cancers. J. Med. Genet. *48*, 856–859.

Harbour, J.W., Onken, M.D., Roberson, E.D., Duan, S., Cao, L., Worley, L.A., Council, M.L., Matatall, K.A., Helms, C., and Bowcock, A.M. (2010). Frequent mutation of BAP1 in metastasizing uveal melanomas. Science *330*, 1410–1413.

Njauw, C.N., Kim, I., Piris, A., Gabree, M., Taylor, M., Lane, A.M., Deangelis, M.M., Gragoudas, E., Duncan, L.M., and Tsao, H. (2012). Germline BAP1 inactivation is preferentially associated with metastatic ocular melanoma and cutaneous-ocular melanoma families. PLoS ONE *7*, e35295.

Testa, J.R., Cheung, M., Pei, J. et al. (2011). Germline BAP1 mutations predispose to malignant mesothelioma. Nat. Genet. *43*, 1022–1025.

Vajdic, C.M., Kricker, A., Giblin, M., Mckenzie, J., Aitken, J., Giles, G. G., and Armstrong, B.K. (2003). Incidence of ocular melanoma in Australia from 1990 to 1998. Int. J. Cancer *105*, 117–122.

Wadt, K., Choi, J., Chung, J.Y. et al. (2012). A cryptic BAP1 splice mutation in a family with uveal and cutaneous melanoma, and paraganglioma. Pigment Cell Melanoma Res. *25*, 815–818.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Methods.

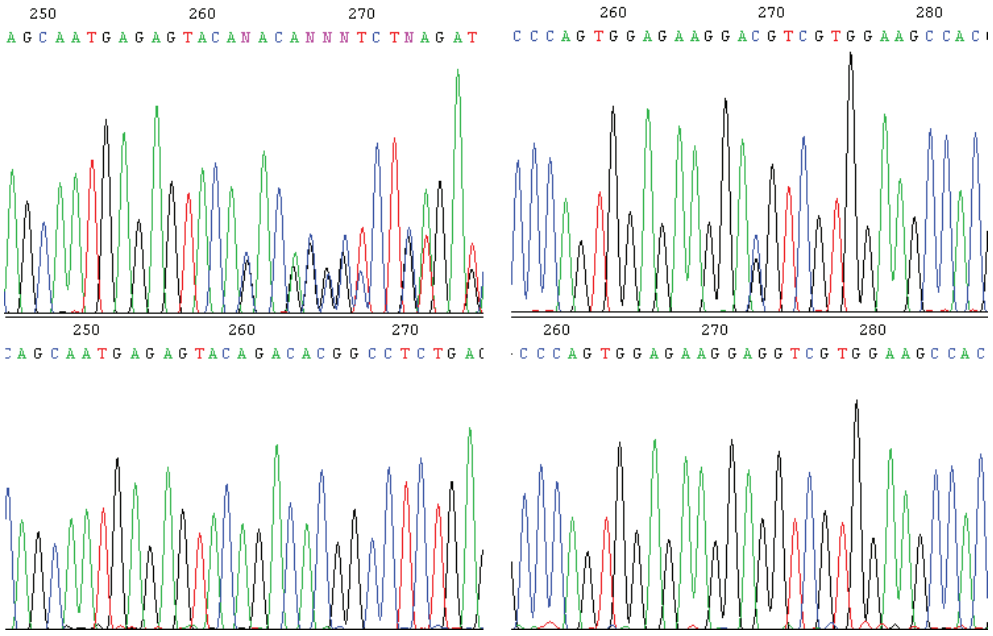
Figure S1. Novel germline *BAP1* mutations. In (a), the top left panel shows the c.1480_1481delga (P494fs) and the bottom left shows the wild type trace. The top right panel of (a) shows the C.1806g > c mutation with the wild type trace directly below it. (b) The truncating and missense mutations in relation to BAP1 protein domains. The figure is not drawn to scale.

Table S1. BAP1 primer sequences.

Table S2. All germline BAP1 nucleotide changes observed in a sample of sporadic uveal melanoma cases.

Supplementary Figure

1a



1b

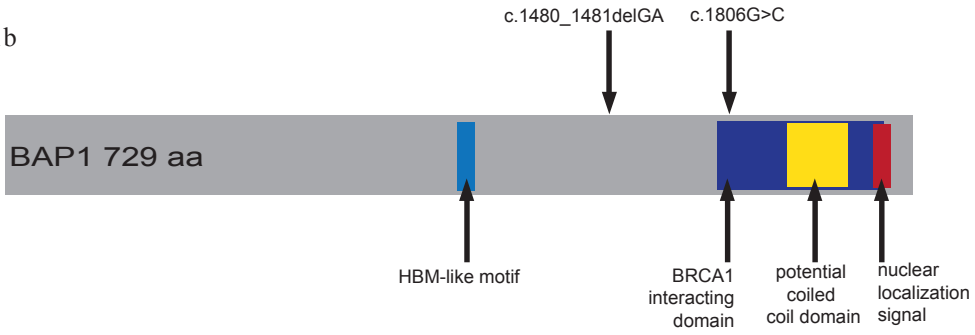


Figure S1: Novel germline *BAP1* mutations

In figure S1a, the top left panel shows the c.1480_1481delga (P494fs) and the bottom left shows the wild type trace. The top right panel of 1a shows the C.1806g>c mutation with the wild type trace directly below it. Figure 1b shows the truncating and missense mutations in relation to BAP1 protein domains. The figure is not drawn to scale.

Chapter 7

A *BAP1* mutation in a Danish family predisposes to uveal melanoma and other cancers

Manuscript 6

Lauren G. Aoude, Karin Wadt, Anders Bojesen, Dorthe Crüger, Åke Borg, Jeffrey M. Trent, Kevin M. Brown, Anne-Marie Gerdes, Göran Jönsson, Nicholas K. Hayward. A *BAP1* mutation in a Danish family predisposes to uveal melanoma and other cancers. *PLOS One*. 2013 Aug; 8(8):e72144

7.1 Relevance to thesis aims

This chapter describes a Danish family with predominantly uveal melanoma but also a range of other tumour types, including lung, neuroendocrine, stomach and breast cancer, as well as pigmented skin lesions. There are a total of 7 UMM cases and one CMM case in the family presented. This study addresses Aims 2 and 3 of my thesis by determining the contribution of novel melanoma risk genes to UMM development in a high-risk UMM family.

An exome sequencing approach identified a *BAP1* splice mutation located at c.581-2A>G, which leads to a premature truncation of the protein in an individual with uveal melanoma. This mutation was carried by several other family members with UMM, CMM, breast cancer and lung cancer. The role of *BAP1* as a tumour suppressor was also supported by LOH experiments which showed loss of the wild type allele in a carrier with UMM. This finding is consistent with other published reports of frameshift aberrations in mutation carriers. The finding expands on the growing profile of *BAP1* as an important uveal and cutaneous melanoma tumour suppressor gene and suggests its involvement in the development of lung, and stomach cancer. The literature describes germline mutations in *BAP1* in relation to several forms of cancer including: UMM, CMM, mesothelioma and renal cell carcinoma, as well as benign melanocytic tumours [103-105, 107-111, 133, 134]. This report suggests the phenotype for germline *BAP1* mutation carriers may extended even further into other cancer types.

Additionally, this study looked at the *BAP1* splice variant in a case-control cohort to assess the possibility that this could contribute to predisposition in the general population (Aim 5). The results showed that this mutation is likely to be 'private' to the discovery family as it was not seen in the 1,655 cases from the QFMP or the 1,596 controls genotyped. Overall, this study expands the role of *BAP1* in cancer development and further describes its role as a high-risk predisposition gene for UMM development.

7.2 Contribution of candidate

I performed the exome data analysis and the Sanger sequencing validation of the variant. I also performed co-segregation analysis to determine whether the variant segregated with disease and whether it would be seen in family members with other cancer types. I genotyped the case-control sample using iPLEX chemistry to determine whether this variant occurred in any other probands. I wrote the manuscript with the guidance of Nicholas Hayward.

7.3 Contribution of other authors

Clinical data on the Danish family was collected by Karin Wadt. Samples were contributed by Anders Bojesen, Dorthe Crüger and Åke Borg. Göran Jönsson performed the RT-PCR, LOH analysis and contributed to the study design. Kevin M. Brown and Jeffrey M. Trent also contributed to the study design. Nicholas Hayward contributed to the concept and design of the study and helped write and edit the manuscript. All authors read and approved of the manuscript.

OPEN ACCESS Freely available online

PLOS ONE

A *BAP1* Mutation in a Danish Family Predisposes to Uveal Melanoma and Other Cancers

Lauren G. Aoude[1,2]*, Karin Wadt[3], Anders Bojesen[4], Dorthe Crüger[4], Åke Borg[5], Jeffrey M. Trent[6], Kevin M. Brown[6,7], Anne-Marie Gerdes[3], Göran Jönsson[5], Nicholas K. Hayward[1]

1 Oncogenomics Laboratory, Queensland Institute of Medical Research, Brisbane, Queensland, Australia, **2** University of Queensland, Brisbane, Queensland, Australia, **3** Department of Clinical Genetics, Rigshospitalet, Copenhagen, Denmark, **4** Department of Clinical Genetics, Vejle Hospital, Vejle, Denmark, **5** Department of Clinical Sciences Lund, Division of Oncology, Lund University, Lund, Sweden, **6** Translational Genomics Research Institute, Phoenix, Arizona, United States of America, **7** Laboratory of Translational Genomics, National Cancer Institute, Bethesda, Maryland, United States of America

Abstract

Truncating germline mutations in the tumor suppressor gene BRCA-1 associated protein-1 (*BAP1*) have been reported in families predisposed to developing a wide range of different cancer types including uveal melanoma and cutaneous melanoma. There has also been an association between amelanotic tumor development and germline *BAP1* mutation suggesting a possible phenotypic characteristic of *BAP1* mutation carriers. Though there have been many types of cancer associated with germline *BAP1* mutation, the full spectrum of disease association is yet to be ascertained. Here we describe a Danish family with predominantly uveal melanoma but also a range of other tumor types including lung, neuroendocrine, stomach, and breast cancer; as well as pigmented skin lesions. Whole-exome sequencing identified a *BAP1* splice mutation located at c.581-2A>G, which leads to a premature truncation of *BAP1* in an individual with uveal melanoma. This mutation was carried by several other family members with melanoma or various cancers. The finding expands on the growing profile of *BAP1* as an important uveal and cutaneous melanoma tumor suppressor gene and implicates its involvement in the development of lung, and stomach cancer.

Citation: Aoude LG, Wadt K, Bojesen A, Crüger D, Borg Å, et al. (2013) A *BAP1* Mutation in a Danish Family Predisposes to Uveal Melanoma and Other Cancers. PLoS ONE 8(8): e72144. doi:10.1371/journal.pone.0072144

Editor: Soheil S. Dadras, University of Connecticut Health Center, United States of America

Received February 27, 2013; **Accepted** July 5, 2013; **Published** August 19, 2013

Funding: This work was supported by the Melanoma Research Alliance, the Swedish Cancer Society, the Swedish Research Council, the Berta Kamprad Foundation, the National Health and Medical Research Council (NHMRC) of Australia, and the Fight for Sight, Denmark. NKH and LGA receive salary support through a NHMRC Senior Principal Research Fellowship and ANZ Ph.D. scholarship, respectively. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: lauren.aoude@qimr.edu.au

Introduction

BRCA-1 associated protein-1 (*BAP1*) is a tumor suppressor gene located on chromosome 3, a region that has been linked to uveal melanoma (UMM). In the first report to make an association between *BAP1* and UMM Harbour and colleagues used a next-generation sequencing approach to exome-sequence two tumors from patients with metastatic UMM [1]. *BAP1* was found to be the only gene that was mutated on chromosome 3 in each of these samples and both mutations led to truncation of the protein. They next interrogated 57 UMM tumors using Sanger sequencing and found that *BAP1* mutations occurred predominantly in tumors that had metastasized. Overall, 26/31 of the metastasizing (high risk) tumors had inactivating *BAP1* somatic mutations compared to only 1/26 of the low risk (non-metastatic) group. For the 20 samples with a matched normal DNA sample, almost all *BAP1* mutations were found to be acquired somatically, the exception being a single case with a matching germline frameshift mutation. This mutation introduced the possibility that *BAP1* defects could predispose to UMM. Leading on from the study by Harbour et. al., several groups have looked at the risk of disease conferred by germline *BAP1* mutation. Testa and colleagues investigated *BAP1* mutations in two American families presenting with mesothelioma and who had no contact with any of the known environmental risk factors for this disease [2]. They found two different frameshift mutations in *BAP1* responsible for the elevated risk of mesothelioma in these individuals. Interestingly, these families also presented with a range of cancers that included UMM and cutaneous melanoma (CMM). A second study also supported these findings by interrogating germline DNA from a family that presented with mesothelioma, meningioma and UMM [3]. In this family a nonsense mutation caused premature truncation of the BAP1 protein in carriers. A report by Wiesner showed *BAP1* mutations in two families that had an autosomal dominant syndrome characterized by UMM, CMM and atypical benign melanocytic nevi [4]. The two families presented with unique frameshift mutations that co-segregated with the disease phenotype. In a follow up report, Wiesner and colleagues identified a third family with a *BAP1* mutation that co-segregated with mesothelioma and also showed evidence of a melanocytic lesion in a mutation carrier [5]. Other groups have also described *BAP1* mutations in individuals presenting with atypical intradermal tumors, proposing that these lesions may be a phenotypic characteristic of *BAP1* mutation carriers [6,7]. A recent study of Portuguese siblings with a rare subtype of epithelioid mesothelioma uncovered a germline *BAP1* mutation as the possible cause of the only known familial clustering of well-differentiated papillary mesothelioma. Notably one of the siblings also developed UMM

[8]. Recently a cryptic splice mutation in *BAP1* was found to co-segregate in a Danish family with UMM, paraganglioma and atypical CMM [9]. *BAP1* mutation has also been implicated in the development of renal cell carcinoma and some of the relatives of these cases have had UMM or CMM also [10]. As evidenced by the literature, there is a large tumor spectrum that appears to accompany *BAP1* germline mutation.

In this paper we report a Danish family predisposed to developing UMM predominantly but also presents with a wide range of seemingly unrelated tumors. There are 7 family members that have had UMM, with the youngest age of onset being 20. One developed UMM at the age of 30 and later went on to develop breast cancer at age 45. There is an individual with lung cancer who had two children that developed melanoma, one UMM and the other CMM. Another individual developed UMM at the age of 69 and has two children that were diagnosed with UMM at ages 27 and 41. There is a single instance of diffuse infiltrating gastric cancer and this individual also had a child with UMM, diagnosed at age 57. Finally, another case of UMM was diagnosed at age 56. Refer to Figure 1 for a family pedigree which identifies the cancer history of each individual.

Aside from the striking incidence of cancer, this family presents with other clinical features that have become indications of *BAP1* mutation. Four of the five individuals with UMM died of liver metastasis derived from the UMM. Individual III:6, who was diagnosed with UMM at age 20, had an isolated iris melanoma and later developed three basal cell carcinomas. She is currently 61 years old and well. Individual III:8, with CMM diagnosed at 35 had a tumor localized in the occipital region, which most likely was a primary CMM but metastasis could not be excluded. She died at 39 years of age with disseminated CMM.

Methods

Ethics Statement

Ethics approval for this project was granted by the Human Research Ethics Committees of the Queensland Institute of Medical Research and Lund University. Written consent was obtained for each individual.

Samples

Whole blood was collected from disease affected individuals in dense uveal melanoma families. Genomic DNA was extracted using standard methods. DNA samples were then used for exome and Sanger sequencing experiments.

Whole-Exome Sequencing

In order to find disease-associated variants, whole-exome sequencing was carried out on key individuals representative of the tumor burden in this family. Two UMM cases that were three meioses apart (III:14 and III:17, Figure 1) were selected and sequencing was carried out using an Illumina Hiseq 2000 and an Illumina TruSeq Exome Enrichment Kit. The reads that were generated were mapped to reference genome UCSC hg19 using the BWA alignment algorithm [11]. SAMTOOLS was then used to detect the SNPs and indels [12]. This process generates large volumes of data and in order to sculpt the long list of resulting variants into a manageable panel of variants several filtering criteria were applied to the data set. Firstly, any variants that were

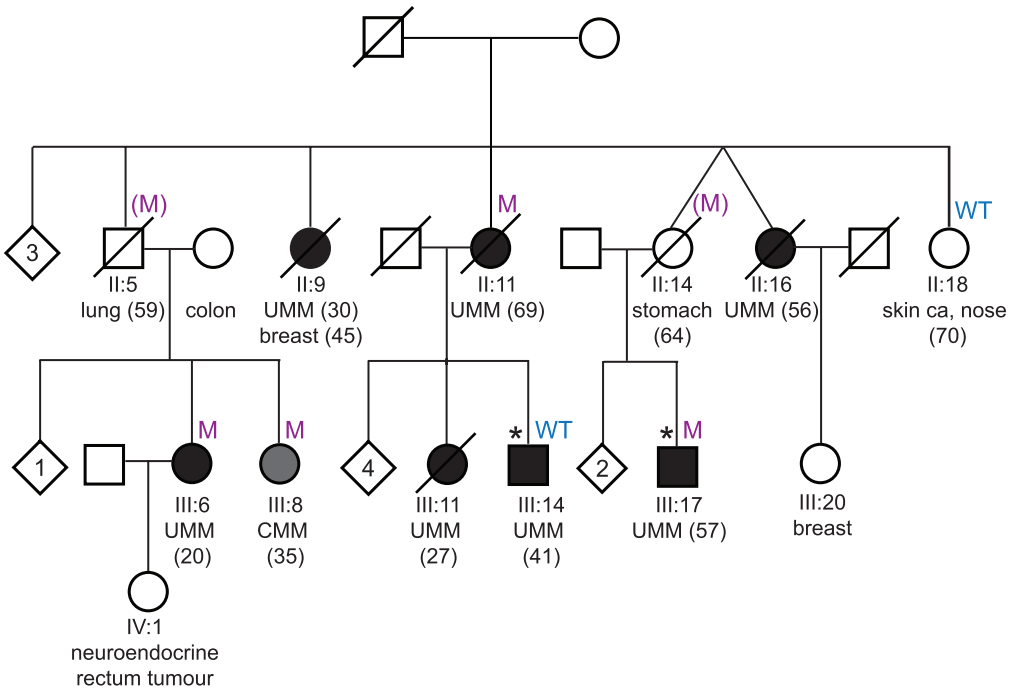


Figure 1. Co-segregation analysis of a *BAP1* splice mutation in a Danish family. In the pedigree individuals that have uveal melanoma (UMM) are represented by black circles (female) or boxes (male) and individuals with cutaneous melanoma (CMM) are indicated by grey circles or boxes. The age of diagnosis of each melanoma is indicated in brackets. A line through a symbol indicates that the person is now deceased. If a person carries the *BAP1* splice mutation it is indicated by an 'M' and if they are wild type for this variant it is indicated with a 'WT'. Where the mutation status is indicated in brackets, the person is a presumed obligate carrier, '(M)'. Other cancer types are also indicated in the pedigree with the age of diagnosis in brackets. Asterisks indicate the two individuals that were exome sequenced. Unaffected siblings are represented by diamonds, with the number in the centre indicating the number of people.
doi:10.1371/journal.pone.0072144.g001

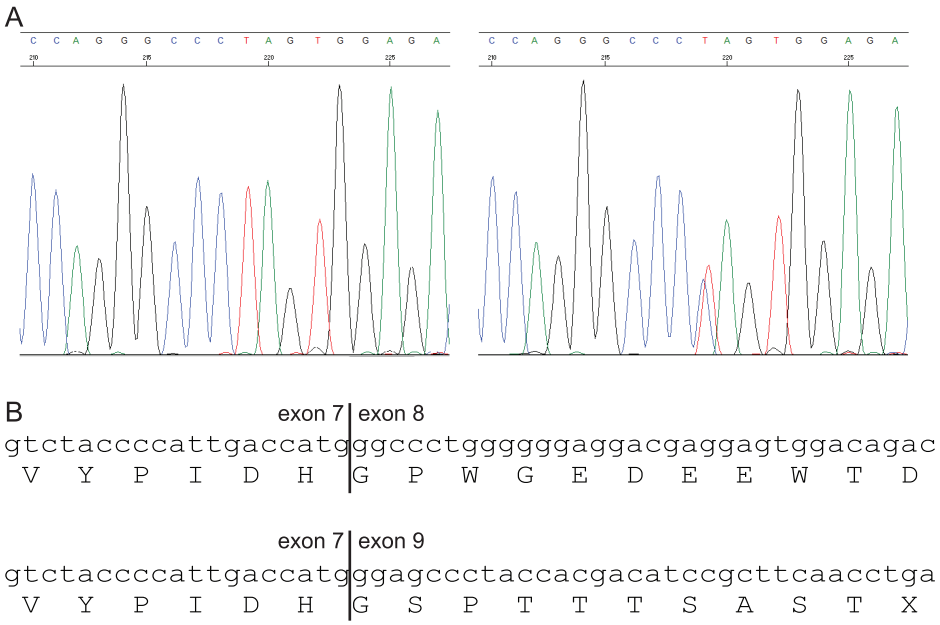


Figure 2. Sanger sequencing trace and amino acid alignment showing the truncated BAP1 protein. (A) The left panel shows the wild type chromatogram while the right panel shows that of the *BAP1* splice mutation. (B) Wild type nucleotide and amino acid sequences are shown in the upper panel while the lower panel shows part of the truncated BAP1 protein resulting from the loss of exon 8.
doi:10.1371/journal.pone.0072144.g002

seen in dbSNP (build 132) or reported by the 1000 Genomes Project (February 2012) were filtered out, leaving only novel variants for consideration. Secondly, synonymous changes were removed, leaving only the protein-altering variants. A cut off of 40 was then applied to the quality score returned for each variant in order to limit the false positive rate amongst the called variants. A 20% cut off was also applied to the number of alternate reads, that is, if a variant did not have a high enough percentage of reads for the alternate allele then it was removed from consideration. What remained after this filtering process were 140 novel variants for individual III:14 and 145 novel variants for individual III:17. Exome data from these individuals may be made available for research purposes upon request to the authors.

Exome-sequencing uncovered a *BAP1* splice mutation, which results in the premature truncation of the protein, in one (III:17) of the two individuals and was seen as a likely candidate responsible for UMM susceptibility due to the supporting literature. The result was verified in the individual using Sanger sequencing (refer to Figure 2). Following this, co-segregation analysis was carried out to determine how well the variant segregated with disease in the family and whether it was present in individuals with different cancer types. Every available blood DNA sample from this family was screened for the variant. For individual II:11 DNA was obtained from tumor tissue. Additionally, a new blood sample was obtained from individual III:14 since the initial sample did not show presence of a germline *BAP1* mutation. The second blood sample was screened for the specific mutation in *BAP1* using Sanger sequencing.

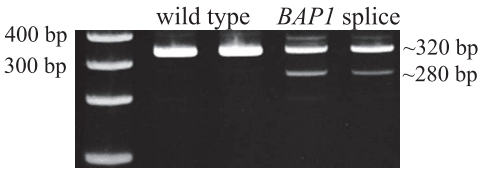


Figure 3. RT-PCR showing the aberrantly spliced *BAP1* transcript. From the left, the first lane shows a size marker, the next two lanes show wild type *BAP1* RT-PCR products, and the last two lanes show the aberrantly spliced product resulting from the c.581-2A>G mutation.
doi:10.1371/journal.pone.0072144.g003

RT-PCR

RT-PCR was performed on cDNA from one carrier (III:6) to verify that the *BAP1* mutation was indeed a splicing variant. RNA was converted to cDNA using Superscript II (Life technologies) and primers flanking the exons affected by the putative splice site were designed. Subsequently, PCR products were run on an acrylamide gel (refer to Figure 3).

Case-Control Genotyping

To more fully understand the role that this variant could play in melanoma risk an Australian case-control sample was genotyped for this novel variant. A Sequenom iPLEX was run on 1655 affected probands from families that had either CMM, UMM or a combination of both, and 1596 controls. The sample set included a wide cross-section of cases with varied histories of melanoma. Some of the probands are from dense melanoma families while others are sporadic cases. This cohort also includes cases with a wide spectrum of ages of onset, which range from childhood disease to late onset melanoma.

Results

Whole-exome-sequencing identified a *BAP1* splice mutation (c.581-2A>G) in an individual with UMM from a Danish family predisposed to developing UMM as well as a host of other cancers. Co-segregation analysis showed that three out of the four UMM cases with DNA available for research purposes were carriers of this mutation, which causes premature truncation of *BAP1*. Additionally, there were four other individuals with different cancer types that also carried the mutation. These cancers were

Table 1. Clinical phenotypes of current and published UMM cases or their families with *BAP1* mutations.

	Scandinavian families			non-Scandinavian families								
	Current Family	Wadt et al [9]	Höiom et al [17]	Testa et al [2]	Wiesner et al [4,5]	Njauw et al [18]	Abdel-Rahman et al [3]	Popova et al [10]	Ribeiro et al [8]	Scandinavian	non-Scandinavian	total
number UMM cases in family	7	3	3	4	2	8	3	14	1	13	32	45
cases with genotype data	4	3	3	3	2	8	2	8	1	10	24	34
cases with *BAP1* mutation	3	3	3	3	2	8	2	8	1	9	24	33
cases with metastatic disease	3[a]	2[b]	3	0[c]	na	4[d]	na	na	0	8	4	12
ages of onset of carriers	20, 57, 69	18, 46, 62	16, 39, 44	55, 59, 63	44, 72	37, 51, 53, 55, 57, 58, 62,*	na	35, 44, 44, 48, 52, 53, 53, 57	56	^	~	§
mean age of onset of carriers	49	42	33	59	58	54	na	49	56	42	53	50

na is not available; a has been observed for 41 years without metastatic disease; b has been observed for 16 years without metastatic disease; c has been observed for 1, 4 and 8 years without metastatic disease; d has no information regarding the 4 other UMM cases;
*refers to UMM case with unknown age of onset;
^refers to values from columns 1–3;
~refers to values from columns 4–9;
§refers to values from columns 1–9.
doi:10.1371/journal.pone.0072144.t001

CMM, lung cancer and stomach cancer. To confirm that individual III:14 was wildtype for the *BAP1* mutation a newly obtained blood sample was screened. Indeed, the analysis confirmed the absence of *BAP1* mutation suggesting that this case is a phenocopy. Loss of heterozygosity (LOH) of the *BAP1* mutation was also observed in DNA from tumor tissue from II:11.

Consistent with data from the 1000 Genomes Project, genotyping results showed that this variant was only present in the current family, that is only 1/1655 melanoma probands carried the variant and 0/1596 controls. Data from the NHLBI GO Exome Sequencing Project (ESP) also reflects this finding as this splicing variant is not found in any of the 6500 individuals sequenced from the USA [13]. Exome data from 200 healthy Danish controls was also examined for *BAP1* mutation and no protein altering variant was seen [14]. This suggests that the individuals described here most likely carry a mutation 'private' to this family.

Discussion

There have been several cancer types associated with *BAP1* germline mutations but the full spectrum of tumor susceptibility is still to be ascertained. We report here a Danish family with UMM and pigmented skin lesions, as well as lung, neuroendocrine, stomach, and breast cancers. Whole-exome sequencing of a UMM case identified a *BAP1* splice mutation (c.581-2A>G), which leads to premature truncation of *BAP1*. This frameshift aberration is consistent with published germline mutations seen in other families predisposed to UMM. In keeping with other Scandinavian germline *BAP1* carriers, we also observe a younger age at diagnosis of UMM (Table 1). In the present and the previously published Scandinavian families, the mean age at diagnosis of UMM was 42 years for the nine affected individuals, compared to 53 years for UMM carriers in non-Scandinavian families (Table 1 and references therein). This compares to a median age at diagnosis for UMM of between 58-62 years in population-based sporadic cases in Scandinavia [15] and the United States [16]. The role of *BAP1* as a tumor suppressor gene was further supported by LOH of the wild type allele in UMM tumor tissue of II:11. The wild type individual that is affected with UMM shows that in dense families it is possible to observe phenocopies even with a tumor type as rare as UMM. The exome sequence data of the person with UMM and wild type *BAP1* did not reveal mutations in other known cancer predisposition genes.

In summary, this finding expands on the growing profile of *BAP1* as an important uveal and cutaneous melanoma suppressor gene. The family described here also had a variety of other cancer types, some of which have been implicated with *BAP1* mutation (lung) and some that have not (stomach, neuroendocrine). To assess the possible involvement of *BAP1* in predisposition to these non-melanoma cancers and to more fully understand the spectrum of disease associated with such mutations, large population-based studies are required.

Supporting Information

Methods S1 Addition information on whole-exome sequencing, Sanger sequencing and Sequenom iPLEX methods.
(DOCX)

Acknowledgments

We thank the family members for their participation in this study.

Author Contributions

Conceived and designed the experiments: NKH GJ KMB JMT. Performed the experiments: LGA GJ. Analyzed the data: LA GJ NKH. Contributed reagents/materials/analysis tools: KW NKH GJ JMT ÅB. Wrote the paper: LGA. Performed clinical work: KW AB DC AMG ÅB. Edited and reviewed the manuscript: LGA KW AB DC ÅB JMT KMB AMG GJ NKH.

References

1. Harbour JW, Onken MD, Roberson ED, Duan S, Cao L, et al. (2010) Frequent mutation of BAP1 in metastasizing uveal melanomas. Science 330: 1410–1413.
2. Testa JR, Cheung M, Pei J, Below JE, Tan Y, et al. (2011) Germline BAP1 mutations predispose to malignant mesothelioma. Nat Genet 43: 1022–1025.
3. Abdel-Rahman MH, Pilarski R, Cebulla CM, Massengill JB, Christopher BN, et al. (2011) Germline BAP1 mutation predisposes to uveal melanoma, lung adenocarcinoma, meningioma, and other cancers. J Med Genet 48: 856–859.
4. Wiesner T, Obenauf AC, Murali R, Fried I, Griewank KG, et al. (2011) Germline mutations in BAP1 predispose to melanocytic tumors. Nat Genet 43: 1018–1021.
5. Wiesner T, Fried I, Ulz P, Stacher E, Popper H, et al. (2012) Toward an improved definition of the tumor spectrum associated with BAP1 germline mutations. J Clin Oncol 30: e337–340.
6. Carbone M, Korb Ferris L, Baumann F, Napolitano A, Lum CA, et al. (2012) BAP1 cancer syndrome: malignant mesothelioma, uveal and cutaneous melanoma, and MBAITs. J Transl Med 10: 179.
7. Busam KJ, Wanna M, Wiesner T (2013) Multiple epithelioid Spitz nevi or tumors with loss of BAP1 expression: a clue to a hereditary tumor syndrome. JAMA Dermatol 149: 335–339.
8. Ribeiro C, Campelos S, Moura CS, Machado JC, Justino A, et al. (2013) Well-differentiated papillary mesothelioma: clustering in a Portuguese family with a germline BAP1 mutation. Ann Oncol.
9. Wadt K, Choi J, Chung JY, Kiilgaard J, Heegaard S, et al. (2012) A cryptic BAP1 splice mutation in a family with uveal and cutaneous melanoma, and paraganglioma. Pigment Cell Melanoma Res.
10. Popova T, Hebert L, Jacquemin V, Gad S, Caux-Moncoutier V, et al. (2013) Germline BAP1 Mutations Predispose to Renal Cell Carcinomas. Am J Hum Genet.
11. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754–1760.
12. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079.
13. Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA (URL: http://evs.gs.washington.edu/EVS/) Accessed 14 June 2013.
14. Li Y, Vinckenbosch N, Tian G, Huerta-Sanchez E, Jiang T, et al. (2010) Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. Nat Genet 42: 969–972.
15. Isager P, Osterlind A, Engholm G, Heegaard S, Lindegaard J, et al. (2005) Uveal and conjunctival malignant melanoma in Denmark, 1943–97: incidence and validation study. Ophthalmic Epidemiol 12: 223–232.
16. Singh AD, Turell ME, Topham AK (2011) Uveal melanoma: trends in incidence, treatment, and survival. Ophthalmology 118: 1881–1885.
17. Hoiom V, Edsgard D, Helgadottir H, Eriksson H, All-Ericsson C, et al. (2013) Hereditary uveal melanoma: a report of a germline mutation in BAP1. Genes Chromosomes Cancer 52: 378–384.
18. Njauw CN, Kim I, Piris A, Gabree M, Taylor M, et al. (2012) Germline BAP1 Inactivation Is Preferentially Associated with Metastatic Ocular Melanoma and Cutaneous-Ocular Melanoma Families. PLoS One 7: e35295.

Supplementary Information

Methods

Whole-exome Sequencing

Whole-exome sequencing was carried out on two members of a Danish family predisposed to uveal melanoma. Both individuals were affected with UMM and had a family history of disease that is not limited to UMM but also includes several other cancer types. The sequencing was performed on the Illumina HiSeq 2000 using the Illumina TruSeq Exome Enrichment Kit. The sequencing was outsourced to Axeq Technologies in Seoul, South Korea. The sequencing was paired-end and had a median coverage of 100x. Sequence data were mapped to the Human Genome build 19 (hg19). Any variant that was found in dbSNP or the 1000 Genomes project was filtered out leaving only the novel mutations. By applying cut offs to the quality score and the rate of mutant calls for each variant, we were left with novel 140 variants in III:14 and 145 novel variants in III:17 for follow up.

Sanger Sequencing

Sanger sequencing was used to validate the *BAP1* splice mutation (c.581-2A>G) identified through whole-exome sequencing. A PCR reaction was run using 0.5 units of Qiagen HotStarTaq, 2ul of 10x PCR Buffer, 4ul of 5x Q-Solution, 1.6ul of 2.5mM dNTP mix, 1mM each of the forward and reverse primers (CCTGGCTCAACTGCTCTTCT and GCCCAGGCAGGAAATAAGAC respectively) and 15ng of genomic DNA, made up to a final reaction volume of 20ul.

The PCR was run on a Biorad Thermocycler according to Qiagen PCR protocol. It is initiated with a 5 minute activation step at a temperature of 95^{o}C, followed by a 3-step cycling process which is repeated 35 times. The cycle starts with a denaturation step at a temperature of 94^{o}C for 45 seconds followed by an annealing step at 65^{o}C for 45 seconds and then an extension step at 72^{o}C for 45 seconds. This 3-step process is repeated with a 65^{o}C annealing temperature for two cycles and then steps down in 2^{o}C increments until it reaches 57^{o}C where the temperature remains until the full 35 cycles are complete. A final 10 minute extension at 72^{o}C finalises the PCR run.

The crude PCR product was sent to Functional Biosciences, Wisconsin, USA, to be cleaned up and run on a sequencer. The AB1 files were analysed using an online chromatogram comparison tool, Multiple SeqDoC (http://research.imb.uq.edu.au/seqdoc/multi.html) which is curated by the University of Queensland. The traces were also analysed using chromatogram viewer, Chromas (version 1.45).

Sequenom iPLEX

The *BAP1* c.581-2A>G splice variant was multiplexed using the Sequenom iPLEX gold system which allows up to 36 variants to be genotyped concurrently using 10 ng of genomic DNA. The iPLEX Gold PCR amplification reactions were carried out in 384-well plates according to supplier protocol and then transferred onto SpectroCHIP arrays using a MassARRAY Nanodispenser. The chips were then transferred to a Sequenom mass spectrometer where the data were generated. The results were analysed using Typer Analyzer software 4.0.

Chapter 8

General discussion and future direction

In this thesis, a next-generation sequencing (NGS) approach has been used to identify novel pathogenic mutations in high-density melanoma families from Australia, the UK, Denmark and Sweden. Using a combination of whole-genome and exome-sequencing we identified three genes in the telomere-associated complex, shelterin, which confer a highly elevated risk of melanoma when mutated. NGS technology has revolutionised our ability to identify cancer predisposition genes and has been particularly valuable in instances where mutations are rare, occurring in only a small fraction of families. By looking at mutations that segregate within specific families and then extending this to a large collection of families, we were able to pinpoint significant genes that contribute to CMM predisposition.

Over the course of this project, a novel mutation occurring in the promoter region of the telomerase reverse transcriptase gene (*TERT*) was found to be associated with familial CMM by Horn and colleagues [130]. The mutation, located -57 bp from the ATG translation start site, was found to segregate with disease in a 14-case family from Germany. Functionally, it created a new binding motif for Ets transcription factors and ternary complex factors (TCF), which resulted in a two-fold increase in transcription of *TERT* (Figure 7). Similar to what was observed in families with shelterin mutations, some carriers of the *TERT* promoter mutation developed multiple cancer types. One individual developed ovarian cancer (at 27 years) and CMM (at 30 years), while another family member developed CMM (at 20 years) followed by ovarian, renal cell, bladder, breast and bronchial carcinoma. Following the identification of the *TERT* mutation in this family, Horn and colleagues analysed tumours derived from sporadic CMM cases and found that 85% of metastatic tumours and 33% of primary tumours carried somatic mutations in the promoter region of *TERT*, with the majority of these mutations occurring at two hotspots (-124 bp and -146 bp; Figure 7). This finding represents the first documentation of a novel high-risk melanoma predisposition gene since *CDK4* in 1996 [6], and despite the huge sequencing efforts of several international research groups, it has taken over 15 years for this breakthrough to occur. This mutation in *TERT* also points to a new pathway responsible for familial melanoma.

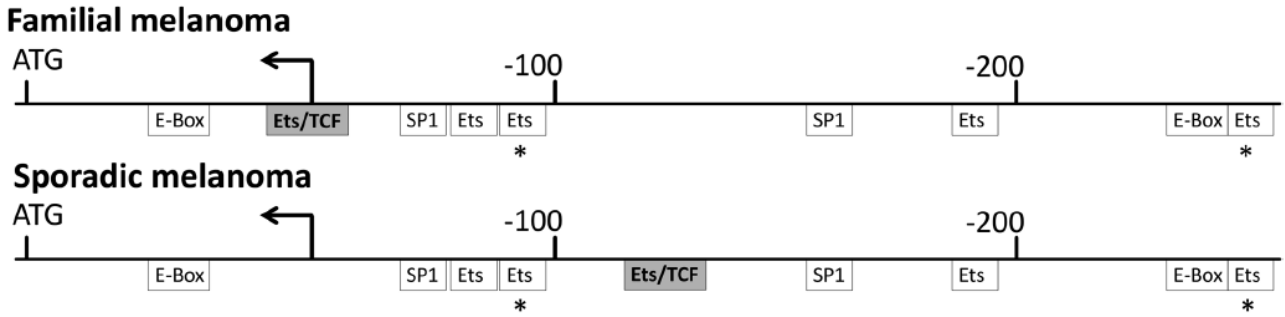


Figure 7: ***TERT*** **promoter mutations in melanoma.** Mutations creating Ets/TCF binding motifs occurred in affected family members at -57 bp (next to the transcription start site). This also occurred in sporadic cases at -124 to -149 bp. Asterisks indicate sites at which Ets2 binding was reported. Adapted from Horn et al., 2013.

This thesis shows mutations in genes *POT1, ACD* and *TERF2IP* confer a high risk of melanoma in carriers. These genes encode members of shelterin, the telomere-specific protein complex that mediates the interaction between telomeres and telomerase (which includes *TERT*). Shelterin is made up of 6 components, encoded by the genes *ACD, POT1, TERF1, TERF2, TERF2IP* and *TINF2*. Collectively, they are necessary for all telomere functions, which include the protection of telomeres from degradation or aberrant recombination, as well as from being inappropriately processed by the DNA-repair pathway [135].

Telomeres are characterised by (TTAGGG)n nucleotide repeats, typically between 50 bp and 300 bp in length at each 3' single-stranded overhang [136]. The length of telomeric DNA is variable between chromosomes and also between individuals, with the range of total telomere length in humans being 0.5 kb to 15 kb depending on age and tissue type [137]. Telomerase is the reverse transcriptase that elongates telomeres by adding telomeric repeats [138]. Shelterin binds to telomeres and mediates telomere length by restructuring the terminal domain through the generation of t-loops in order to protect chromosomes from unwarranted cleavage (Figure 8) [135]. These t-loops are also able to physically block telomerase from elongating the telomeres to facilitate continued cell proliferation [139, 140]. Without this t-loop mediation it is feasible that cells will not enter a natural state of senescence and therefore avoid normal regulation through the cell cycle.

POT1 is the most conserved of the shelterin complex genes and in conjunction with *ACD*, enables the formation of t-loops [141]. POT1 specifically binds to telomeric single-stranded DNA (ssDNA) through two OB (oligonucleotide/oligosaccharide-binding) domains in order to mediate the access of telomerase [142]. It has been shown that the ACD-POT1 sub-complex also has a higher binding affinity to ssDNA than POT1 does on its own [143]. Literature shows that if the function of *POT1* is inhibited the length of telomeres is increased through telomerase activity, indicating that this sub-complex is able to regulate the elongation of chromosome ends [144-148]. In our study, we show that missense mutations in the OB-domains of POT1 result in an elongation of telomeres in carriers. This was measured in both exome data and through qPCR. In a separate study, Bataille and colleagues report that a large naevus count is positively correlated with longer telomere length in women of Caucasian descent by measuring the telomere length of white blood cells. They have also shown that individuals (N=1,897) with naevi larger than 5 mm in diameter had an average telomere length that was 150 bp longer than their age-matched controls [149]. High naevus count and naevus diameter has been shown to be a strong predictive marker of melanoma risk with members of

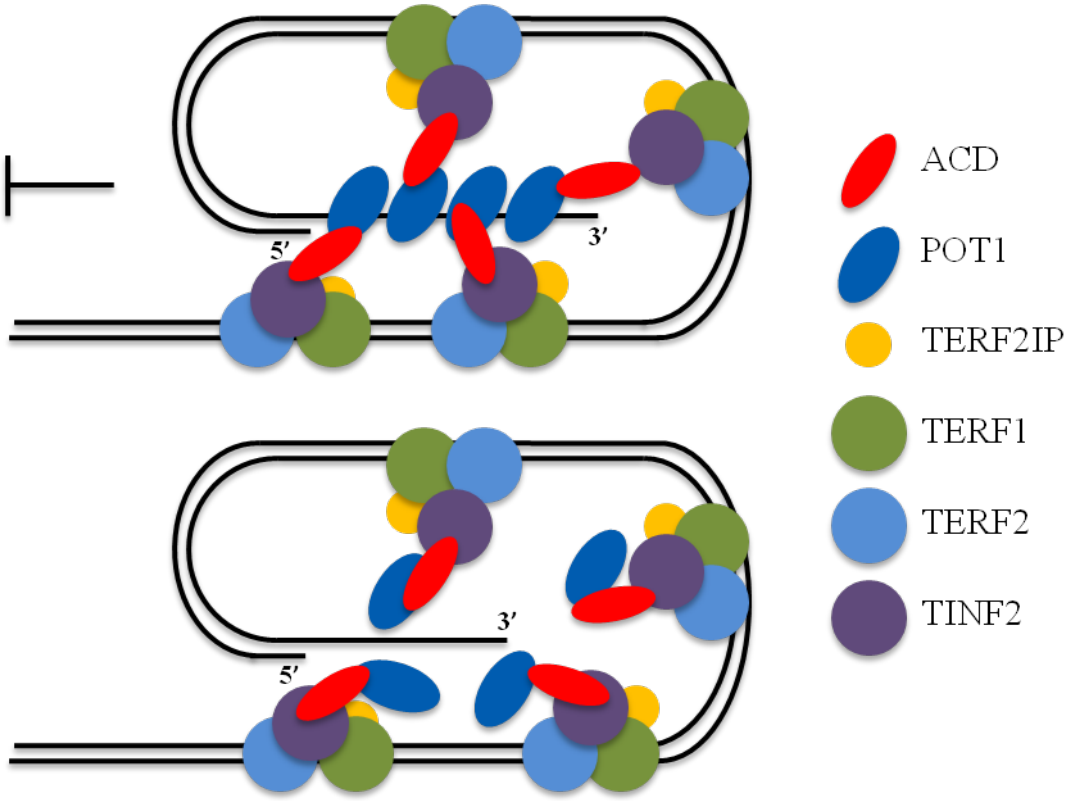


Figure 8: The shelterin complex. The top panel depicts the formation of the t-loop, blocking the interaction with telomerase. The lower frame shows the shelterin complex in the open position, allowing telomerase access to telomeres.

large melanoma families often having large naevus counts [150, 151]. The research supports the association between telomere length and naevus development. A future avenue of research will be to investigate whether aberrant function of the shelterin complex gives rise to irregular naevi and a higher risk of CMM development.

The main role of TERF1 and TERF2 (telomeric repeat-binding factor 1 and telomeric repeat-binding factor 2) is to bind telomeric double-stranded DNA (dsDNA) and mediate recruitment of different factors to telomeres [152-156]. TINF2 (TERF1-interacting nuclear factor 2) interacts with the TERF1-TERF2 dsDNA binding complex. TINF2 is at the centre of the complex, tethering the ACD-POT1 partnership to the TERF1-TERF2 partnership. It is crucial for the assembly of the shelterin complex. TERF2IP (TERF2-interacting telomeric protein 1) binds to a central region of TERF2 via its C-terminus and is necessary for the repression of homology-directed repair of double strand chromosomal break at the telomere [157].

Unlike the mutations we have reported in *POT1*, mutations in *ACD* and *TERF2IP* do not appear to alter the length of telomeres. Quantitative PCR data showed no observable difference in the length of telomeres between mutation carriers and their wild type relatives. The data suggests that these mutations are most likely disrupting the complex and destabilising chromosomes, in carriers rather than specifically altering their telomere length. More comprehensive functional experiments will be needed to deduce the exact mechanism through which mutations in *ACD* and *TERF2IP* confer elevated risk of CMM development in the families we describe.

The initial association between mutations in the shelterin complex and cancer were first published in 2013 by Ramsey et al [129]. Exome sequencing data showed that a recurrent somatic mutation in *POT1* occurred in approximately 3.5% of chronic lymphocytic leukaemia (CLL) patients, with these mutations occurring at a frequency of 9% when considering only the more aggressive disease subtype. The authors showed that CLL cells with mutated *POT1* acquired numerous chromosomal abnormalities in comparison to *POT1* wild type controls. Interestingly, the majority of the mutations detected in these cases lie within the OB-domains of *POT1*, similarly to what we report in the germline (Chapter 4). This indicates that *POT1* plays an important role in cancer biology, specifically in CLL and CMM. The functional consequences of the *POT1* mutations in CLL lends to the hypothesis that mutations may also lead to destabilisation of chromosomes in CMM and that this may be an important mechanism driving tumour development in families. Further functional analysis is required to test this theory in CMM development.

Currently, the only other disease to report germline mutations in the shelterin complex is dyskeratosis congenita (DKC). It is a rare congenital disorder with phenotypic characteristics

resembling premature aging. Its clinical features include abnormal skin pigmentation with cases presenting with hyper or hypo skin pigmentation, nail dystrophy, oral leukoplakia and bone marrow failure [158]. Cases also have an elevated risk of malignancy with an over representation of basal cell carcinoma, Hodgkin lymphoma, gastrointestinal cancers and bronchial cancers. One type of DKC is caused by germline mutations in exon 6 of *TINF2* [158-162]. It is of note that the mutation we report in Chapter 5, Supplementary methods, occurs outside of this region (in exon 1). In the first report linking *TINF2* mutations to DKC, the authors described shortened telomeres leading to destabilisation of telomeres in carriers [160]. This is the opposite of what we have seen, as CMM cases with *POT1* mutations are shown to have increased telomere length. Overall, this indicates that both very short and very long telomeres are pathogenic and can lead to different malignancies; clearly the role of shelterin is multifaceted and we are yet to fully understand its biological complexities. It would be valuable to determine the functional mechanism through which *ACD* and *TERF2IP* mutations are leading to CMM development. Obvious experiments (protein modelling; qPCR length assay; protein-protein interaction through octet; exome data analysis) have not been able to pinpoint the biological mechanism that causes these variants to be pathogenic in our families.

When looking collectively at the families with shelterin mutations in conjunction with the family carrying the *TERT* promoter mutation [130], we see an enrichment of cancers in mutation carriers. This may be an indication that this pathway is important in general tumour suppression and that mutations in these genes may lead to a cancer phenotype that extends beyond CMM. It is possible that, like *BAP1*, mutations in these genes cause a predisposition to a host of malignancies. This would be a valuable follow-up to the body of work presented in Chapters 4 and 5. Looking at CMM families that also present with CLL may be a good starting point for association to a general cancer phenotype.

We have identified shelterin mutations in cases that lack mutation in any of the known melanoma predisposition genes (*CDKN2A* and *CDK4*). Collectively, with the identification of the *TERT* promoter mutation [130], the findings of this thesis have uncovered a novel pathway that leads to CMM development in the familial setting. It would be valuable to more thoroughly interrogate the role of the remaining three shelterin members in melanoma predisposition as well as other genes in this pathway as it is becoming clear that telomere dysregulation is a major contributor to melanoma development in some high-density families.

An alternative method of gene discovery presented in this thesis was the candidate gene approach, whereby a literature-based search was used to identify genes of interest in melanoma susceptibility.

The protein-coding regions of the chosen genes were Sanger sequenced to identify novel variants in probands from melanoma families for mutations in *PALB2, SOX10* and *MITF* (Chapter 3). This method did not prove overly successful in identifying causative mutations amongst high-risk CMM families. Though the literature is suggestive of a role of *SOX10* in melanomagenesis, no evidence of it was found in the familial setting. There was also a lack of evidence for further *MITF* variants in melanoma predisposition. This consolidates the notion that a more high-throughput approach to sequencing entire exonic/genomic regions in cases is a more time-effective method for gene discovery in cancer, particularly seeing as the costs associated with NGS have dropped dramatically over the past two years.

Over the course of my PhD, BRCA-1 associated protein-1 (*BAP1)* was identified as a melanoma risk gene in the familial setting. The initial association was made with UMM predisposition but since its discovery the disease profile has extended far beyond this phenotype [103]. In this thesis, I sought to determine the contribution of *BAP1* to both CMM and UMM population-based samples from Australia (Chapter 2 and Chapter 7). This was achieved through a combination of high-throughput, targeted pull-down of the selected genes using the Ion Torrent PGM, and also SNP analysis using Sequenom mass spectrometry. We also estimated the contribution of known melanoma risk genes *CDKN2A*, *CDK4* and *BAP1* to melanoma in a Queensland population-based sample (N=1,109). We showed that *CDKN2A* accounts for around 1.31% of Queensland melanoma cases, *BAP1* accounts for around 0.63%, and *CDK4* mutations are extremely rare, not being found in the Queensland-based sample. This is the best estimate of the contribution of *CDKN2A* and *CDK4* to CMM in the Queensland population reported to date. It is also the only reported estimate of *BAP1* contribution to CMM and UMM in any population.

In the familial setting, *BAP1* has become the first (and currently the only) predisposition gene for UMM. We have shown that a mutation of this gene is present in a high-density UMM family from Denmark. Interestingly, this family presented with a host of other cancer types which support *BAP1* as a general tumour suppressor that predisposes to a range of cancers that includes both UMM and CMM. In terms of risk genes that are associated with UMM, there is still a great deal that is unknown. In comparison to CMM, the discovery rate of UMM predisposition genes has been slow, with only a single gene attributed to predisposition in 2011. This may be in part due to the smaller availability of samples. Given that UMM affects only 5.6 people per million in the USA, the collection of samples for this type of study is much more difficult. In the data that we have examined to date, it is likely that the mutations driving UMM in families are going to be rare; therefore we will need to sequence larger numbers of families to see the recurrence of genes being mutated in families. It is feasible that like CMM, mutations in particular pathways will predispose

to UMM and therefore we may need to interrogate families of genes to identify the causative pathways rather than stand-alone genes.

In Australian clinics, it is not yet routine for sporadic UMM cases to be screened for mutations in BAP1. As families with UMM are rare, there may be only a few individuals that present with a family history. A better guide for determining whether or not a person may have a causative germline mutation in *BAP1* would be to look at the extended cancer history of the family. Since there have been such a broad range of cancer implicated with germline *BAP1* mutations (see Chapter 1 and Chapter 7) it is likely that these families will present with a more complex cancer history than just UMM. In the future, clinicians could benefit from comprehensive guidelines for identifying a *BAP1* mutation phenotype so these families can be better identified. This will lead to earlier detection and better management of cancers arising by germline mutations in this gene. If we can identify individuals carrying germline BAP1 mutations in the clinical setting we can improve the diagnosis and prognosis of cancer in their family members.

Overall, this project has made a major contribution to the field of familial melanoma genetics. Firstly, we have analysed the population-based frequency of melanoma risk genes *CDKN2A*, *CDK4* and *BAP1* in a CMM sample from Queensland, Australia. We have also determined the fraction of UMM cases that can be attributed to *BAP1* mutation in a population-based sample from New South Wales, Australia. Following on from this, we described a Danish family with a germline *BAP1* mutation responsible for the development of UMM and a host of other cancers. Finally, we have described novel truncating mutations in three members of the shelterin complex in CMM families, which collectively accounts for 9% of families lacking mutations in previously known risk genes. In combination with a reported finding of a mutation in the *TERT* promoter, the research presented in this thesis has led to the identification of a novel pathway related to melanoma predisposition and thus a significant advance in the field of familial melanoma genetics.

The findings of this thesis could be translated to positive outcome in the clinical setting. Currently, genetic counselling clinics in Australia only screen CMM families for mutations in CDKN2A and CDK4. In the future, the hope is that individuals presenting with a strong family history of CMM will be able to be tested for mutations in POT1, ACD and TERF2IP as well. By identifying those individuals with the greatest risk of CMM development, dermatologists can monitor them and their family members more rigorously for changes to their skin. Furthermore these families could benefit from education that allows them to better monitor changes to their own skin. As melanoma is quite curable when detected in the early stages, the key for the future will be identifying the families that

carry these high-risk germline mutations so they can be thoroughly monitored to improve CMM prognosis.

Next-generation sequencing techniques have been an invaluable tool in melanoma predisposition gene discovery. Understanding the genetic mutations behind melanoma development provides an insight into the disease mechanism and, in the future, will aid early detection and inform more effective prevention strategies. Ultimately, the identification of these genes will lead to more therapeutic options for patients with melanoma. This is of particular importance to a population like Queensland's where the burden of melanoma is so great.

List of References

1. AIHW, *Cancer in Australia: an overview 2012.* . cancer series no. 74, Canberra 2012(cat no. CAN70).

2. Dickson, P.V. and J.E. Gershenwald, *Staging and prognosis of cutaneous melanoma.* Surg Oncol Clin N Am, 2011. **20**(1): p. 1-17.

3. MacKie, R.M., A. Hauschild, and A.M. Eggermont, *Epidemiology of invasive cutaneous melanoma.* Ann Oncol, 2009. **20 Suppl 6**: p. vi1-7.

4. AIHW, *Australian Cancer Incidence and Mortality*. 2011: Canberra.

5. AIHW, *Cancer in adolescents and young adults in Australia*, in *Cancer series no. 62. Cat. no. CAN 59*. 2011: Canberra.

6. Zuo, L., et al., *Germline mutations in the p16INK4a binding domain of CDK4 in familial melanoma.* Nat Genet, 1996. **12**(1): p. 97-9.

7. Soufir, N., et al., *Prevalence of p16 and CDK4 germline mutations in 48 melanoma-prone families in France. The French Familial Melanoma Study Group.* Hum Mol Genet, 1998. **7**(2): p. 209-16.

8. Hussussian, C.J., et al., *Germline p16 mutations in familial melanoma.* Nat Genet, 1994. **8**(1): p. 15-21.

9. MacGeoch, C., et al., *Genetic heterogeneity in familial malignant melanoma.* Hum Mol Genet, 1994. **3**(12): p. 2195-200.

10. Gruis, N.A., et al., *Homozygotes for CDKN2 (p16) germline mutation in Dutch familial melanoma kindreds.* Nat Genet, 1995. **10**(3): p. 351-3.

11. Walker, G.J., et al., *Mutations of the CDKN2/p16INK4 gene in Australian melanoma kindreds.* Hum Mol Genet, 1995. **4**(10): p. 1845-52.

12. Clark, W.H., Jr., et al., *The histogenesis and biologic behavior of primary human malignant melanomas of the skin.* Cancer Res, 1969. **29**(3): p. 705-27.

13. Clark, W.H., Jr., et al., *The developmental biology of primary human malignant melanomas.* Semin Oncol, 1975. **2**(2): p. 83-103.

14. Chudnovsky, Y., P.A. Khavari, and A.E. Adams, *Melanoma genetics and the development of rational therapeutics.* J Clin Invest, 2005. **115**(4): p. 813-24.

15. Clark, W.H., Jr. and M.C. Mihm, Jr., *Lentigo maligna and lentigo-maligna melanoma.* Am J Pathol, 1969. **55**(1): p. 39-67.

16. Kallini, J.R., S.K. Jain, and A. Khachemoune, *Lentigo maligna: review of salient characteristics and management.* Am J Clin Dermatol, 2013. **14**(6): p. 473-80.

17. Singh, A.D., M.E. Turell, and A.K. Topham, *Uveal melanoma: trends in incidence, treatment, and survival.* Ophthalmology, 2011. **118**(9): p. 1881-5.

18. Cust, A.E., et al., *Population-based, case-control-family design to investigate genetic and environmental influences on melanoma risk: Australian Melanoma Family Study.* Am J Epidemiol, 2009. **170**(12): p. 1541-54.

19. Goldstein, A.M., et al., *High-risk melanoma susceptibility genes and pancreatic cancer, neural system tumors, and uveal melanoma across GenoMEL.* Cancer Res, 2006. **66**(20): p. 9818-28.

20. Kraft, P., et al., *Beyond odds ratios--communicating disease risk based on genetic profiles.* Nat Rev Genet, 2009. **10**(4): p. 264-9.

21. Yokoyama, S., et al., *A novel recurrent mutation in MITF predisposes to familial and sporadic melanoma.* Nature, 2011. **480**(7375): p. 99-103.

22. Woodman, S.E., *BAP1tism of a Tumor Suppressor.* Clin Cancer Res, 2012. **18**(2): p. 323-5.

23. Kamb, A., et al., *Analysis of the p16 gene (CDKN2) as a candidate for the chromosome 9p melanoma susceptibility locus.* Nat Genet, 1994. **8**(1): p. 23-6.

24. Bahuau, M., et al., *Germ-line deletion involving the INK4 locus in familial proneness to melanoma and nervous system tumors.* Cancer Res, 1998. **58**(11): p. 2298-303.

25. Nobori, T., et al., *Deletions of the cyclin-dependent kinase-4 inhibitor gene in multiple human cancers.* Nature, 1994. **368**(6473): p. 753-6.

26. Palmieri, G., et al., *Main roads to melanoma.* J Transl Med, 2009. **7**: p. 86.

27. Campisi, J., *Cancer and ageing: rival demons?* Nat Rev Cancer, 2003. **3**(5): p. 339-49.

28. Randerson-Moor, J.A., et al., *A germline deletion of p14(ARF) but not CDKN2A in a melanoma-neural system tumour syndrome family.* Hum Mol Genet, 2001. **10**(1): p. 55-62.

29. Rizos, H., et al., *A melanoma-associated germline mutation in exon 1beta inactivates p14ARF.* Oncogene, 2001. **20**(39): p. 5543-7.

30. Harland, M., et al., *A mutation hotspot at the p14ARF splice site.* Oncogene, 2005. **24**(28): p. 4604-8.

31. Mistry, S.H., et al., *Prevalence of 9p21 deletions in UK melanoma families.* Genes Chromosomes Cancer, 2005. **44**(3): p. 292-300.

32. Harland, M., et al., *Mutation screening of the CDKN2A promoter in melanoma families.* Genes Chromosomes Cancer, 2000. **28**(1): p. 45-57.

33. Liu, L., et al., *Mutation of the CDKN2A 5' UTR creates an aberrant initiation codon and predisposes to melanoma.* Nat Genet, 1999. **21**(1): p. 128-32.

34. Harland, M., et al., *A deep intronic mutation in CDKN2A is associated with disease in a subset of melanoma pedigrees.* Hum Mol Genet, 2001. **10**(23): p. 2679-86.

35. de Snoo, F.A. and N.K. Hayward, *Cutaneous melanoma susceptibility and progression genes.* Cancer Lett, 2005. **230**(2): p. 153-86.

36. Begg, C.B., et al., *Lifetime risk of melanoma in CDKN2A mutation carriers in a population-based sample.* J Natl Cancer Inst, 2005. **97**(20): p. 1507-15.

37. Aitken, J., et al., *CDKN2A variants in a population-based sample of Queensland families with melanoma.* J Natl Cancer Inst, 1999. **91**(5): p. 446-52.

38. Gandini, S., et al., *Meta-analysis of risk factors for cutaneous melanoma: III. Family history, actinic damage and phenotypic factors.* Eur J Cancer, 2005. **41**(14): p. 2040-59.

39. Nelson, A.A. and H. Tsao, *Melanoma and genetics.* Clin Dermatol, 2009. **27**(1): p. 46-52.

40. Suzuki, I., et al., *Participation of the melanocortin-1 receptor in the UV control of pigmentation.* J Investig Dermatol Symp Proc, 1999. **4**(1): p. 29-34.

41. Sturm, R.A., R.D. Teasdale, and N.F. Box, *Human pigmentation genes: identification, structure and consequences of polymorphic variation.* Gene, 2001. **277**(1-2): p. 49-62.

42. Duffy, D.L., et al., *Interactive effects of MC1R and OCA2 on melanoma risk phenotypes.* Hum Mol Genet, 2004. **13**(4): p. 447-61.

43. Thody, A.J., et al., *Pheomelanin as well as eumelanin is present in human epidermis.* J Invest Dermatol, 1991. **97**(2): p. 340-4.

44. Robinson, S., et al., *Protection against UVR involves MC1R-mediated non-pigmentary and pigmentary mechanisms in vivo.* J Invest Dermatol, 2010. **130**(7): p. 1904-13.

45. Kraemer, K.H., et al., *The role of sunlight and DNA repair in melanoma and nonmelanoma skin cancer. The xeroderma pigmentosum paradigm.* Arch Dermatol, 1994. **130**(8): p. 1018-21.

46. Hayward, N.K., *Genetics of melanoma predisposition.* Oncogene, 2003. **22**(20): p. 3053-62.

47. Stahl, S., et al., *Genetics in melanoma.* Isr Med Assoc J, 2004. **6**(12): p. 774-7.

48. Valverde, P., et al., *Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans.* Nat Genet, 1995. **11**(3): p. 328-30.

49. Box, N.F., et al., *MC1R genotype modifies risk of melanoma in families segregating CDKN2A mutations.* Am J Hum Genet, 2001. **69**(4): p. 765-73.

50. Levy, C., M. Khaled, and D.E. Fisher, *MITF: master regulator of melanocyte development and melanoma oncogene.* Trends Mol Med, 2006. **12**(9): p. 406-14.

51. Tassabehji, M., V.E. Newton, and A.P. Read, *Waardenburg syndrome type 2 caused by mutations in the human microphthalmia (MITF) gene.* Nat Genet, 1994. **8**(3): p. 251-5.

52. Cronin, J.C., et al., *Frequent mutations in the MITF pathway in melanoma.* Pigment Cell Melanoma Res, 2009. **22**(4): p. 435-44.

53. Garraway, L.A., et al., *Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma.* Nature, 2005. **436**(7047): p. 117-22.

54. Williams, P.F., et al., *Melanocortin 1 receptor and risk of cutaneous melanoma: a meta-analysis and estimates of population burden.* Int J Cancer, 2011. **129**(7): p. 1730-40.

55. Bertolotto, C., et al., *A SUMOylation-defective MITF germline mutation predisposes to melanoma and renal carcinoma.* Nature, 2011. **480**(7375): p. 94-8.

56. Hoek, K.S., et al., *In vivo switching of human melanoma cells between proliferative and invasive states.* Cancer Res, 2008. **68**(3): p. 650-6.

57. Moynahan, M.E., A.J. Pierce, and M. Jasin, *BRCA2 is required for homology-directed repair of chromosomal breaks.* Mol Cell, 2001. **7**(2): p. 263-72.

58. Xia, F., et al., *Deficiency of human BRCA2 leads to impaired homologous recombination but maintains normal nonhomologous end joining.* Proc Natl Acad Sci U S A, 2001. **98**(15): p. 8644-9.

59. Sharan, S.K., et al., *Embryonic lethality and radiation hypersensitivity mediated by Rad51 in mice lacking Brca2.* Nature, 1997. **386**(6627): p. 804-10.

60. Wong, A.K., et al., *RAD51 interacts with the evolutionarily conserved BRC motifs in the human breast cancer susceptibility gene brca2.* J Biol Chem, 1997. **272**(51): p. 31941-4.

61. Welcsh, P.L. and M.C. King, *BRCA1 and BRCA2 and the genetics of breast and ovarian cancer.* Hum Mol Genet, 2001. **10**(7): p. 705-13.

62. Breast Cancer Linkage, C., *Cancer risks in BRCA2 mutation carriers.* J Natl Cancer Inst, 1999. **91**(15): p. 1310-6.

63. Moran, A., et al., *Risk of cancer other than breast or ovarian in individuals with BRCA1 and BRCA2 mutations.* Fam Cancer, 2012. **11**(2): p. 235-42.

64. Meyle, K.D. and P. Guldberg, *Genetic risk factors for melanoma.* Hum Genet, 2009. **126**(4): p. 499-510.

65. Brown, K.M., et al., *Common sequence variants on 20q11.22 confer melanoma susceptibility.* Nat Genet, 2008. **40**(7): p. 838-40.

66. Amos, C.I., et al., *Genome-wide association study identifies novel loci predisposing to cutaneous melanoma.* Hum Mol Genet, 2011. **20**(24): p. 5012-23.

67. Barrett, J.H., et al., *Genome-wide association study identifies three new melanoma susceptibility loci.* Nat Genet, 2011. **43**(11): p. 1108-13.

68. Fernandez, L.P., et al., *SLC45A2: a novel malignant melanoma-associated gene.* Hum Mutat, 2008. **29**(9): p. 1161-7.

69. Bishop, D.T., et al., *Genome-wide association study identifies three loci associated with melanoma risk.* Nat Genet, 2009. **41**(8): p. 920-5.

70. Falchi, M., et al., *Genome-wide association study identifies variants at 9p21 and 22q13 associated with development of cutaneous nevi.* Nat Genet, 2009. **41**(8): p. 915-9.

71. Macgregor, S., et al., *Genome-wide association study identifies a new melanoma susceptibility locus at 1q21.3.* Nat Genet, 2011. **43**(11): p. 1114-8.

72. Rafnar, T., et al., *Sequence variants at the TERT-CLPTM1L locus associate with many cancer types.* Nat Genet, 2009. **41**(2): p. 221-7.

73. Gudbjartsson, D.F., et al., *ASIP and TYR pigmentation variants associate with cutaneous melanoma and basal cell carcinoma.* Nat Genet, 2008. **40**(7): p. 886-91.

74. Kim, H.K. and S.J. Chanock, *Genome-wide association studies in melanoma: off to a good start.* Pigment Cell Melanoma Res, 2012.

75. Xia, B., et al., *Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2.* Mol Cell, 2006. **22**(6): p. 719-29.

76. Tischkowitz, M. and B. Xia, *PALB2/FANCN: recombining cancer and Fanconi anemia.* Cancer Res, 2010. **70**(19): p. 7353-9.

77. Rahman, N., et al., *PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene.* Nat Genet, 2007. **39**(2): p. 165-7.

78. Tischkowitz, M., et al., *Rare germline mutations in PALB2 and breast cancer risk: A population-based study.* Hum Mutat, 2012.

79. Jones, S., et al., *Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene.* Science, 2009. **324**(5924): p. 217.

80. Stadler, Z.K., et al., *Germline PALB2 mutation analysis in breast-pancreas cancer families.* J Med Genet, 2011. **48**(8): p. 523-5.

81. Sabbaghian, N., et al., *Mutation analysis of the PALB2 cancer predisposition gene in familial melanoma.* Fam Cancer, 2011. **10**(2): p. 315-7.

82. Goding, C.R., *Mitf from neural crest to melanoma: signal transduction and transcription in the melanocyte lineage.* Genes Dev, 2000. **14**(14): p. 1712-28.

83. Harris, M.L., et al., *Sox proteins in melanocyte development and melanoma.* Pigment Cell Melanoma Res, 2010. **23**(4): p. 496-513.

84. Davies, H., et al., *Mutations of the BRAF gene in human cancer.* Nature, 2002. **417**(6892): p. 949-54.

85. Wellbrock, C., M. Karasarides, and R. Marais, *The RAF proteins take centre stage.* Nat Rev Mol Cell Biol, 2004. **5**(11): p. 875-85.

86. Wan, P.T., et al., *Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF.* Cell, 2004. **116**(6): p. 855-67.

87. Colombino, M., et al., *BRAF/NRAS mutation frequencies among primary tumors and metastases in patients with melanoma.* J Clin Oncol, 2012. **30**(20): p. 2522-9.

88. Platz, A., et al., *Human cutaneous melanoma; a review of NRAS and BRAF mutation frequencies in relation to histogenetic subclass and body site.* Mol Oncol, 2008. **1**(4): p. 395-405.

89. Omholt, K., et al., *NRAS and BRAF mutations arise early during melanoma pathogenesis and are preserved throughout tumor progression.* Clin Cancer Res, 2003. **9**(17): p. 6483-8.

90. Shinozaki, M., et al., *Incidence of BRAF oncogene mutation and clinical relevance for primary cutaneous melanomas.* Clin Cancer Res, 2004. **10**(5): p. 1753-7.

91. Albino, A.P., et al., *Transforming ras genes from human melanoma: a manifestation of tumour heterogeneity?* Nature, 1984. **308**(5954): p. 69-72.

92. Albino, A.P., et al., *Analysis of ras oncogenes in malignant melanoma and precursor lesions: correlation of point mutations with differentiation phenotype.* Oncogene, 1989. **4**(11): p. 1363-74.

93. Goydos, J.S., et al., *Detection of B-RAF and N-RAS mutations in human melanoma.* J Am Coll Surg, 2005. **200**(3): p. 362-70.

94. Tamura, M., et al., *Inhibition of cell migration, spreading, and focal adhesions by tumor suppressor PTEN.* Science, 1998. **280**(5369): p. 1614-7.

95. Curtin, J.A., et al., *Somatic activation of KIT in distinct subtypes of melanoma.* J Clin Oncol, 2006. **24**(26): p. 4340-6.

96. Harbour, J.W., *The genetics of uveal melanoma: an emerging framework for targeted therapy.* Pigment Cell Melanoma Res, 2012.

97. Lucas, R.M., *An epidemiological perspective of ultraviolet exposure--public health concerns.* Eye Contact Lens, 2011. **37**(4): p. 168-75.

98. Onken, M.D., et al., *Gene expression profiling in uveal melanoma reveals two molecular classes and predicts metastatic death.* Cancer Res, 2004. **64**(20): p. 7205-9.

99. Worley, L.A., et al., *Transcriptomic versus chromosomal prognostic markers and clinical outcome in uveal melanoma.* Clin Cancer Res, 2007. **13**(5): p. 1466-71.

100. Jonsson, G., et al., *Mapping of a novel ocular and cutaneous malignant melanoma susceptibility locus to chromosome 9q21.32.* J Natl Cancer Inst, 2005. **97**(18): p. 1377-82.

101. Van Raamsdonk, C.D., et al., *Frequent somatic mutations of GNAQ in uveal melanoma and blue naevi.* Nature, 2009. **457**(7229): p. 599-602.

102. Van Raamsdonk, C.D., et al., *Mutations in GNA11 in uveal melanoma.* N Engl J Med, 2010. **363**(23): p. 2191-9.

103. Harbour, J.W., et al., *Frequent mutation of BAP1 in metastasizing uveal melanomas.* Science, 2010. **330**(6009): p. 1410-3.

104. Testa, J.R., et al., *Germline BAP1 mutations predispose to malignant mesothelioma.* Nat Genet, 2011. **43**(10): p. 1022-5.

105. Abdel-Rahman, M.H., et al., *Germline BAP1 mutation predisposes to uveal melanoma, lung adenocarcinoma, meningioma, and other cancers.* J Med Genet, 2011. **48**(12): p. 856-9.

106. Bott, M., et al., *The nuclear deubiquitinase BAP1 is commonly inactivated by somatic mutations and 3p21.1 losses in malignant pleural mesothelioma.* Nat Genet, 2011. **43**(7): p. 668-72.

107. Wiesner, T., et al., *Germline mutations in BAP1 predispose to melanocytic tumors.* Nat Genet, 2011. **43**(10): p. 1018-21.

108. Njauw, C.N., et al., *Germline BAP1 Inactivation Is Preferentially Associated with Metastatic Ocular Melanoma and Cutaneous-Ocular Melanoma Families.* PLoS One, 2012. **7**(4): p. e35295.

109. Carbone, M., et al., *BAP1 cancer syndrome: malignant mesothelioma, uveal and cutaneous melanoma, and MBAITs.* J Transl Med, 2012. **10**(1): p. 179.

110. Hoiom, V., et al., *Hereditary uveal melanoma: a report of a germline mutation in BAP1.* Genes Chromosomes Cancer, 2013. **52**(4): p. 378-84.

111. Cheung, M., et al., *Further evidence for germline BAP1 mutations predisposing to melanoma and malignant mesothelioma.* Cancer Genet, 2013. **206**(5): p. 206-10.

112. Yoshida, K., M. Sanada, and S. Ogawa, *Deep sequencing in cancer research.* Jpn J Clin Oncol, 2013. **43**(2): p. 110-5.

113. Pleasance, E.D., et al., *A small-cell lung cancer genome with complex signatures of tobacco exposure.* Nature, 2010. **463**(7278): p. 184-90.

114. Mardis, E.R., et al., *Recurring mutations found by sequencing an acute myeloid leukemia genome.* N Engl J Med, 2009. **361**(11): p. 1058-66.

115. Berger, M.F., et al., *The genomic complexity of primary human prostate cancer.* Nature, 2011. **470**(7333): p. 214-20.

116. Bamshad, M.J., et al., *Exome sequencing as a tool for Mendelian disease gene discovery.* Nat Rev Genet, 2011. **12**(11): p. 745-55.

117. Pleasance, E.D., et al., *A comprehensive catalogue of somatic mutations from a human cancer genome.* Nature, 2010. **463**(7278): p. 191-6.

118. Wei, X., et al., *Exome sequencing identifies GRIN2A as frequently mutated in melanoma.* Nat Genet, 2011. **43**(5): p. 442-6.

119. Walia, V., et al., *Delving into somatic variation in sporadic melanoma.* Pigment Cell Melanoma Res, 2012.

120. Stark, M.S., et al., *Frequent somatic mutations in MAP3K5 and MAP3K9 in metastatic melanoma identified by exome sequencing.* Nat Genet, 2011. **44**(2): p. 165-9.

121. Nikolaev, S.I., et al., *Exome sequencing identifies recurrent somatic MAP2K1 and MAP2K2 mutations in melanoma.* Nat Genet, 2011. **44**(2): p. 133-9.

122. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.

123. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.

124. Aitken, J.F., et al., *The Queensland Familial Melanoma Project: study design and characteristics of participants.* Melanoma Res, 1996. **6**(2): p. 155-65.

125. Baxter, A.J., et al., *The Queensland Study of Melanoma: environmental and genetic associations (Q-MEGA); study design, baseline characteristics, and repeatability of phenotype and sun exposure measures.* Twin Res Hum Genet, 2008. **11**(2): p. 183-96.

126. McGregor, B., et al., *Genetic and environmental contributions to size, color, shape, and other characteristics of melanocytic naevi in a sample of adolescent twins.* Genet Epidemiol, 1999. **16**(1): p. 40-53.

127. Pingault, V., et al., *SOX10 mutations in patients with Waardenburg-Hirschsprung disease.* Nat Genet, 1998. **18**(2): p. 171-3.

128. Arias, S., *Genetic heterogeneity in the Waardenburg syndrome.* Birth Defects Orig Artic Ser, 1971. **07**(4): p. 87-101.

129. Ramsay, A.J., et al., *POT1 mutations cause telomere dysfunction in chronic lymphocytic leukemia.* Nat Genet, 2013. **45**(5): p. 526-30.

130. Horn, S., et al., *TERT promoter mutations in familial and sporadic melanoma.* Science, 2013. **339**(6122): p. 959-61.

131. Huang, F.W., et al., *Highly recurrent TERT promoter mutations in human melanoma.* Science, 2013. **339**(6122): p. 957-9.

132. Vajdic, C.M., et al., *Incidence of ocular melanoma in Australia from 1990 to 1998.* Int J Cancer, 2003. **105**(1): p. 117-22.

133. Popova, T., et al., *Germline BAP1 Mutations Predispose to Renal Cell Carcinomas.* Am J Hum Genet, 2013.

134. Aoude, L.G., et al., *Prevalence of germline BAP1 mutation in a population-based sample of uveal melanoma cases.* Pigment Cell Melanoma Res, 2013. **26**(2): p. 278-9.

135. de Lange, T., *Shelterin: the protein complex that shapes and safeguards human telomeres.* Genes Dev, 2005. **19**(18): p. 2100-10.

136. Moyzis, R.K., et al., *A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes.* Proc Natl Acad Sci U S A, 1988. **85**(18): p. 6622-6.

137. Lansdorp, P.M., et al., *Heterogeneity in telomere length of human chromosomes.* Hum Mol Genet, 1996. **5**(5): p. 685-91.

138. Autexier, C. and N.F. Lue, *The structure and function of telomerase reverse transcriptase.* Annu Rev Biochem, 2006. **75**: p. 493-517.

139. Griffith, J.D., et al., *Mammalian telomeres end in a large duplex loop.* Cell, 1999. **97**(4): p. 503-14.

140. Stansel, R.M., T. de Lange, and J.D. Griffith, *T-loop assembly in vitro involves binding of TRF2 near the 3' telomeric overhang.* EMBO J, 2001. **20**(19): p. 5532-40.

141. Xin, H., et al., *TPP1 is a homologue of ciliate TEBP-beta and interacts with POT1 to recruit telomerase.* Nature, 2007. **445**(7127): p. 559-62.

142. Zhong, F.L., et al., *TPP1 OB-fold domain controls telomere maintenance by recruiting telomerase to chromosome ends.* Cell, 2012. **150**(3): p. 481-94.

143. Wang, F., et al., *The POT1-TPP1 telomere complex is a telomerase processivity factor.* Nature, 2007. **445**(7127): p. 506-10.

144. Loayza, D. and T. De Lange, *POT1 as a terminal transducer of TRF1 telomere length control.* Nature, 2003. **423**(6943): p. 1013-8.

145. Ye, J.Z., et al., *POT1-interacting protein PIP1: a telomere length regulator that recruits POT1 to the TIN2/TRF1 complex.* Genes Dev, 2004. **18**(14): p. 1649-54.

146. Ye, J.Z. and T. de Lange, *TIN2 is a tankyrase 1 PARP modulator in the TRF1 telomere length control complex.* Nat Genet, 2004. **36**(6): p. 618-23.

147. Liu, D., et al., *PTOP interacts with POT1 and regulates its localization to telomeres.* Nat Cell Biol, 2004. **6**(7): p. 673-80.

148. Kendellen, M.F., K.S. Barrientos, and C.M. Counter, *POT1 association with TRF2 regulates telomere length.* Mol Cell Biol, 2009. **29**(20): p. 5611-9.

149. Bataille, V., et al., *Nevus size and number are associated with telomere length and represent potential markers of a decreased senescence in vivo.* Cancer Epidemiol Biomarkers Prev, 2007. **16**(7): p. 1499-502.

150. Greene, M.H., *The genetics of hereditary melanoma and nevi. 1998 update.* Cancer, 1999. **86**(11 Suppl): p. 2464-77.

151. Newton Bishop, J.A., M. Harland, and D.T. Bishop, *The genetics of melanoma: the UK experience.* Clin Exp Dermatol, 1998. **23**(4): p. 158-61.

152. Hanaoka, S., A. Nagadoi, and Y. Nishimura, *Comparison between TRF2 and TRF1 of their telomeric DNA-bound structures and DNA-binding activities.* Protein Sci, 2005. **14**(1): p. 119-30.

153. Court, R., et al., *How the human telomeric proteins TRF1 and TRF2 recognize telomeric DNA: a view from high-resolution crystal structures.* EMBO Rep, 2005. **6**(1): p. 39-45.

154. Smogorzewska, A., et al., *Control of human telomere length by TRF1 and TRF2.* Mol Cell Biol, 2000. **20**(5): p. 1659-68.

155. Broccoli, D., et al., *Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2.* Nat Genet, 1997. **17**(2): p. 231-5.

156. Nishikawa, T., et al., *Structure of the DNA-binding domain of human telomeric protein, TRF1 and its interaction with telomeric DNA.* Nucleic Acids Res Suppl, 2001(1): p. 273-4.

157. Sfeir, A., et al., *Loss of Rap1 induces telomere recombination in the absence of NHEJ or a DNA damage signal.* Science, 2010. **327**(5973): p. 1657-61.

158. Yang, D., et al., *TIN2 protein dyskeratosis congenita missense mutants are defective in association with telomerase.* J Biol Chem, 2011. **286**(26): p. 23022-30.

159. Yamaguchi, H., et al., *Identification of TINF2 gene mutations in adult Japanese patients with acquired bone marrow failure syndromes.* Br J Haematol, 2010. **150**(6): p. 725-7.

160. Savage, S.A., et al., *TINF2, a component of the shelterin telomere protection complex, is mutated in dyskeratosis congenita.* Am J Hum Genet, 2008. **82**(2): p. 501-9.

161. Sasa, G.S., et al., *Three novel truncating TINF2 mutations causing severe dyskeratosis congenita in early childhood.* Clin Genet, 2012. **81**(5): p. 470-8.

162. Sarper, N., E. Zengin, and S.C. Kilic, *A child with severe form of dyskeratosis congenita and TINF2 mutation of shelterin complex.* Pediatr Blood Cancer, 2010. **55**(6): p. 1185-6.

163. Flores, J.F., ey at., *Analysis of the CDKN2A, CDKN2B and CDK4 genes in 48 Australian melanoma kindreds.* Oncogene, 1997. **15**(24): 2999-3005.

164. Whiteman, D.C., et al. Germline *CDKN2A mutations in childhood melanoma.* J Natl Cancer Inst, 1999. **89(**19): 1460.

165. Holland, E.A., et al., *CDKN2A ($P16^{INK4a}$) and CDK4 mutation analysis in 131 Australian melanoma probands: effect of family history and multiple primary melanomas.* Genes Chromosomes Cancer, 1999. **25**(4): 339-48.

166. Rizos, H., et al., *Differential expression of p16INK4a and p16beta transcripts in B-lymphoblastoid cells from members of hereditary melanoma families without CDKN2A exon mutations.* Oncogene, 1997. **15**(5): 515-23.

167. Ghiorzo, P., et al., *Characterization of ligurian melanoma families and risk of occurrence of other neoplasia.* Int J Cancer, 1999. **83**(4): 441-8.

168. Landi, M.T., et al., *Genetic susceptibility in familial melanoma from northeastern Italy.* J Med Genet, 2004. **41**(7): 557-66.

169. Bergman, W., et al., *The Dutch FAMMM family material: clinical and genetic data.* Cytogenet Cell Genet, 1992. **59**(2-3): 161-4.

170. Ruiz, A., et al., *CDKN2A mutations in Spanish cutaneous malignant melanoma families and patients with multiple melanomas and other neoplasia*. J Med Genet, 1999. **36**(6): 490-3.

171. Platz, A., et al., *Screening of germline mutations in the CDK4, CDKN2C and TP53 genes in familial melanoma: a clinic-based population study*. Int J Cancer, 1998. **78**(1): 13-5.

172. Platz, A., et al., *Screening of germline mutations in the CDKN2A and CDKN2B genes in Swedish families with hereditary cutaneous melanoma.* Int J Cancer, 1997. **89**(10): 697-702.

173. Hashemi, J., et al., *Melanoma development in relation to non-functional p16/INK4A protein and dysplastic naevus syndrome in Swedish melanoma kindreds*. Melanoma Res, 1999. **9**(1): 21-30.

174. Borg, A., et al., *High frequency of multiple melanomas and breast and pancreas carcinomas in CDKN2A mutation-positive melanoma families.* J Nat Cancer Ist, 2000. **92**(15) 1260-6.

175. Borg, A., et al., *Novel germline p16 mutation in familial malignant melanoma in southern Sweden*. Cancer Res, 1996. **56**(11): 2497-500.

176. MacKie, R.M., *CDKN2A germline mutations in U.K. patients with familial melanoma and multiple primary melanomas*. J Invest Dermatol, 1998. **111**(2): 269-72.

177. Lang, J., *CDKN2A mutations in Scottish families with cutaneous melanoma: results from 32 newly identified families*. Br J Dermatol, 2005. **153**(6): 1121-5.

178. Newton Bishop, J.A., *Mutation testing in melanoma families: INK4A, CDK4 and INK4D.* Br J Cancer, 1999. **80**(1-2): 295-300.

179. Yang, G., *A novel methionine-53-valine mutation of p16 in a hereditary melanoma kindred.* J Invest Dermatol, 2004. **123**(3): 574-5.

180. Goldstein, A.M., *Genotype-phenotype relationships in U.S. melanoma-prone families with CDKN2A and CDK4 mutations.* J Natl Cancer Inst, 2000. **92**(12): 1006-10.

181. Eliason, M.J., *Population-based prevalence of CDKN2A mutations in Utah melanoma families.* J Invest Dermatol, 2006. **126**(3): 660-6.

Appendix

LETTER

doi:10.1038/nature10630

A novel recurrent mutation in *MITF* predisposes to familial and sporadic melanoma

Satoru Yokoyama[1]*, Susan L. Woods[2]*, Glen M. Boyle[2]*, Lauren G. Aoude[2]*, Stuart MacGregor[2]*, Victoria Zismann[3]*, Michael Gartside[2], Anne E. Cust[4], Rizwan Haq[1], Mark Harland[5], John C. Taylor[5], David L. Duffy[2], Kelly Holohan[2], Ken Dutton-Regester[2], Jane M. Palmer[2], Vanessa Bonazzi[2], Mitchell S. Stark[2], Judith Symmons[2], Matthew H. Law[2], Christopher Schmidt[2], Cathy Lanagan[2], Linda O'Connor[2], Elizabeth A. Holland[6], Helen Schmid[6], Judith A. Maskiell[7], Jodie Jetann[8], Megan Ferguson[8], Mark A. Jenkins[7], Richard F. Kefford[6], Graham G. Giles[9], Bruce K. Armstrong[4], Joanne F. Aitken[8], John L. Hopper[7], David C. Whiteman[2], Paul D. Pharoah[10], Douglas F. Easton[10], Alison M. Dunning[10], Julia A. Newton-Bishop[5], Grant W. Montgomery[2], Nicholas G. Martin[2], Graham J. Mann[6], D. Timothy Bishop[5], Hensin Tsao[11]*, Jeffrey M. Trent[3]*, David E. Fisher[1]*, Nicholas K. Hayward[2]* & Kevin M. Brown[3,12]*

So far, two genes associated with familial melanoma have been identified, accounting for a minority of genetic risk in families. Mutations in *CDKN2A* account for approximately 40% of familial cases[1], and predisposing mutations in *CDK4* have been reported in a very small number of melanoma kindreds[2]. Here we report the whole-genome sequencing of probands from several melanoma families, which we performed in order to identify other genes associated with familial melanoma. We identify one individual carrying a novel germline variant (coding DNA sequence c.G1075A; protein sequence p.E318K; rs149617956) in the melanoma-lineage-specific oncogene microphthalmia-associated transcription factor (*MITF*). Although the variant co-segregated with melanoma in some but not all cases in the family, linkage analysis of 31 families subsequently identified to carry the variant generated a log of odds (lod) score of 2.7 under a dominant model, indicating E318K as a possible intermediate risk variant. Consistent with this, the E318K variant was significantly associated with melanoma in a large Australian case–control sample. Likewise, it was similarly associated in an independent case–control sample from the United Kingdom. In the Australian sample, the variant allele was significantly over-represented in cases with a family history of melanoma, multiple primary melanomas, or both. The variant allele was also associated with increased naevus count and non-blue eye colour. Functional analysis of E318K showed that MITF encoded by the variant allele had impaired sumoylation and differentially regulated several MITF targets. These data indicate that *MITF* is a melanoma-predisposition gene and highlight the utility of whole-genome sequencing to identify novel rare variants associated with disease susceptibility.

Cutaneous malignant melanoma is predominantly a disease of fair-skinned individuals. Aetiology is complex, with environmental (mainly ultraviolet radiation exposure) and genetic factors affecting disease risk. Phenotypic risk factors, which are largely heritable, include pigmentation (fair skin, blue or green eyes, blonde or red hair), sun sensitivity, an inability to tan[3–6], high number of melanocytic naevi[7,8], or the presence of clinically atypical naevi[7]. Candidate-gene studies and genome-wide association studies (GWAS) for melanoma and these melanoma-associated phenotypes have identified several variants associated with melanoma risk in the general population[9–13]. Family studies, on the other hand, have identified only two high-penetrance melanoma genes, *CDKN2A* (ref. 1) and *CDK4* (ref. 2), accounting for a minority of genetic risk in melanoma families.

As part of a larger sequencing effort to identify novel melanoma risk genes, we sequenced the genome of an affected individual from an eight-case melanoma family negative for alterations in *CDKN2A* or *CDK4* (Fig. 1, FAM1) using a nanoarray-based short-read sequencing-by-ligation strategy[14]. From among the 410 novel variants predicted to affect protein structure, we prioritized for follow-up a single nucleotide polymorphism (SNP) resulting in a glutamic acid to lysine substitution

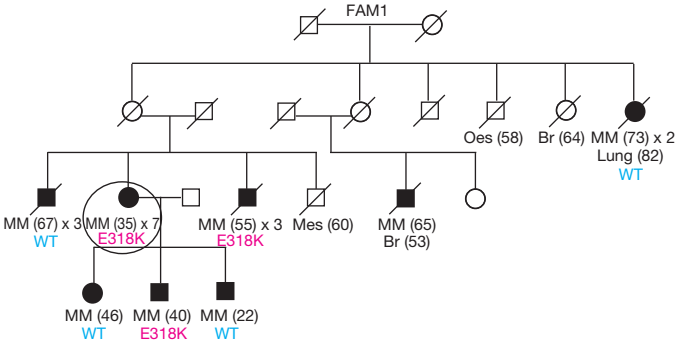


Figure 1 | Co-segregation analysis of the *MITF* E318K variant in the family in which it was identified. The pedigree shows individuals that have had melanoma (shaded circles or boxes), with the age of first melanoma diagnosis indicated in brackets and the number of melanomas that have occurred in the individual so far (for example, × 2 indicates two primary melanomas). If the number of melanomas is not stated, the individual has had a single melanoma. A diagonal line through the symbol indicates that the person is deceased. The genotype for the *MITF* E318K variant for individuals with an available DNA sample for testing is annotated 'E318K' if a carrier or wild type 'WT'. Other cancer types are also indicated with the age of first diagnosis indicated in brackets if known. Br, breast; Mes, mesothelioma; MM, melanoma; Oes, oesophagus. The individual circled in Family 1 (FAM1) is the melanoma case in which the *MITF* E318K variant was discovered through whole-genome sequencing. See Supplementary Fig. 3 for pedigrees of all other families identified as carrying E318K.

[1]Department of Dermatology, Cutaneous Biology Research Center, Massachusetts General Hospital, Harvard Medical School, Massachusetts 02114, USA. [2]Queensland Institute of Medical Research, Brisbane, Queensland 4029, Australia. [3]Translational Genomics Research Institute, Phoenix, Arizona 85004, USA. [4]Cancer Epidemiology and Services Research, Sydney School of Public Health, Sydney Medical School, The University of Sydney, New South Wales 2006, Australia. [5]Section of Epidemiology and Biostatistics, Leeds Institute of Molecular Medicine, Cancer Research UK Clinical Centre at Leeds, St James's University Hospital, Leeds LS9 7TF, UK. [6]Westmead Institute of Cancer Research, University of Sydney at Westmead Millennium Institute and Melanoma Institute Australia, Westmead, New South Wales 2145 Australia. [7]Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, School of Population Health, University of Melbourne, Victoria 3052 Australia. [8]Viertel Centre for Research in Cancer Control, The Cancer Council Queensland, Brisbane, Queensland 4004 Australia. [9]Cancer Epidemiology Centre, The Cancer Council Victoria, Carlton, Victoria 3053 Australia. [10]University of Cambridge, Cambridge CB1 8RN, England, United Kingdom. [11]Wellman Center for Photomedicine, Massachusetts General Hospital, Harvard Medical School, Massachusetts 02114, USA. [12]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Gaithersburg, Maryland 20892, USA.
*These authors contributed equally to this work.

Table 1 | *MITF* E318K association with melanoma

Population	Group	No. of individuals with variant	No. of individuals without variant	Variant carrier frequency	OR	95% CI	One-sided exact *P*	Two-sided exact *P*
Australia	Cases	34	2,025	0.0165	2.33	1.21–4.70	0.0045	0.0083
	Controls	14	1,939	0.0072				
UK	Cases	34	1,895	0.0176	2.09	1.14–3.94	0.0074	0.0115
	Controls	18	2,097	0.0085				
Australia + UK	Cases	68	3,920	0.0171	2.19	1.41–3.45	0.0001	0.0003
	Controls	32	4,036	0.0079				

It should be noted that the reported allele frequencies for *MITF* E318K in the population-based samples are without removing individuals with *CDKN2A* or *CDK4* mutations, as screening for these genes was not routinely performed.

in MITF (E318K, codon numbering based on the melanocyte-specific MITF-M isoform; c.G1075A, NCBI accession NM_000248.3; p.E318K, NCBI accession NP_000239.1; rs149617956). Although linkage[15] and GWAS studies[9,10] have not provided evidence implicating *MITF* in either predisposition to melanoma or the melanoma-associated phenotypes of pigmentation and naevogenesis[11,12,16–19], MITF is known to regulate a broad repertoire of genes whose functions in melanocytes range from development, differentiation, survival, cell-cycle regulation and pigment production. *MITF* is somatically amplified[20,21] or mutated[22] in a subset of melanomas, and strongly over-expressed in others[20], making it an attractive candidate despite the lack of prior evidence for involvement in germline risk.

We evaluated whether *MITF* E318K is a high-penetrance melanoma susceptibility variant in Family 1 by genotyping the remaining affected individuals available for study. The *MITF* variant allele was found in 3/7 melanoma cases assessed in this family (Fig. 1), consistent with it being a medium-penetrance melanoma risk variant. To assess further this possibility, we genotyped two large Australian melanoma case–control samples for *MITF* E318K. The variant was found in 14/1,953 controls (carrier frequency = 0.0072) and thus represents a rare population variant (Table 1). We observed a significantly higher frequency (34/2,059) in cases (carrier frequency = 0.0165) than controls (Fisher exact $P = 0.008$, odds ratio (OR) 2.33, 95% confidence interval (CI) 1.21–4.70), indicating that the variant correlates with increased melanoma risk in the general population. The effect size for E318K is larger than those reported for variants from melanoma GWAS[9,10] and similar to that observed for red-hair-colour-associated variants of the melanocortin 1 receptor (*MC1R*) gene (OR for most populations ~2.4)[23]. Among cases, the *MITF* E318K variant was enriched in those with multiple primary melanomas (OR 4.22, 95% CI 1.52–10.91), a family history of melanoma (OR 2.95, 95% CI 1.23–6.92), or both (OR 8.37, 95% CI 2.58–23.80), but not in cases with earlier age of onset (comparing diagnosis before age 40 versus after 40 years) (Table 2).

We replicated these findings in two independent population-based case–control samples from the United Kingdom. In the combined UK sample, the variant allele frequency was also significantly higher in cases (carrier frequency = 0.0176) than controls (carrier frequency = 0.0085, $P = 0.012$, OR 2.09, 95% CI 1.14–3.94, Table 1). The association with melanoma in the pooled UK and Australian data was highly significant (combined $P = 0.0003$, OR 2.19, 95% CI 1.41–3.45). In the UK cases there were also trends towards family history, earlier age of onset, and the occurrence of more than one primary melanoma in variant carriers (Table 2).

To extend assessment of the MITF variant in melanoma-prone families, we screened for E318K in 182 UK families with at least two melanoma cases and 88 Australian families with at least three cases, all of which are negative for mutations in *CDKN2A* or *CDK4*. Six families (2.2%) were found to carry the variant. In the UK, E318K was enriched in the more melanoma-dense families; 4/54 (7.4%) families with at least three melanoma cases versus 1/128 (0.8%) families with two melanoma cases (Fisher's exact $P = 0.013$). We subsequently evaluated whether *MITF* E318K co-segregated with melanoma in these as well as additional multiple-case families identified from the case–control sample. In total, we identified 31 unrelated cases carrying *MITF* E318K from Australia and the UK with at least one first- or second-degree relative diagnosed with melanoma (listed in Supplementary Table 1; Supplementary Fig. 3), 22 of which had DNA available from additional affected family members for genotyping. In 9/31 families (five three-case and four two-case families) the variant was found in all affected individuals (Supplementary Fig. 3a; non-segregating families shown in Supplementary Fig. 3b), whereas in 12 additional families, the variant co-segregated with melanoma in the available cases, but DNA from all cases was not available for screening (Supplementary Fig. 3c). To test more formally for linkage of melanoma with *MITF* E318K in these families, we calculated a lod score of 2.7 under a dominant model, again consistent with an incompletely penetrant medium risk variant.

To assess whether the *MITF* variant is related to known melanoma-associated risk phenotypes of pigmentation and naevus count, we tested for association both in cases and controls from the Australian and British populations. The *MITF* variant allele is significantly associated with increased naevus count (combined $P = 0.002$, OR 2.54, 95% CI 1.42–4.55; Supplementary Table 2) and non-blue eye colour (combined $P = 0.018$, OR 2.01, 95% CI 1.11–3.81; Supplementary Table 3). It was not associated with skin colour, hair colour, or freckling (Supplementary Table 4). Reassessing the case–control analysis accounting for naevus count and eye colour gave a slightly reduced effect size for association of the variant with melanoma (OR 1.82, 95% CI 0.85–3.92), suggesting that the risk of melanoma attributable to *MITF* E318K may be mediated at least in part via one or both of these phenotypes, but that there is a substantial residual risk conferred by the variant through an as yet undetermined mechanism.

We next sought to evaluate whether the E318K mutation alters MITF function. The E318K variant is located within one of two IKXE consensus sites on MITF previously shown to be post-translationally modified by the addition of the small ubiquitin-like-modifier SUMO[24].

Table 2 | Association of *MITF* E318K with melanoma-associated variables

Case set	Australia				UK			
	Carrier frequency	Carrier counts (carrier/wild type)	OR vs Aus controls	95% CI	Carrier frequency	Carrier counts (carrier/wild type)	OR vs UK controls	95% CI
All cases	0.0165	34/2,025	2.33	1.21–4.70	0.0176	34/1,895	2.09	1.14–3.94
Age of onset <40 years	0.0139	14/996	1.95	0.86–4.42	0.0174	5/283	2.10	0.60–5.91
Family history of melanoma	0.0209	12/563	2.95	1.23–6.92	0.0273	3/107	3.36	0.62–11.77
Multiple primary melanomas	0.0296	8/262	4.22	1.52–10.91	0.0225	2/87	2.74	0.30–11.74
Multiple melanomas and family history	0.0571	6/99	8.37	2.58–23.80	0.0000	0/10	-	-

Aus, Australia.

Mutation of the residue to which SUMO is covalently attached in this motif (K316R) has previously been shown to abrogate MITF sumoylation and significantly increase MITF transcriptional activity *in vitro*[24,25]. We thus hypothesized that E318K would similarly alter sumoylation and transcriptional activity of MITF. To test this we constructed a cDNA encoding His-tagged MITF carrying the E318K mutation. We evaluated the effects of E318K on sumoylation in comparison to the wild type and previously characterized synthetic mutations of the two known MITF sumoylation sites (K316R and K182R) by co-transfecting with haemagglutinin (HA)-tagged SUMO1 in COS-7 cells (Fig. 2a). Wild-type MITF shows two SUMO1-modified forms, whereas MITF mutants K182R or K316R each show only one modified form (Fig. 2a). Similar to the synthetic K316R and K182R mutants, E318K abrogates sumoylation, resulting in complete loss of the doubly sumoylated form of MITF and reducing the mono-sumoylated form. When the second site is mutated (K182R) simultaneously with E318K, MITF sumoylation is completely abolished. Immunoprecipitation of endogenously expressed MITF E318K from melanoma cells homozygous for E318K (NAE) when transfected with SUMO similarly revealed only bands corresponding to mono- and non-sumoylated isoforms of MITF on western blot (Fig. 2b).

We then looked for differences between mutant and wild-type MITF transcriptional activity using a reporter construct containing the MITF-responsive *TRPM1* promoter[25]. At two concentrations tested, the E318K mutant exhibited 1.34–1.40 fold induction of the *TRPM1* luciferase reporter relative to wild-type MITF (Fig. 2c). This fold induction is similar to that observed previously on multiple MITF target promoters using single or double artificial sumoylation-site MITF mutants[24,25] and suggests that the E318K variant found in melanoma changes the transcriptional potency of MITF. To study this in greater detail, we determined the effect of the E318K mutation on global MITF target gene transcription. We developed a tetracycline-inducible system for expression of wild-type MITF or the E318K variant in melanoma cell lines with constitutively low or undetectable levels of endogenous MITF (HT144 and C32, respectively[26], Fig. 2d). At the phenotypic level, induction of wild-type or E318K MITF led to increased proliferation compared to uninduced controls for each of the cell lines, although there was no significant difference in growth rate between the cells expressing the different isoforms (data not shown). We examined whole-genome expression profiles in these cells following induction of either wild-type or E318K MITF for 48 h. Of the 37 genes commonly regulated by wild-type and E318K MITF in both cell lines (Supplementary Fig. 1a, b; see Methods for analysis details), 28 (76%) had previously been identified as MITF target genes (Supplementary Table 5)[27,28], and 17 showed ≥1.25-fold differences in expression between the wild-type and E318K isoforms (Supplementary Fig. 1b). We also identified two gene products that were uniquely differentially regulated compared to uninduced cells by the induction of wild-type MITF but not MITF E318K in both parental cell lines, and 16 gene products after induction of MITF E318K but not wild-type MITF (Supplementary Table 6). Of these, 61% (11/18) have previously been reported as MITF targets (Supplementary Table 6)[27,28]. Collectively, these data indicate that the MITF E318K mutant exhibits differential transcriptional activity against some, although not all, target genes. In agreement with the reporter assays (Fig. 2c), we identified transcriptional differences in gene products known to be involved in pigmentation (*DCT*, *MLANA*), in which the differences were more marked with expression of the E318K variant in comparison to wild-type MITF. These were validated by quantitative polymerase chain reaction with reverse transcription (qRT–PCR) in the cell lines used for microarray analysis (Fig. 2e), as well as in an additional melanoma cell line constitutively expressing wild-type or E318K mutant MITF (Fig. 2f and Supplementary Fig. 2). In keeping with the increase in expression of these pigmentation genes, we detected a 22% increase in melanin content in HT-144 melanoma cells 72 h after induction of MITF E318K compared to wild-type MITF

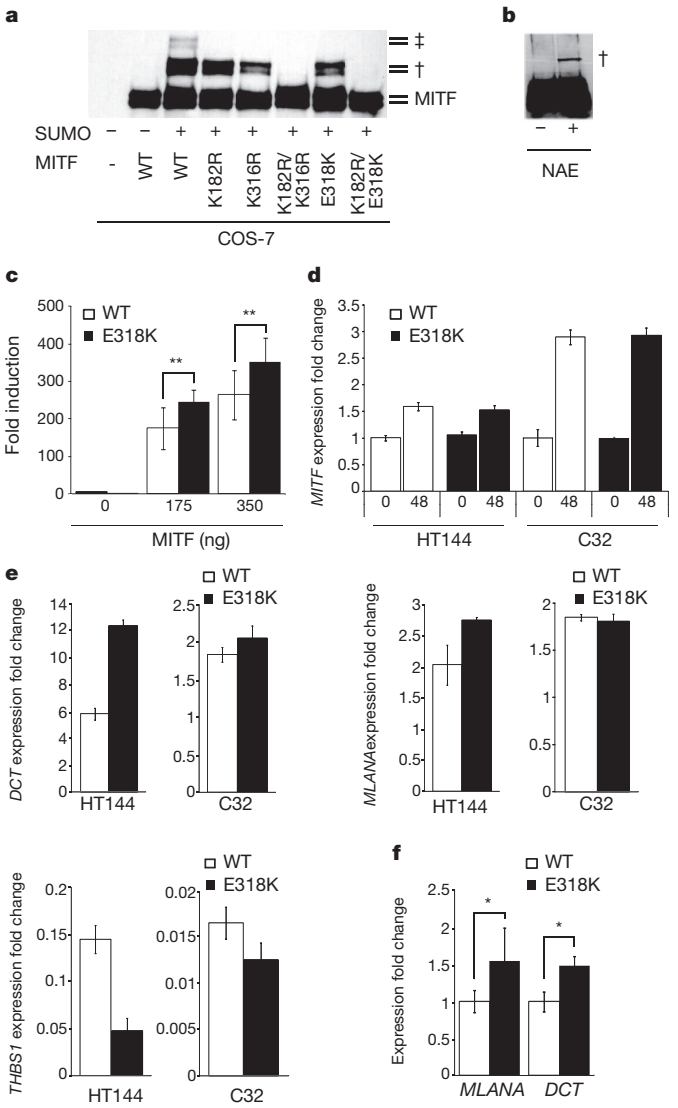


Figure 2 | E318K prevents MITF sumoylation and results in differential expression of MITF target genes. **a**, His-tagged wild-type MITF or the indicated single or double point mutants were co-transfected with HA–SUMO1 in COS-7 cells or **b**, HA–SUMO was transfected alone into homozygous mutant E318K MITF melanoma cells (NAE). Single- and double-sumoylated forms of MITF are indicated by a dagger and double dagger, respectively. The doublet bands are caused by MAPK-mediated phosphorylation at serine 73 (ref. 30). **c**, UACC62 human melanoma cells were transfected with *TRPM1*-promoter constructs with indicated amounts of expression vector encoding wild-type or mutant forms of MITF. Fold induction is shown as the ratio to the average of no MITF transfection (0 ng). Data are mean ± s.d. of at least four independent experiments. **d**, Expression of *MITF* in two melanoma cell lines (HT144 and C32) engineered to inducibly express wild-type (WT) or mutant (E318K) MITF after treatment with tetracycline for 48 h (48), as determined by qRT–PCR. Performed in triplicate, error bars depict s.d. **e**, Expression of MITF target genes *DCT* (top left), *MLANA* (top right) and *THBS1* (bottom left) determined by qRT–PCR in melanoma cell lines 48 h after induction of wild-type or E318K MITF. Gene expression is normalized to *GAPDH* and shown as fold change compared to uninduced cells. Performed in triplicate, error bars denote s.d. **f**, qRT–PCR analysis of total RNA isolated from UACC62 human melanoma cells, which were transfected with expression vector encoding wild-type or mutant forms of MITF. The expression level of each target gene was normalized to *MITF* mRNA. Fold induction is shown as the ratio to each mRNA expression with wild-type *MITF*. Data are mean ± s.d. of at least three independent experiments. $^{*}P < 0.05$, $^{**}P < 0.01$.

(data not shown). This is also consistent with our observation that carriers are more likely to have darker (that is, non-blue) eye colour (Supplementary Table 3) but, notably, these data contrast with other

previously reported 'fair-skin-associated' melanoma risk variants, such as those in *MC1R* or *TYR*. It is uncertain whether the enhanced expression of pigment genes may contribute to melanomagenesis, perhaps by increasing oxidative stress and an increase in oxidative DNA damage[29], or alternatively may simply reflect increased MITF activity, which (separately) promotes tumorigenesis, as *MITF* is a previously recognized amplified melanoma oncogene[20].

We adopted the approach of whole-genome sequencing of patients from melanoma families and identified a novel germline mutation of *MITF*. This mutation was found to be present in numerous melanoma families, as well as the general population, in which its association with melanoma has an effect size similar to red-hair-causing variants of *MC1R*[23]. The melanoma susceptibility genes discovered through GWAS so far account for only a minority of inherited disease risk. A proportion of this 'missing heritability' may be due to rare sequence variants, which are poorly detected by GWAS using SNP arrays. The new *MITF* variant reported here shows reasonably strong linkage to melanoma (lod score 2.7) but crucially not a high enough signal to be clearly visible in previous genome-wide linkage scans. We also provide *in vitro* data supporting a functional mechanism by which this mutation may mediate melanoma risk, specifically abrogation of MITF sumoylation and differential transcription of select MITF target genes. Although the individual changes in transcription induced by the mutant E318K MITF in comparison to wild-type MITF are modest, the orchestrated change in the levels of multiple MITF target genes is likely to be biologically important, especially over the lifetime of a person. This study offers a rare glimpse of a complex functionality whereby a risk-conferring SNP affects the post-translational processing of a crucial lineage-specific survival and differentiation gene. This study demonstrates the utility of performing whole-genome and exome resequencing in appropriate affected individuals to identify such novel rare disease-specific variants and functionally characterize variants associated with complex disease not otherwise detectable via GWAS or linkage approaches.

METHODS SUMMARY

The collection of the Australian melanoma families used for the study, as well as the Queensland and AMFS case–control sets are described elsewhere and in Methods. Likewise the UK studies from Leeds and Cambridge as well as the panel of melanoma cell lines. Whole-genome sequencing, assembly and variant calling were performed by Complete Genomics, as described previously[14]. Genotyping of *MITF* E318K was performed using the Sequenom MassArray system (Australian studies) or a custom TaqMan assay (UK studies), with DNA from the affected family member in which E318K was identified included multiple times as a positive control. Statistical analyses are described in detail in Methods. Co-segregation analyses were performed in melanoma families via Sanger sequencing using the primers: forward, 5′-CAGGCTCGAGCTCATGGA-3′; reverse, 5′-TGGGGACACTATAGGCTTGG-3′. MITF sumoylation and *TRPM1* reporter assays were performed as previously described[25].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 March; accepted 13 October 2011.
Published online 13 November 2011.

1. Goldstein, A. M. *et al.* Features associated with germline CDKN2A mutations: a GenoMEL study of melanoma-prone families from three continents. *J. Med. Genet.* **44,** 99–106 (2007).
2. Zuo, L. *et al.* Germline mutations in the p16INK4a binding domain of CDK4 in familial melanoma. *Nature Genet.* **12,** 97–99 (1996).
3. Naldi, L. *et al.* Cutaneous malignant melanoma in women. Phenotypic characteristics, sun exposure, and hormonal factors: a case–control study from Italy. *Ann. Epidemiol.* **15,** 545–550 (2005).
4. Titus-Ernstoff, L. *et al.* Pigmentary characteristics and moles in relation to melanoma risk. *Int. J. Cancer* **116,** 144–149 (2005).
5. Holly, E. A., Aston, D. A., Cress, R. D., Ahn, D. K. & Kristiansen, J. J. Cutaneous melanoma in women. I. Exposure to sunlight, ability to tan, and other risk factors related to ultraviolet light. *Am. J. Epidemiol.* **141,** 923–933 (1995).
6. Holly, E. A., Aston, D. A., Cress, R. D., Ahn, D. K. & Kristiansen, J. J. Cutaneous melanoma in women. II. Phenotypic characteristics and other host-related factors. *Am. J. Epidemiol.* **141,** 934–942 (1995).
7. Bataille, V. *et al.* Risk of cutaneous melanoma in relation to the numbers, types and sites of naevi: a case–control study. *Br. J. Cancer* **73,** 1605–1611 (1996).
8. Chang, Y. M. *et al.* A pooled analysis of melanocytic nevus phenotype and the risk of cutaneous melanoma at different latitudes. *Int. J. Cancer* **124,** 420–428 (2009).
9. Bishop, D. T. *et al.* Genome-wide association study identifies three loci associated with melanoma risk. *Nature Genet.* **41,** 920–925 (2009).
10. Brown, K. M. *et al.* Common sequence variants on 20q11.22 confer melanoma susceptibility. *Nature Genet.* **40,** 838–840 (2008).
11. Duffy, D. L. *et al. IRF4* variants have age-specific effects on nevus count and predispose to melanoma. *Am. J. Hum. Genet.* **87,** 6–16 (2010).
12. Falchi, M. *et al.* Genome-wide association study identifies variants at 9p21 and 22q13 associated with development of cutaneous nevi. *Nature Genet.* **41,** 915–919 (2009).
13. Gudbjartsson, D. F. *et al. ASIP* and *TYR* pigmentation variants associate with cutaneous melanoma and basal cell carcinoma. *Nature Genet.* **40,** 886–891 (2008).
14. Drmanac, R. *et al.* Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* **327,** 78–81 (2010).
15. Gillanders, E. *et al.* Localization of a novel melanoma susceptibility locus to 1p22. *Am. J. Hum. Genet.* **73,** 301–313 (2003).
16. Han, J. *et al.* A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS Genet.* **4,** e1000074 (2008).
17. Nan, H. *et al.* Genome-wide association study of tanning phenotype in a population of European ancestry. *J. Invest. Dermatol.* **129,** 2250–2257 (2009).
18. Sulem, P. *et al.* Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nature Genet.* **39,** 1443–1452 (2007).
19. Zhu, G. *et al.* A genome-wide scan for naevus count: linkage to CDKN2A and to other chromosome regions. *Eur. J. Hum. Genet.* **15,** 94–102 (2007).
20. Garraway, L. A. *et al.* Integrative genomic analyses identify *MITF* as a lineage survival oncogene amplified in malignant melanoma. *Nature* **436,** 117–122 (2005).
21. Stark, M. & Hayward, N. Genome-wide loss of heterozygosity and copy number analysis in melanoma using high-density single-nucleotide polymorphism arrays. *Cancer Res.* **67,** 2632–2642 (2007).
22. Cronin, J. C. *et al.* Frequent mutations in the MITF pathway in melanoma. *Pigment Cell Melanoma Res.* **22,** 435–444 (2009).
23. Williams, P. F., Olsen, C. M., Hayward, N. K. & Whiteman, D. C. Melanocortin 1 receptor and risk of cutaneous melanoma: a meta-analysis and estimates of population burden. *Int. J. Cancer* **129,** 1730–1740 (2010).
24. Murakami, H. & Arnheiter, H. Sumoylation modulates transcriptional activity of MITF in a promoter-specific manner. *Pigment Cell Res.* **18,** 265–277 (2005).
25. Miller, A. J., Levy, C., Davis, I. J., Razin, E. & Fisher, D. E. Sumoylation of MITF and its related family members TFE3 and TFEB. *J. Biol. Chem.* **280,** 146–155 (2005).
26. Boyle, G. M. *et al.* Melanoma cell invasiveness is regulated by miR-211 suppression of the BRN2 transcription factor. *Pigment Cell Melanoma Res.* **24,** 525–537 (2011).
27. Strub, T. *et al.* Essential role of microphthalmia transcription factor for DNA replication, mitosis and genomic stability in melanoma. *Oncogene* **30,** 2319–2332 (2011).
28. Hoek, K. S. *et al.* Novel MITF targets identified using a two-step DNA microarray strategy. *Pigment Cell Melanoma Res.* **21,** 665–676 (2008).
29. Smit, N. P. *et al.* Increased melanogenesis is a risk factor for oxidative DNA damage—study on cultured melanocytes and atypical nevus cells. *Photochem. Photobiol.* **84,** 550–555 (2008).
30. Hemesath, T. J., Price, E. R., Takemoto, C., Badalian, T. & Fisher, D. E. MAP kinase links the transcription factor Microphthalmia to c-Kit signalling in melanocytes. *Nature* **391,** 298–301 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by team science awards by the Melanoma Research Alliance (J.M.T., N.K.H., H.T. and D.E.F.), the American Cancer Society (K.M.B., RSG-08-200-01), the National Institutes of Health (NIH; D.E.F., AR043369-14; N.K.H., CA88363; H.T., K24CA149202 and P50CA93683), Doris Duke Medical Foundation (D.E.F.), Dr Miriam and Sheldon G. Adelson Medical Research Foundation (D.E.F.), US-Israel Binational Science Foundation (D.E.F.), and the Division of Cancer Epidemiology and Genetics of the National Cancer Institute (K.M.B.). N.K.H., D.L.D., S.M. and G.W.M. are supported by National Health and Medical Research Council of Australia (NHMRC) research fellowships. M.H.L. is supported by Cancer Australia grant 1011143. The collection of samples in the Leeds-based case–control study (the Melanoma Cohort Study) was funded by Cancer Research UK (Project Grant C8216/A6129 and Programme awards C588/A4994 and C588/A10589) and by the NIH (R01 CA83115). Recruitment was facilitated by the UK National Cancer Research Network. We thank S. Leake, S. Haynes, S. Waseem for Leeds case–control data collection; and H. Snowdon and C. Taylor from the Leeds Cancer Research UK Cancer Centre Genomics Facility for the genotyping of the UK samples. AMFS was supported by the NHMRC (project grants 566946, 107359, 211172 and program grant number 402761 to G.J.M. and R.F.K.); the Cancer Council New South Wales (project grant 77/00, 06/10), the Cancer Council Victoria and the Cancer Council Queensland (project grant 371); and the NIH (via RO1 grant CA-83115-01A2 to the international Melanoma Genetics Consortium—GenoMEL). The University of Cambridge SEARCH study was supported by Cancer Research UK Programme awards (C490/A11021 and C8197/A10123). A.E.C. is the recipient of an NHMRC public health postdoctoral fellowship (520018) and a Cancer Institute NSW Early Career Development Fellowship (10/ECF/2-06). B.K.A. is supported by a University of Sydney Medical Foundation Program Grant and J.L.H. is an Australia Fellow of the NHMRC. We gratefully acknowledge all of the participants, the work and dedication of the research coordinators, interviewers, examiners and data management staff, including J. Arbuckle, S. Columbus, M. Lang,

H. Rodais, C. Ellis (Centre for MEGA Epidemiology); E. A. Holland, C. Agha-Hamilton, C. El Hayek, L. Morgan, J. Roland, E. Tyler, J. Barton, C. Watts and L. Porter (Westmead Institute of Cancer Research); M. Hillcoat, K. Holland, P. Saunders, J. Roberts and S. Tait (Viertel Centre for Research in Cancer Control); A. Kurien, C. Patterson, C. Thoo, S. de Zwaan, A. Sklavos, S. Manoharan, J. Cahill and S. Brennand (skin examiners).

Author Contributions K.M.B., N.K.H. and J.M.T. designed, analysed and managed sequencing and genotyping aspects of the study. K.M.B., S.M., S.L.W. and N.K.H. wrote the paper. D.E.F. designed functional studies and contributed to manuscript preparation. J.M.T. designed genetic studies. H.T. designed functional studies, sequenced *MITF* in melanoma cell lines and contributed to manuscript preparation. Analysis of whole-genome sequence data was performed by K.H., M.S.S., N.K.H. and K.M.B. Genotyping of MITF E318K in all Australian samples was performed by V.Z., with statistical analyses performed by S.M. and M.H.L. D.C.W., D.L.D., G.W.M., N.K.H. and N.G.M oversaw collection of the Queensland samples and contributed to statistical analyses, data interpretation and manuscript preparation. G.J.M., E.A.H., H.S., J.A.M., J.J., M.F., M.A.J., R.F.K., G.G.G., B.K.A, J.F.A. and J.L.H. oversaw collection and contributed to phenotypic analyses of the AMFS study, with statistical analysis performed by A.E.C. D.T.B., J.A.N.-B., P.D.P., D.F.E. and A.M.D. designed and managed aspects of genotyping in UK case–control studies, with genotyping performed by M.H. and statistical analyses performed by J.C.T. J.M.P. assisted in analysis of Australian melanoma pedigrees. Sequencing of MITF in and analysis of Australian melanoma pedigrees and cell lines were performed by L.G.A., S.L.W., M.G., K.D.-R., V.B., M.S.S., J.S., C.S., C.L. and L.O'C. S.Y. and R.H. performed the sumoylation and MITF-transactivation studies and contributed to manuscript preparation. G.M.B. generated inducible-MITF melanoma cell lines, performed phenotypic and global transcript analyses and contributed to manuscript preparation.

Author Information Expression microarray data are available through the NCBI GEO website under accession GSE31269. Data for the full genome sequenced from FAM1 has been deposited in NCBI dbGAP under accession phs000419.v1.p1. Correspondence and requests for materials should be addressed to K.M.B. (brownkm2@mail.nih.gov).

METHODS

Melanoma families. Ascertainment and clinical evaluation of Australian and UK pedigrees with familial melanoma is described elsewhere[31–35]. In an effort to maximize the chance of identifying a novel melanoma susceptibility gene, we restricted inclusion in our study to pedigrees that were mutation negative for both *CDKN2A* and *CDK4*, by single-strand conformation polymorphism (SSCP) and/or direct sequencing.

Melanoma case–control studies. Australia. Cases were drawn from (1) a population-based sample from Queensland, unselected for age at onset (Queensland study of Melanoma: Environment and Genetic Associations; Q-MEGA[36]) and (2) a population-based case–control family study of melanoma diagnosed before age 40 years, ascertained in Brisbane, Melbourne and Sydney (AMFS)[37]. Two sets of Australian controls were used. First, a sample of unrelated individuals was ascertained; they were mainly parents of adolescent twins (80% of the sample), together with a smaller number of twins and their siblings, recruited through schools to participate in the Brisbane Adolescent Twin Study[38]. Additionally, a set of controls from AMFS were available for genotyping[37]. The AMFS and twin controls did not have a personal history of melanoma. Approval for these as well as the melanoma family studies was obtained from the Human Research Ethics Committees of Queensland Institute of Medical Research (QIMR), University of Sydney, University of Melbourne and cancer registries of New South Wales, Victoria and Queensland. Informed consent was obtained from all participants. Most samples had: questionnaire-based naevus count based on a four point scale: 'none', 'a few', 'moderate' and 'very many' (Queensland); or 'none', 'few', 'some' and 'many' for AMFS; data on pigmentation variables, including: hair colour (fair, light brown, red, dark brown, black), eye colour (blue/grey, green hazel, brown/black), skin colour (light, medium, dark) and number of freckles (none, few, some/moderate, many).

United Kingdom. The Leeds-based case-control study recruited population-based incident melanoma cases diagnosed between September 2000 and December 2006 from a geographically defined area of Yorkshire and the Northern region of the UK (63% response rate)[9,12,39]. Cases were identified by clinicians and pathology registers and via the Northern and Yorkshire Cancer Registry and Information Service to ensure overall ascertainment. For all but 18 months of the study period, recruitment was restricted to patients with Breslow thickness of at least 0.75 mm. Controls were ascertained by contacting general practitioners to identify eligible individuals. These controls were frequency-matched with cases for age and sex from general practitioners who had also had cases as a part of their patient register. Overall there was a 55% response rate for controls. The first 960 of the cases recruited and all controls were examined by trained interviewers who performed a standardized examination of the skin, recording naevi by anatomical site and size. For subsequent cases, self-reported information on naevi was obtained. In total we utilized 1,549 cases and 495 controls from this study. We also used 380 cases and 373 controls recruited by the University of Cambridge who were genotyped in the replication series. The cases and controls were recruited as a part of the SEARCH study[40,41], an ongoing population-based study in Eastern England. Cases were ascertained through the Eastern Cancer Registry and Information Centre, and were aged between 18 and 70 years at diagnosis. Controls were drawn from SEARCH and EPIC-Norfolk. The control set was supplemented with 1,245 additional controls from the Wellcome Trust Blood Services control set. Details of these studies have been previously published[40–42]. Approvals for the Leeds-based family studies, the Leeds case–control study and the Leeds Melanoma Cohort study were obtained from the National Research Ethics Service, Northern and Yorkshire Research Ethics Committee. Approvals for Cambridge case–control studies were obtained from the Cambridgeshire 4 Research Ethics Committee and the Norwich Local Research Ethics Committee. Informed consent was obtained from all participants.

Melanoma cell lines. The melanoma tumour cell lines used in this study were derived from primary cutaneous melanomas or melanoma metastases, as described previously[43]. DNA was extracted using QIAGEN QIamp Blood Maxi kits (Qiagen) according to the manufacturer's instructions.

DNA Sequencing (whole genome). One affected member from Family 1 was sequenced by Complete Genomics using a nanoarray-based short-read sequencing-by-ligation strategy. Sequencing, genome assembly (to genome build hg18/NCBI36.1), and variant calling were performed as described previously[14].

Sequenom MassARRAY genotyping of E318K. *MITF* E318K was genotyped in Australian samples as part of a multiplex assay designed using the Sequenom MassARRAY Assay Design software (version 3.0). The SNP was typed using Sequenom iPLEX chemistry on a MALDI-TOF Compact Mass Spectrometer (Sequenom). The 2.5 µl PCR reactions were performed in standard 384-well plates using 12.5 ng genomic DNA, 0.8 units of Taq polymerase (HotStarTaq, Qiagen), 500 µmol of each dNTP, 1.625 mM of $MgCl_2$ and 100 nmol of each PCR primer (Bioneer). Standard PCR thermal cycling conditions and post-PCR extension reactions were carried out as described previously[10]. The iPLEX reaction products were desalted by diluting samples with 15 µl of water and adding 3 µl of resin. The products were spotted on a SpectroChip (Sequenom), and data were processed and analysed by MassARRAY TYPER 3.4 software (Sequenom). DNA from the Family 1 sample in which E318K was identified was included multiple times as a positive control. As an additional quality control measure, we also confirmed the genotype of multiple individuals from the Australian case–control study, particularly those with a family history of melanoma, via Sanger sequencing, identifying no genotyping errors.

TaqMan genotyping of E318K. UK samples were genotyped for E318K using a custom TaqMan assay. DNA from the Australian family member in which E318K was identified was included as a positive control. Leeds melanoma pedigrees also were re-screened as a quality control measure using high-resolution melting (HRM) primers designed to assay E318K, with the results from HRM and TaqMan matching exactly.

Statistical analyses. Parametric linkage analysis on the multiple case melanoma families was conducted assuming a dominant model of inheritance, with penetrances specified as 5%, 95%, 95% for the three genotype classes. The allele frequency at the disease and marker locus was specified to be 0.004. Computations were done using the standard (lod score) parametric linkage routine implemented in GENEHUNTER-MODSCORE[44]. For melanoma case–control status, number of cases carrying the variant was compared to number of controls carrying the variant using Fisher's exact test. *P* values given are two-sided except where noted otherwise. Sub-phenotypes were tested similarly, with family history, age at onset and multiple primary melanoma carrier rates compared with that in controls. Number of blue-eyed individuals carrying the variant was compared to number of non-blue eyed individuals carrying the variant using Fisher's exact test. Meta-analysis for disease/sub-phenotype status and eye colour was done by combining the raw data. Association with mole count (high versus low where high was Australian Q-MEGA 'moderate/many', Australian AMFS 'some/many', UK '>25 moles' and low was Australian Q-MEGA 'none/few', Australian AMFS 'none/few', UK '<25 moles') was tested using Fisher's exact test for each sample individually, due to differences in the exact definition of low/high. Meta-analysis results for mole count were generated by combining the individual sample ORs (where the natural log of the odds ratio was weighted for each sample by the inverse variance of the natural log of the odds ratio). The case–control analysis corrected for naevus count and eye colour was computed by fitting a logistic regression with naevus count and eye colour categories as covariates.

Sanger sequencing for E318K. Australian pedigrees for which the proband was identified as carrying the E318K variant via Sequenom, as well as the panel of melanoma cell lines, were screened for E318K via Sanger sequencing using the following primers: forward, 5′-CAGGCTCGAGCTCATGGA-3′; reverse, 5′- TGGGGACACTATAGGCTTGG-3′.

MITF sumoylation assay. His–MITF and/or HA–SUMO1 expression vectors, previously described[25], were transfected into COS-7 or NAE cells. The cells were harvested after three days and 10 µg total protein was applied to each lane. Western blots were performed with anti-MITF monoclonal antibody (C5).

***TRPM1* reporter assay.** *TRPM1* reporter plasmid (100 ng), *Renilla* luciferase vector (1 ng) (as internal control), and MITF expression vector (0, 175 or 350 ng) were transfected into UACC62 human melanoma cells as previously described[25]. After 2 days, luciferase activity was determined. Transfection efficiency was normalized to *Renilla* luciferase activity. Results are reported as the average of at least four independent experiments.

Construction of inducible MITF-expressing cell lines. Stable inducible expression of MITF was achieved using the ViraPower T-REx Lentiviral Expression System (Invitrogen). Briefly, cell pools were transduced with lentiviral particles containing pLENTI6/TR, and selected with 3 µg ml^{-1} blasticidin. Wild-type MITF was cloned into pLENTI4/TO/V5-DEST using Gateway recombination. Site-directed mutagenesis of the putative sumolyation site was carried out using a QuickChange II XL site-directed mutagenesis kit (Stratagene/Agilent Technologies), and the following primer pair: E318K change: E318K forward, 5′- GGTGAATCGGATCATCAAGCAAAAACCCGTTCTTG -3′ and E318K reverse, 5′- CAAGAACGGGTTTTTGCTTGATGATCCGATTCACC -3′. The negative control construct was identical except that it encodes the β-galactosidase gene (Invitrogen). Lentivirus particles were packaged in 293FT cells, before being titred using MM96L cells. Target melanoma cell lines were transduced with *MITF* and mutant particles using a multiplicity of infection of less than 1, and selected with 500 µg ml^{-1} zeocin and 1 µg ml^{-1} blasticidin for 3 weeks. Cells were maintained on 100 µg ml^{-1} zeocin and 0.5 µg ml^{-1} blasticidin for all experiments.

Induction of MITF. Cells were treated the day after seeding with 1 µg ml^{-1} tetracycline to induce *MITF* expression, and harvested 24 or 48 h afterwards. RNA was extracted using Qiagen RNeasy mini kit as per manufacturer's instructions. RNA concentration was determined using the NanoDrop 2000 (Thermo

Scientific). For mRNA analyses 0.5 μg of total RNA was reverse transcribed using Superscript III according to the manufacturer's instructions (Invitrogen) and *MITF*, *DCT*, *MLANA*, *TBSP1* and *GAPDH* expression analysed using SYBR Green PCR Master Mix (Applied Biosystems). Primer sequences were *MITF* forward, 5′-CCAAGTACCACATACAGCAAGC-3′ and *MITF* reverse, 5′-TCA TCCATCTGCATACAGGACG-3′, *DCT* forward 5′-CGACTCTGATTAGTCG GAACTCA-3′ and *DCT* reverse, 5′-GGTGGTTGTAGTCATCCAAGC-3′, *MLANA* forward, 5′- GAGAAAAACTGTGAACCTGTGGT-3′ and *MLANA* reverse, 5′- GACTGTTCTGCAGAGAGTTTCTCAT-3′, *THBS1* forward, 5′-CGGTCCAGACACGGACCTGC-3′ and *THBS1* reverse 5′-GGCTTTGG TCTCCCGCGCTT-3′, *GAPDH* forward, 5′-GGCTCTCCAGAACATCATCC CTGC-3′ and *GAPDH* reverse, 5′-GGGTGTCGCTGTTGAAGTCAGAGG -3′. qRT–PCR was performed using a 7900HT Fast Real Time PCR System (Applied Biosystems).

Generation of constitutive MITF-expressing melanoma cell lines. pcDNA4.1-HisMax-hMITF (wild type, E318K or K182R/K316R) and pEGFP-C1 (for selection with G418) were transfected into the UACC62 melanoma cell line. After 5 days of selection with G418, total RNA was collected and *MITF* target gene transcript levels assessed by quantitative RT–PCR using the Pfaffl method[45].

Expression profiling and analysis. Biotinylated cRNA was prepared with the Illumina TotalPrep RNA Amplification Kit (Ambion). Labelled cRNA was hybridized to HumanHT-12 v4 BeadChip Arrays (Illumina), and then washed and scanned according to standard Illumina protocols. Data were extracted in GenomeStudio (Illumina) using default analysis settings and no normalization method. Resulting data were imported into GeneSpring GX v11.5 (Agilent Technologies). Expression values were normalized using quantile normalization with default settings.

Entities at least 1.5 fold up- or downregulated following induction of either wild-type or E318K MITF were determined (C32 wild type, 491; C32 E318K, 705; HT144 wild type, 273; HT144 E318K, 450). Targets similarly up-or downregulated only by E318K and not wild-type MITF in both of the melanoma cell lines were identified using Venn diagram ($n = 16$ entities). Also targets similarly up- or downregulated only by wild-type and not E318K MITF in both of the melanoma cell lines were identified using Venn diagram ($n = 2$ entities). Entities similarly up- or downregulated by induction of wild-type and E318K MITF were identified (C32 wild type and E318K, 244; HT144 wild type and E318K, 150), and those targets consistently differentially expressed in both of the cell lines identified by Venn diagram ($n = 43$ entities representing 37 gene products).

NCBI GEO (http://www.ncbi.nlm.nih.gov/geo/) accession numbers are as follows: GEO Series reference, GSE31269; GSM774914 HT144 wild-type uninduced; GSM774915 HT144 wild-type 48 h induced; GSM774916 HT144 E318K uninduced; GSM774917 HT144 E318K 48 h induced; GSM774918 C32 wild-type uninduced; GSM774919 C32 wild-type 48 h induced; GSM774920 C32 E318K uninduced; GSM774921 C32 E318K 48 h induced.

Melanin quantification assay. Melanin levels were determined[46] after 72 h induction of wild-type or E318K MITF, and normalized to amounts in uninduced control cells. Values are from duplicate readings from triplicate independent experiments.

31. Kefford, R. F., Salmon, J., Shaw, H. M., Donald, J. A. & McCarthy, W. H. Hereditary melanoma in Australia. Variable association with dysplastic nevi and absence of genetic linkage to chromosome 1p. *Cancer Genet. Cytogenet.* **51,** 45–55 (1991).
32. Nancarrow, D. J. *et al.* Linkage mapping of melanoma (MLM) using 172 microsatellite markers. *Genomics* **14,** 939–947 (1992).
33. Harland, M. *et al.* Germline mutations of the *CDKN2* gene in UK melanoma families. *Hum. Mol. Genet.* **6,** 2061–2067 (1997).
34. Harland, M. *et al.* Mutation screening of the *CDKN2A* promoter in melanoma families. *Genes Chromosom. Cancer* **28,** 45–57 (2000).
35. Harland, M., Mistry, S., Bishop, D. T. & Bishop, J. A. A deep intronic mutation in *CDKN2A* is associated with disease in a subset of melanoma pedigrees. *Hum. Mol. Genet.* **10,** 2679–2686 (2001).
36. Baxter, A. J. *et al.* The Queensland Study of Melanoma: environmental and genetic associations (Q-MEGA); study design, baseline characteristics, and repeatability of phenotype and sun exposure measures. *Twin Res. Hum. Genet.* **11,** 183–196 (2008).
37. Cust, A. E. *et al.* Population-based, case-control-family design to investigate genetic and environmental influences on melanoma risk: Australian Melanoma Family Study. *Am. J. Epidemiol.* **170,** 1541–1554 (2009).
38. McGregor, B. *et al.* Genetic and environmental contributions to size, color, shape, and other characteristics of melanocytic naevi in a sample of adolescent twins. *Genet. Epidemiol.* **16,** 40–53 (1999).
39. Newton-Bishop, J. A. *et al.* Melanocytic nevi, nevus genes, and melanoma risk in a large case–control study in the United Kingdom. *Cancer Epidemiol. Biomarkers Prev.* **19,** 2043–2054 (2010).
40. Pooley, K. A. *et al.* Common single-nucleotide polymorphisms in DNA double-strand break repair genes and breast cancer risk. *Cancer Epidemiol. Biomarkers Prev.* **17,** 3482–3489 (2008).
41. Pooley, K. A. *et al.* No association between TERT-CLPTM1L single nucleotide polymorphism rs401681 and mean telomere length or cancer risk. *Cancer Epidemiol. Biomarkers Prev.* **19,** 1862–1865 (2010).
42. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447,** 661–678 (2007).
43. Pavey, S. *et al.* Microarray expression profiling in melanoma reveals a BRAF mutation signature. *Oncogene* **23,** 4060–4067 (2004).
44. Strauch, K. Parametric linkage analysis with automatic optimization of the disease model parameters. *Am. J. Hum. Genet.* **73** (Suppl. 1), A2624 (2003).
45. Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT–PCR. *Nucleic Acids Res.* **29,** e45 (2001).
46. Oancea, E. *et al.* TRPM1 forms ion channels associated with melanin content in melanocytes. *Sci. Signal.* **2,** ra21 (2009).

doi:10.1038/nature10630

Supplementary Results

Association of MITF E318K with body site and Breslow thickness

When we analysed the available data on body site for Australian cases, there was some indication of differences by body site for cases that carried the *MITF* variant compared to those without the *MITF* variant. Specifically, there was a higher proportion of melanomas on the head and neck and a lower proportion on the legs ($P = 0.03$). However no clear differences by body site were observed for the UK cases ($P = 0.46$), nor when the UK and Australian data were combined ($P = 0.16$). Similar to our findings by body site, when we analysed the available data on Breslow thickness for Australian cases, those who carried the *MITF* variant had thinner primary tumours (median 0.38 mm) than those who did not (median 0.50 mm, $P = 0.003$). Once again, however, no clear differences by Breslow thickness were observed for the UK cases ($P = 0.80$) nor when the UK and Australian data were combined (Breslow thickness for cases with variant = 0.80 mm versus without variant = 0.85 mm; $P = 0.18$).

MITF E318K in melanoma cell lines

Screening of a panel of 205 cutaneous melanoma cell lines identified seven that carried the *MITF* E318K variant; one was homozygous (the only such sample in the entire study). The carrier frequency (0.034) was higher than in any of the case or control populations from Australia or the UK, thus providing further support for an association with melanoma and suggesting that *MITF* E318K may provide an advantage to the establishment of melanoma cell lines in culture.

Tumour-associated functions of select genes differentially regulated by MITF E318K

Several of the non-pigmentation genes that were differentially regulated between wild-type MITF and MITF E318K in melanoma cells (Supplementary Figure 1b) have been associated with melanoma or other cancer types. *SPINK1* (*serine peptidase inhibitor, Kazal type I*) expression was activated to a greater degree by MITF E318K than wild-type MITF, and is associated with poor survival in ovarian, kidney, colorectal, bladder and prostate cancer[1,2 and references therein]. *SPINK1* also acts as an oncogene to regulate proliferation, invasion and intravasation in prostate cancer[2]. We also observed more pronounced down-regulation of the anti-angiogenic gene (*THBS1*)[3] (validated by qRT-PCR, Fig. 2eiii) and cysteine-rich protein-61 (CYR61) with expression of E318K. *THBS1* and *CYR61* have recently been identified as tumour suppressors in melanoma[4, Bonazzi et al., submitted].

Supplementary Tables

Supplementary Table 1. Age of onset, multiple primary melanoma status of probands and *MITF* E318K cosegregation data in melanoma families.

Family ID	MM cases in family	MM cases with E318K	MM cases with unknown genotype	Multiple primary MM in proband	Age at onset of cases with E318K	Age at onset of other cases in family*
FAM1	8	3	1	Y	35, 40, 55	22, 46, 65, 67, 73
FAM2	7	5	1	Y	40, 50, 50, 54, <65	17~~, <69
FAM3	5	3	1	Y	45, 59, 65	59, <36
FAM4	4	1	3	N	67	20, 27, 50
FAM5	4	1	2	Y	34	42, 44, 72
FAM6	4	1	2	Y	53	31, 79, 51
FAM7	3	3	0	N	30, 32, 47	
FAM8	3	3	0	N	35, 62, 68	
FAM9	3	3	0	N	11, 49, 55	
FAM10	3	3~	0	Y	45, 58~, 76	
FAM11	3	3	0	N	20, 53, 57	
FAM12	3	2	0	Y	33, 45	79
FAM13	3	2	1	N	58,61	15
FAM14	3	2	1	N	65, 68	61
FAM15	3	1	2	N	57	79, <79
FAM16	3	2	1	Y	60, 72	88
FAM17	3	1	1	N	65	67, 61
FAM18	3	1	0	N	13	51, 46
FAM19	2	2	0	Y	18,32	
FAM20	2	2	0	Y	17, 51	
FAM21	2	2	0	N	24, 43	
FAM22	2	2	0	Y	24, 63	
FAM23	2	1	1	Y	?	?
FAM24	2	1	1	N	64	46
FAM25	2	1	1	Y	54	53
FAM26	2	1	1	N	15	<64^
FAM27	2	1	1	Y	38	72
FAM28	2	1	1	N	62	42
FAM29	2	1	1	N	35	49
FAM30	2	1	0	Y	52	59
FAM31	2	1	0	N	71	?

* includes typed & untyped cases
^ age at death from melanoma but age of onset unknwown
~includes 1 presumed obligate carrier
~~presumed wt
? denotes age unknown
E318K segregates
E318K may segregate

Supplementary Table 2. Nevus association with *MITF* E318K.

Sample	OR	se(ln(OR))	*P*	OR lower CI	OR upper CI
Q-MEGA	3.72	0.49	0.005	1.34	11.88
AMFS	1.69	0.60	0.574	0.47	7.55
UK	2.29	0.48	0.081	0.90	5.81
META	2.54	0.30	0.002	1.42	4.55

Meta-analysis was done on the basis of ORs and their standard errors, not on raw data, since there were differences in definition of "high"/"low" mole count

Supplementary Table 3. Eye colour association with *MITF* E318K.

	Blue Eyes			Non-Blue Eyes					
Sample	Variant Carriers	Non-Variant Carriers	Carrier Frequency	Variant Carriers	Non-Variant Carriers	Carrier Frequency	*P*	OR	95% CI
Aus (Q-MEGA+AMFS)	7	1265	0.0055	27	1721	0.0154	0.013	2.83	1.20-7.73
UK	9	681	0.0130	19	1069	0.0175	0.560	1.34	0.58-3.40
TOTAL	16	1946	0.0082	46	2790	0.0162	0.018	2.01	1.11-3.81

Supplementary Table 4. Lack of hair colour, skin colour and freckling associations with *MITF* E318K.

	Freckling		Skin Colour		Hair Colour	
Sample	**OR**	***P***	**OR**	***P***	**OR**	***P***
Aus-QIMR	1.59	0.326	1.77	0.360	0.87	1.000
Aus-AMFS	1.58	0.636	1.58	0.705	1.25	0.753
UK	1.31	0.228	1.10	0.484	0.98	0.771

Supplementary Table 5. The majority of gene products commonly regulated by wild-type and E318K mutant MITF have been previously reported to be MITF target genes.

			(Strub et al. 2011)		*(Hoek et al. 2008)*
		Gene Symbol	*Direct MITF Target (MITF ChIP-Seq)*	*Indirect MITF Target*	*Direct or Indirect MITF Target*
Gene products with >1.5-fold difference in expression compared to uninduced parental cell line AND <1.25-fold difference in expression between wild-type and E318K mutant MITF	***Increased with induction of MITF***	*SILV*	Yes	No	Yes
		CAPN3	Yes	No	Yes
		HSPA2	No	No	No
		SPRY1	No	No	No
		SPATS1	No	Yes	Yes
		LGALS3	Yes	No	Yes
		LOC730525	No	No	No
		MBP	Yes	No	Yes
		GYG2	Yes	No	Yes
		LAMA1	Yes	No	Yes
		ENO2	Yes	No	Yes
	Decreased with MITF induction	*KRT81*	No	Yes	N.D.
		ASNS	No	Yes	N.D.
		FLRT3	No	Yes	N.D.
		COL1A2	No	No	Yes
		KRT80	No	No	N.D.
		NPTX2	No	No	Yes
		JUN	Yes	Yes	N.D.
		PRKCDBP	No	No	Yes
		FST	No	No	N.D.
		PCLO	No	No	Yes
		TNFRSF12A	No	No	Yes
		CTGF	No	Yes	Yes
Gene products with >1.5-fold difference in expression compared to uninduced parental cell line AND ≥1.25-fold difference in expression between wild-type and E318K mutant MITF	***Increased with induction of MITF***	*SPINK1*	No	No	Yes
		PHACTR1	Yes	No	Yes
		TNFRSF14	Yes	No	Yes
		DCT	Yes	No	Yes
		CAPN3	Yes	No	Yes
		HSPA2	No	No	N.D.
		TSPAN10	Yes	No	Yes
		C18orf51	No	No	Yes
		TBC1D7	Yes	No	Yes
		Hs.406790	No	No	No
		PI15	Yes	No	Yes
		NDRG1	No	No	No
		MLANA	Yes	No	Yes
	Decreased with MITF induction	*TMEM16A*	No	No	Yes
		CYR61	Yes	No	Yes
		FST	No	No	N.D.
		THBS1	No	No	Yes

N.D., not determined as Hoek et al., 2008 did not include MITF downregulated genes.

Supplementary Table 6. Gene products uniquely regulated by either (a) E318K mutant MITF or (b) wild-type MITF in melanoma cell lines.

A. Gene products uniquely regulated by E318K mutant MITF and not wild-type MITF in both C-32 and HT144 melanoma cell lines as determined by microarray analysis.

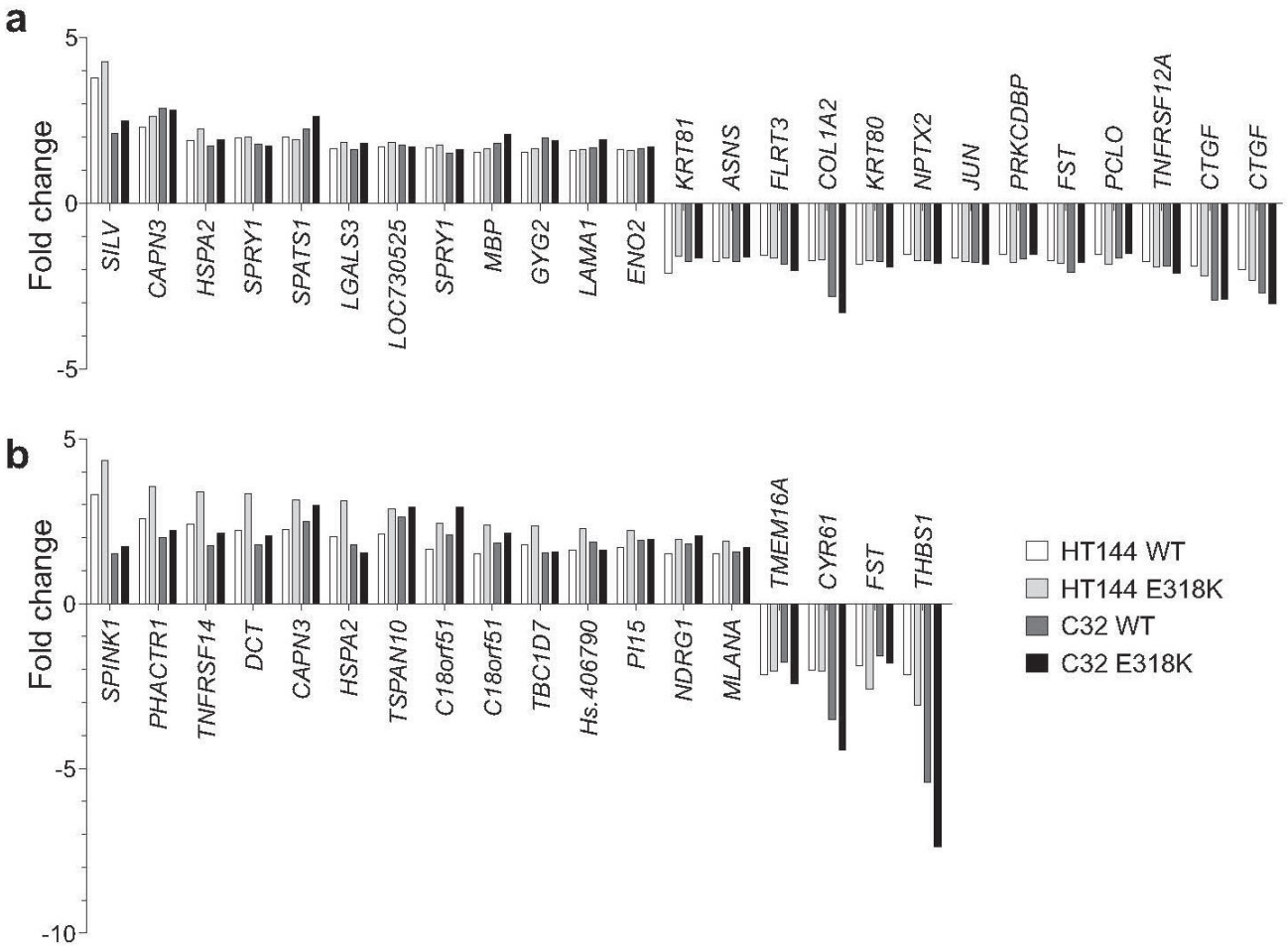
			(Strub et al. 2011)		*(Hoek et al. 2008)*
Gene Symbol	*Fold change in gene expression with induction of E318K mutant MITF in HT144 cells*	*Fold change in gene expression with induction of E318K mutant MITF in C32 cells*	*Direct MITF Target (MITF ChIP-Seq)*	*Indirect MITF Target*	*Direct or Indirect MITF Target*
NAGLU	1.85	1.58	Yes	No	Yes
FCGR2A	1.71	1.55	Yes	No	Yes
CASP1	1.68	1.50	No	No	Yes
TPD52L1	1.66	1.52	No	No	No
FAM69B	1.64	1.55	Yes	No	No
LASS4	1.59	1.70	Yes	No	No
CABC1	1.58	1.52	No	No	No
PRSS33	1.58	1.58	No	No	No
PNCK	1.52	1.55	No	No	No
CABLES1	1.51	1.68	Yes	No	Yes
FAM75A3	1.51	-1.52	No	No	No
ENC1	-1.70	-1.72	No	No	N.D.
KIF11	-1.63	-1.59	No	No	N.D.
LYPD1	-1.63	-1.51	No	No	Yes
PRNP	-1.58	-1.89	No	No	Yes
CAPN2	-1.54	-1.52	Yes	No	No

B. Gene products uniquely regulated by wild-type MITF and not E318K mutant MITF in both C-32 and HT144 melanoma cell lines as determined by microarray analysis.

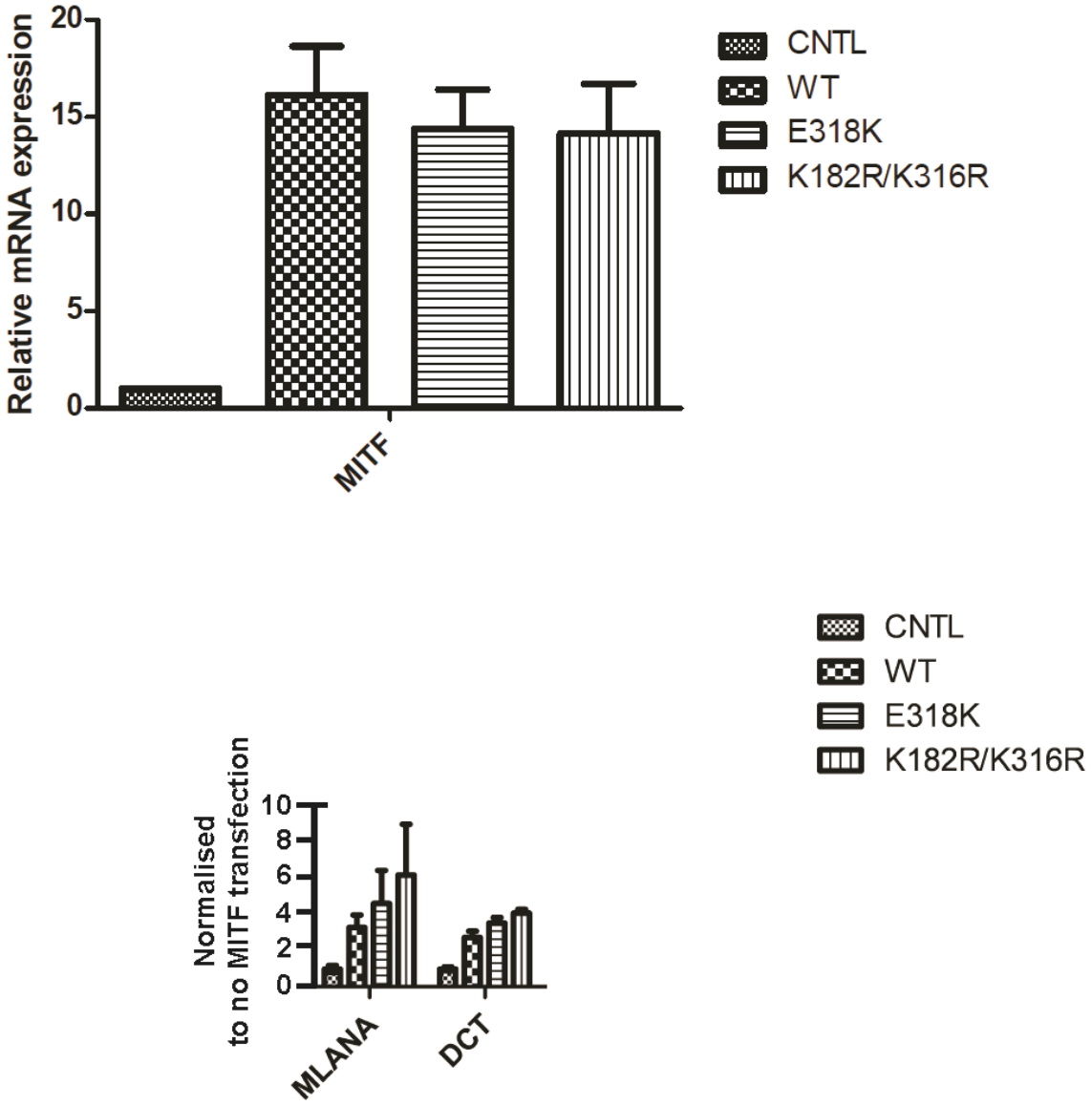
			(Strub et al. 2011)		*(Hoek et al. 2008)*
Gene Symbol	*Fold change in gene expression with induction of wildtype MITF in HT144 cells*	*Fold change in gene expression with induction of wildtype MITF in C32 cells*	*Direct MITF Target (MITF ChIP-Seq)*	*Indirect MITF Target*	*Direct or Indirect MITF Target*
TBC1D14	1.54	1.50	Yes	No	Yes
FXYD3	1.65	1.76	Yes	No	Yes

N.D., not determined as Hoek *et al.* (2008) did not include MITF downregulated genes.

Supplementary Figures and Legends

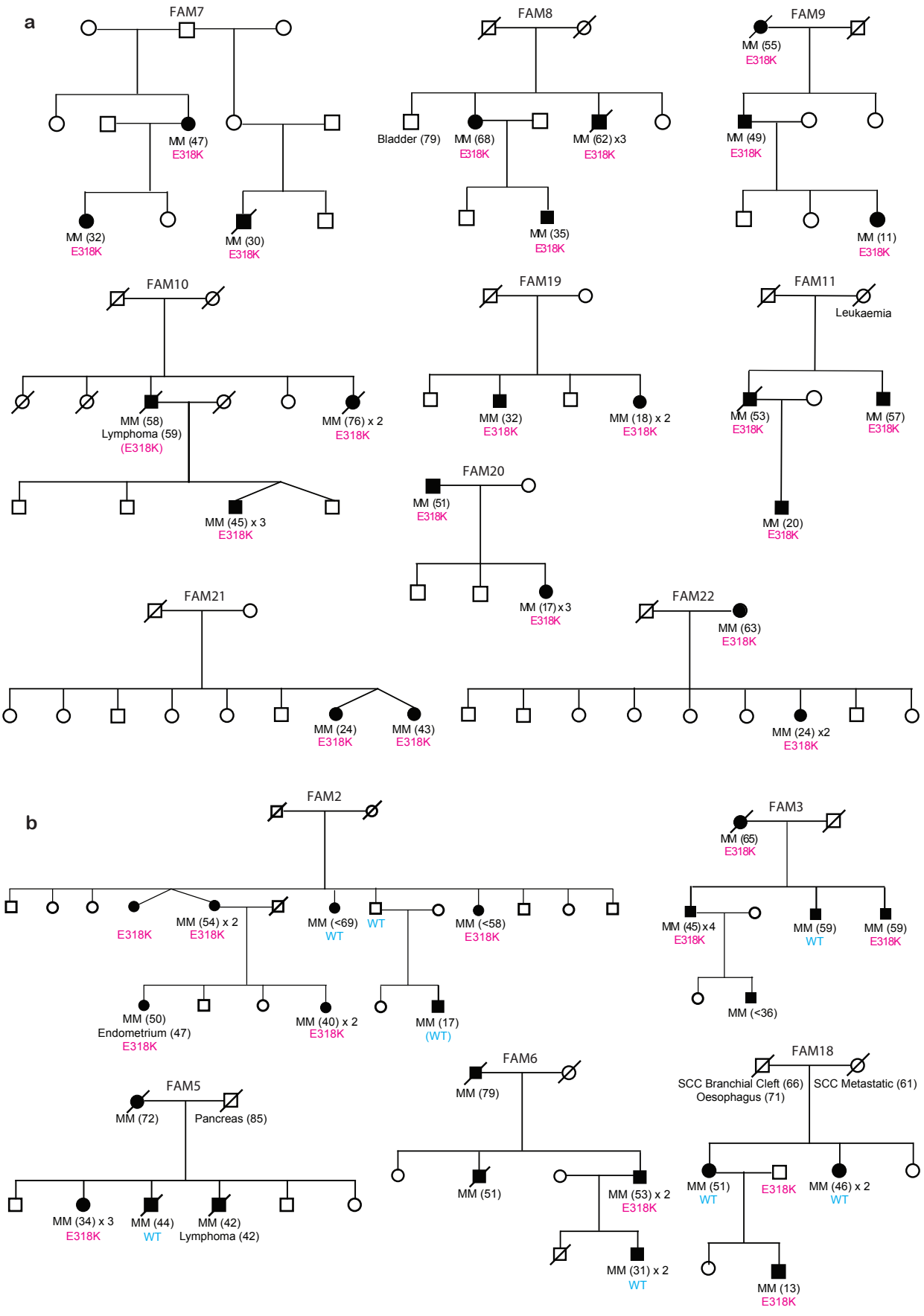


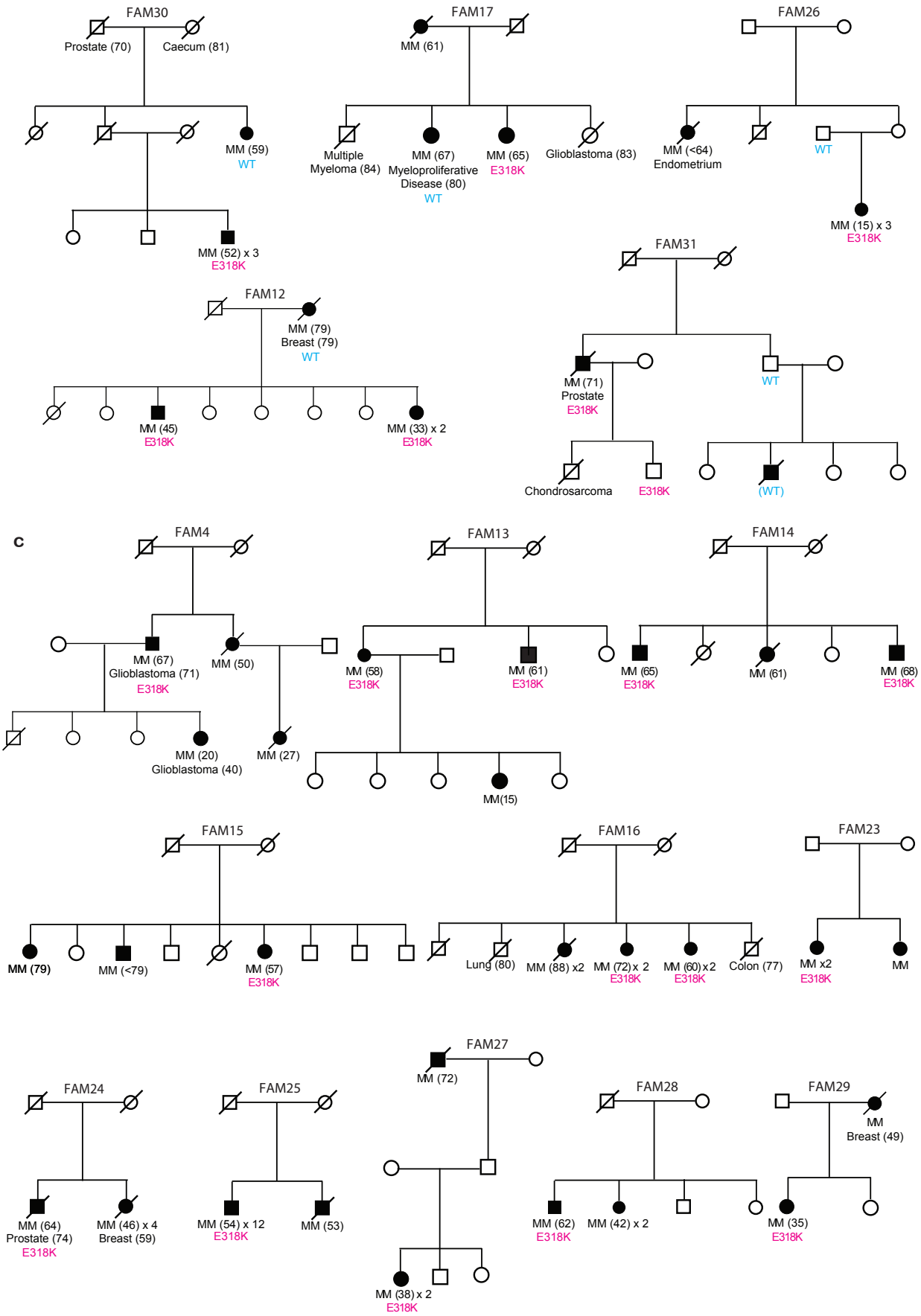
Supplementary Figure 1. Gene products commonly regulated by wild-type and E318K mutant MITF. Expression of either wild-type or E318K mutant MITF was induced by treatment with tetracycline for 48 hours prior to RNA extraction and whole genome expression profiling. A total of 43 entities representing 37 gene products were identified as being commonly regulated by wild-type and E318K mutant MITF in both of the cell lines. **a**, Gene products with less than 1.25-fold difference in expression between wild-type and E318K mutant MITF. **b**, Gene products with greater than 1.25-fold difference between E318K mutant and wild-type MITF in at least one of the two melanoma cell lines following induction. Fold changes are shown compared to the level of gene expression from un-induced control cells.



Supplementary Figure 2. Quantitative RT-PCR analysis of total RNA isolated from UACC62 human melanoma cells which were transfected with control vector, wild-type *MITF*, or the indicated mutants. Each target gene expression level was normalized to ACTB mRNA (Figure 2f shows data normalized to MITF expression). Fold induction is shown as the ratio to each mRNA expression with vector control. Data are mean ± SD of at least three independent experiments.

Supplementary Figure 3. Cosegregation analysis of the *MITF* E318K variant in melanoma families. Legend as for Fig. 1. **a,** Families in which all melanoma cases carry the *MITF* E318K variant (Excluding FAM1). **b.** Families in which the MITF E318K variant does not segregate perfectly with melanoma. **c.** Families in which the genotype of some melanoma cases is not known.


a
FAM7
MM (47)
E318K
MM (32)
E318K
MM (30)
E318K
FAM8
Bladder (79)
MM (68)
E318K
MM (62) x3
E318K
MM (35)
E318K
FAM9
MM (55)
E318K
MM (49)
E318K
MM (11)
E318K
FAM10
MM (58)
Lymphoma (59)
(E318K)
MM (76) x 2
E318K
MM (45) x 3
E318K
FAM19
MM (32)
E318K
MM (18) x 2
E318K
FAM11
Leukaemia
MM (53)
E318K
MM (57)
E318K
MM (20)
E318K
FAM20
MM (51)
E318K
MM (17) x 3
E318K
FAM21
MM (24)
E318K
MM (43)
E318K
FAM22
MM (63)
E318K
MM (24) x2
E318K
b
FAM2
E318K
MM (54) x 2
E318K
MM (<69)
WT
WT
MM (<58)
E318K
MM (50)
Endometrium (47)
E318K
MM (40) x 2
E318K
MM (17)
(WT)
FAM3
MM (65)
E318K
MM (45) x4
E318K
MM (59)
WT
MM (59)
E318K
MM (<36)
FAM5
MM (72)
Pancreas (85)
MM (34) x 3
E318K
MM (44)
WT
MM (42)
Lymphoma (42)
FAM6
MM (79)
MM (51)
MM (53) x 2
E318K
MM (31) x 2
WT
FAM18
SCC Branchial Cleft (66)
Oesophagus (71)
SCC Metastatic (61)
MM (51)
WT
E318K
MM (46) x 2
WT
MM (13)
E318K


c
FAM30
Prostate (70)
Caecum (81)
MM (59)
WT
MM (52) x 3
E318K
FAM17
MM (61)
Multiple Myeloma (84)
MM (67) Myeloproliferative Disease (80)
WT
MM (65)
E318K
Glioblastoma (83)
FAM26
MM (<64) Endometrium
WT
MM (15) x 3
E318K
FAM12
MM (79)
Breast (79)
WT
MM (45)
E318K
MM (33) x 2
E318K
FAM31
MM (71)
Prostate
E318K
WT
Chondrosarcoma
E318K
(WT)
FAM4
MM (67)
Glioblastoma (71)
E318K
MM (50)
MM (20)
Glioblastoma (40)
MM (27)
FAM13
MM (58)
E318K
MM (61)
E318K
MM(15)
FAM14
MM (65)
E318K
MM (61)
MM (68)
E318K
FAM15
MM (79)
MM (<79)
MM (57)
E318K
FAM16
Lung (80)
MM (88) x2
MM (72) x 2
E318K
MM (60) x2
E318K
Colon (77)
FAM23
MM x2
E318K
MM
FAM24
MM (64)
Prostate (74)
E318K
MM (46) x 4
Breast (59)
FAM25
MM (54) x 12
E318K
MM (53)
FAM27
MM (72)
MM (38) x 2
E318K
FAM28
MM (62)
E318K
MM (42) x 2
FAM29
MM
Breast (49)
MM (35)
E318K

Supplementary References

1 Gaber, A. *et al.* High expression of tumour-associated trypsin inhibitor correlates with liver metastasis and poor prognosis in colorectal cancer. *Br J Cancer* **100**, 1540-1548 (2009).

2 Ateeq, B. *et al.* Therapeutic targeting of SPINK1-positive prostate cancer. *Sci Transl Med* **3**, 72ra17 (2011).

3 Bornstein, P. Thrombospondins as matricellular modulators of cell function. *J Clin Invest* **107**, 929-934 (2001).

4 Dobroff, A. S. *et al.* Silencing cAMP-response element-binding protein (CREB) identifies CYR61 as a tumor suppressor gene in melanoma. *J Biol Chem* **284**, 26194-26206 (2009).