

Drivers of melanoma susceptibility



Carla Daniela Robles Espinoza

Wellcome Trust Sanger Institute

University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

Para quienes siempre estuvieron allí:
María Elena, Gabriel, Gabrielito, José y Daniel.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other University. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text. This dissertation contains less than 60,000 words as per the requirements of the Degree Committee for the Faculty of Biology.

Carla Daniela Robles Espinoza
September 2014

Acknowledgements

First and foremost, I would like to thank my supervisor Dave Adams for welcoming me into his lab after a rather “interesting” first meeting, and for his kindness and patience over these last few years. I have learnt from him not only things related to my project and science but also important lessons in life. He is a true mentor to me.

I would like to thank all members of the Experimental Cancer Genetics Team at the Sanger Institute for filling every working day with fun and interesting lunch-time discussions, and for their disposition to help whenever I asked: James, Alistair, Mamun, Martín, Clara, Stefan, Marco, Chi, Louise, Rebecca, Marcela, Mia, Gabi, Theo and Richard, and to Ximena who is practically a member of the team as well. I also thank Thomas and Sendu, from the Vertebrate Resequencing team, for their patience and help with analysis pipelines.

Also, none of this work would have been possible without my wonderful collaborators: Julia, Tim and Mark, from the University of Leeds, Nelleke, from the Leiden University Medical Centre, and Nick, Lauren and Antonia, from QIMR Berghofer. I thank them for trusting me with the analysis of their melanoma families, for patiently answering all questions I asked, and for very valuable advice regarding this project. I would also like to thank Andrew, Víctor and Carlos, from the University of Oviedo, for their expertise in protein analyses (and for answering e-mails at strange times during the night!).

I would like to acknowledge Moyra, Yip and Lucy, whom I met the first week I arrived in Cambridge and who have become some of my closest friends. I would like to thank them for keeping me sane, making sure I ate at adequate times and generally taking care of me during these last years. Equally, I owe much of my present happiness to Abi, Rachael and Sarah, whom I met on my first day at Sanger. I thank them for their care, for listening, and for the fun weekend trips we had during these past four years. I would like to thank Carlos, Mariana and Claudia, for their everlasting friendship and support across different time zones.

I would like to acknowledge the Wellcome Trust and the Consejo Nacional de Ciencia y Tecnología of Mexico, my funding bodies, for making sure I was not homeless and

hungry during these last years, and for trusting in me. This work would not have been possible without their generous support.

I would like to thank, from the bottom of my heart, those who stayed back home and have not forgotten me: My childhood friends, María and Linda, for awaiting me every time I go back to Mexico, for their life advice, help and understanding. I would like to thank my parents, María Elena and Gabriel, and my little brothers, Gabrielito and José, for their love, their neverending support and enthusiasm.

Finally, with all my love, I would like to thank Daniel. Above all, for understanding my constant travelling and moving to different countries. I will be forever thankful to him for filling every day with laughter, and for his love, support and patience.

Abstract

Cutaneous melanoma is a cancer of melanocytes, the pigment-producing cells in our skin. It is one of the most aggressive human malignancies, constituting only about 2% of all dermatological cancers but being responsible for over 75% of all deaths from skin cancer. It has recently become a major public health problem, as it is now the fifth most common cancer in the United Kingdom after its incidence more than quadrupled in the last three decades. For these reasons, understanding the biological processes that are involved in its development is of great importance for devising novel treatments and for the management of patients in the clinic.

The study of the genetic factors that influence melanoma risk can uncover mechanisms that are relevant in the transition from a benign melanocyte to a malignant melanoma. Approximately 10% of all melanoma cases are familial, and about half of these familial cases can be explained by pathogenetic variants in genes such as cyclin-dependent kinase inhibitor 2A (*CDKN2A*), cyclin-dependent kinase 4 (*CDK4*), breast cancer 2 (*BRCA2*), BRCA1-associated protein-1 (*BAP1*) and in the promoter of the telomerase reverse transcriptase (*TERT*). However, about 50% of all familial melanoma cannot be explained by mutations in known genes. In this dissertation, I detail the methodology I followed in an effort to uncover additional high-penetrance melanoma susceptibility genes.

I analysed exome and genome sequence data from a total of 184 individuals that belong to 105 melanoma-prone families from the United Kingdom, The Netherlands and Australia that did not have any pathogenetic variants in known susceptibility genes. I applied different gene prioritisation strategies and developed novel software tools in order to devise a list of plausible melanoma susceptibility candidate genes; these analyses suggested that genes regulating telomere function could be influencing melanoma risk. After performing functional experimental analyses, our research team was able to determine that carriers of rare variants in the protection of telomeres (*POT1*) gene, a member of the shelterin complex that safeguards telomere integrity, are at high risk for developing melanoma. We successfully described the mechanism by which this

happens, showing that the variants identified either disrupt *POT1* mRNA splicing or abolish the ability of POT1 to bind to telomeres, and lead to increased telomere length in carriers when compared to melanoma cases with wild-type *POT1*.

The main finding of the work described in this dissertation is the identification of telomere dysfunction as an important contributor to the risk of developing melanoma, and possibly other cancers. Our analyses suggest that *POT1* is the second most commonly mutated high-penetrance melanoma susceptibility gene reported thus far, and moreover, that rare variants in this gene constitute the first hereditary mechanism for telomere lengthening in humans.

Contents

Contents	i
List of Figures	v
List of Tables	ix
Nomenclature	xii
1 Introduction	13
1.1 From peas to man: A (very) brief history of genetics	13
1.2 The human genome: Its structure and gene content	14
1.3 The basis of cancer aetiology in humans	17
1.3.1 Environmental risk factors and their role in cancer development . .	18
1.3.2 Genetic risk factors and their influence on cancer development . .	25
1.4 Melanoma: Facts, origin and biology	38
1.4.1 Melanocytes: How they originate, where they reside, what they do	39
1.4.2 From melanocytes to melanoma: Different models of disease pro- gression	49
1.4.3 Clinical classification and staging systems	52
1.5 The human melanoma genome	57
1.6 Risk factors for melanoma development	61
1.6.1 Environmental factors	61
1.6.2 Physical characteristics and medical history	62
1.6.3 Genetic risk factors	65
1.7 Familial melanoma	67
1.8 Unanswered questions in melanoma genetics	68
1.8.1 Overview of whole-exome sequencing as a tool for providing answers	69

1.8.2	Overview of the methodology followed to study melanoma predisposition genes	70
2	Familial melanoma sequencing: European phase	73
2.1	Discovery phase	75
2.1.1	Patient selection	75
2.1.2	Exome sequencing	75
2.1.3	Data processing	76
2.1.4	Gene prioritisation for replication phase	78
2.1.5	Custom probe design	82
2.2	Replication phase	83
2.2.1	Targeted sequencing and data processing	83
2.3	Gene prioritisation strategy	85
2.3.1	Gene ranking methodology	85
2.3.2	Validation of next-generation sequencing-detected variants	89
2.3.3	Development of a visualisation tool for germline variants	91
2.3.4	Other analyses	92
2.4	Summary and conclusion	99
3	Familial melanoma sequencing: Integrative phase	101
3.1	Patient selection	101
3.2	Exome sequencing and data processing	103
3.3	Gene prioritisation strategy	105
3.4	Frequency of <i>POT1</i> and <i>ACD</i> variants in a control dataset	112
3.4.1	Variants in <i>POT1</i>	113
3.4.2	Variants in <i>ACD</i>	115
3.5	Summary and conclusion	116
4	Biological validation of candidate melanoma susceptibility genes	119
4.1	<i>SMG1</i> , the top candidate gene from the European phase of the study	120
4.1.1	Summary of variants found in <i>SMG1</i>	120
4.1.2	Biological function of the SMG1 protein	121
4.1.3	Description of biological assays	124
4.1.4	Conclusion	125
4.2	Candidate genes from the integrative phase	126
4.2.1	The biological role of <i>POT1</i> , <i>ACD</i> and the shelterin complex	126

4.2.2	The effect of <i>POT1</i> variants found in melanoma pedigrees	129
4.2.3	The effect of <i>ACD</i> variants found in melanoma pedigrees	144
4.2.4	Other genes in the shelterin complex	146
4.3	Summary and conclusion	147
5	Discussion	149
5.1	Summary of the work described in this dissertation	149
5.2	The identification of telomere dysregulation as an important contributor to familial melanoma	151
5.3	Future directions	155
	References	159
A	Electronic files and supplementary information	201
A.1	Description of electronic files	201
A.1.1	Pedigree structures for 24 melanoma-prone families sequenced as part of the discovery phase	201
A.1.2	Detailed information about recurrently mutated genes	201
A.1.3	List of biological pathways ordered by <i>P</i> -value after hypergeomet- ric tests on list of recurrently mutated genes	202
A.1.4	All genes included for capture in the replication phase	202
A.1.5	List of ranked genes after prioritisation stage	202
A.1.6	List of variants in the analysis of founder mutations and those in single pedigrees	203
A.1.7	Genes found with only one variant in multi-case pedigrees	203
A.1.8	Pedigrees sequenced as part of the integrative phase	203
A.1.9	Genes with co-segregating variants from the 28 pedigrees for which we had sequence data for 3 or more members and their GO terms	203
A.1.10	Wide variation in telomere measurements when samples have not been processed in the same manner	204
A.2	Supplementary tables, figures and notes	204
A.2.1	Removal of genes likely to be false positives after filtering	204
B	Articles published during my PhD	207
B.1	Articles directly associated with this dissertation	207
B.2	Other articles	209

List of Figures

1.1	Types of DNA damage caused by UV radiation	24
1.2	Paul Broca's pedigree evidences familial predisposition to cancer	26
1.3	Known melanoma susceptibility genes and their functions	35
1.4	Polygenic model of the distribution of cancer risk in the population and in individual cases	37
1.5	Percentage change in European age-standardised three-year average in- cidence rates, in the United Kingdom, for the ten most common cancers .	39
1.6	Origin of the melanocytic lineage	41
1.7	Skin structure and melanocyte location	42
1.8	Stages of melanosome maturation	45
1.9	Melanin biosynthesis pathway	46
1.10	Interactions between keratinocytes and melanocytes	48
1.11	The Clark model for the progression from naevi to malignant melanoma .	50
1.12	Types of melanoma	52
1.13	Survival curves from the American Joint Committee on Cancer Melanoma Staging Database grouped by TNM stage	57
1.14	Frequency and type of somatically-acquired mutations in melanoma gen- omes	59
1.15	Effect sizes and allele frequencies of loci influencing melanoma risk	65
1.16	Hybridisation methods for target DNA capture	70
1.17	Methodology followed during my PhD to study melanoma susceptibility genes	71
2.1	Flowchart of analysis steps followed in the search for melanoma suscept- ibility genes, European phase	74
2.2	Number of variants per exome remaining after filtering for common vari- ation and quality, discovery phase	78

2.3	Number of variants per pedigree remaining after filtering for co-segregation and protein-altering changes, discovery phase	78
2.4	Distribution of pairwise identity-by-descent values for samples in the discovery and replication sets	84
2.5	Principal component analysis plot showing that cases and controls are matched by ancestry, European phase	87
2.6	Percentage of variants confirmed by PCR to be real grouped by quality score	91
2.7	Example of a protein plot showing NGS-detected variants	93
2.8	Example of a putative founder variant found in melanoma pedigrees	97
3.1	Flowchart of analysis steps followed in the search for melanoma susceptibility genes, integrative phase	102
3.2	Pedigree structure for families with <i>POT1</i> variants	108
3.3	Pedigree structure for families with <i>ACD</i> variants.	109
3.4	Plot of the <i>POT1</i> and <i>ACD</i> variants found in familial melanoma pedigrees	111
3.5	Principal component analysis plot showing that cases and controls are matched by ancestry, integrative phase	114
4.1	Summary of variants detected in the SMG1 protein	122
4.2	Cell cycle profile for Mel-ST melanoma cells with a <i>SMG1</i> knockdown	125
4.3	The shelterin complex and its biological functions	127
4.4	MaxEntScan scores for the splice acceptor and splice donor variants detected in familial melanoma pedigrees	131
4.5	The <i>POT1</i> splice acceptor variant affects transcript splicing	132
4.6	Highly conserved residues of POT1 are altered in familial melanoma	133
4.7	Missense variants in <i>POT1</i> disrupt the interaction between POT1 and single-stranded DNA	134
4.8	³⁵ S gel showing the <i>in vitro</i> translation products of <i>POT1</i> wild-type (WT) and OB domain mutants	136
4.9	Linear model used to adjust bioinformatically calculated telomere lengths for age at blood draw and sex	138
4.10	Missense variants in <i>POT1</i> lead to elongated telomeres	139
4.11	Linear model used to adjust PCR mean ΔC_t values for age at blood draw and sex	141

4.12 PCR-based estimate of telomere length showing the carrier of a <i>POT1</i> intronic splice donor variant	142
4.13 Biolayer interferometry investigation of wild-type and mutated ACD in- teractions with POT1	145

List of Tables

1.1	Clark levels used to score tumour depth	54
1.2	Breslow thickness measurements used by the 2009 American Joint Committee on Cancer melanoma tumour staging system	55
1.3	Most recent pathologic stage categories for cutaneous melanoma as determined by the American Joint Committee on Cancer	56
2.1	Pedigrees sequenced as part of the discovery phase	76
2.2	Consequences of variants kept for further analyses, discovery phase	79
2.3	Top recurrently mutated genes, discovery phase	80
2.4	Significantly overrepresented biological pathways in set of recurrently mutated genes, discovery set	82
2.5	Control cohorts used in the melanoma gene prioritisation stage, European phase	85
2.6	Top 30 genes after prioritisation, European phase	90
2.7	Gene prioritisation after validation of NGS-detected variants	94
2.8	List of <i>MC1R</i> variants found in samples from the replication dataset	99
3.1	Summary of pedigrees sequenced by Institute and sequencing centre, integrative phase	104
3.2	Consequences of variants kept for further analyses, integrative phase	106
3.3	Genes with co-segregating variants in more than one pedigree with three members sequenced	107
3.4	Variants in <i>POT1</i> and <i>ACD</i> identified in melanoma pedigrees	110
4.1	Summary of all variants found in <i>SMG1</i>	121
4.2	Genomic location of bases encoding the 24 OB domain residues in close proximity to telomeres	135

A.1 Tools and parameters used for read alignment and variant calling in the
discovery and replication phases 205

Nomenclature

Roman Symbols

ACD Adrenocortical dysplasia homolog

BAP1 BRCA1 associated protein-1

BRCA2 Breast cancer 2, early onset

CDK4 Cyclin-dependent kinase 4

CDKN2A Cyclin-dependent kinase inhibitor 2A

CLL Chronic lymphocytic leukaemia

GWAS Genome-wide association studies

IARC International Agency for Research on Cancer

MC1R Melanocortin 1 receptor

MITF Microphthalmia-associated transcription factor

MPM Multiple primary melanomas

NGS Next-generation sequencing

OB Oligonucleotide/oligosaccharide-binding

POT1 Protection of telomeres

QFMP Queensland Familial Melanoma Project

RB1 Retinoblastoma protein 1

SMG1 SMG1 phosphatidylinositol 3-kinase-related kinase

TERT Telomerase reverse transcriptase

WES Whole-exome sequencing

WGS Whole-genome sequencing