Chapter 4

Biological validation of candidate melanoma susceptibility genes

Methods and results of this chapter have been published or accepted for publication (refs. [376] and [414]). Additionally, a review I wrote on sections covered by this chapter has also been accepted for publication (ref. [425]). Some parts of the text have been reproduced from these references; I confirm I have ownership of copyright for reproduction in this work.

In previous chapters, I have discussed the different methodologies we followed to pinpoint genes that might have a role in melanoma susceptibility. We found several genes with rare variants that co-segregate with melanoma in all affected members sequenced from each carrier family. These genes have roles in DNA repair upon UV irradiation or telomere protection, and the detected variants are predicted to alter protein structure or function. Therefore, these genes represent plausible candidates given their biological roles, and moreover, some of them are mutated at a higher-than-expected frequency in the melanoma cohort than in controls matched by ancestry.

However, *in silico* predictions of the consequences of these variants are insufficient to establish the causality of these genes in melanoma susceptibility, if any, and thus it is necessary to experimentally demonstrate their biological relevance. This Chapter explains the experimental procedures we performed to test the effect of these variants, and the consequences they may have in variant carriers.

4.1 *SMG1*, the top candidate gene from the European phase of the study

SMG1 was the highest-ranked candidate in the European phase of this study, and it also had a rare variant co-segregating with melanoma in multiple members of an Australian pedigree, which made this gene one of the top candidates of the integrative phase as well (Table 3.3). In order to investigate the role of *SMG1* in melanoma origin and progression, we attempted to assess whether *SMG1* had any influence on melanocyte transformation, to determine if changes of expression had any effect on cell cycle or proliferation.

4.1.1 Summary of variants found in *SMG1*

During the discovery phase of this study (Section 2.1), we had identified four variants in *SMG1* that segregated with the disease in four different families (Table 2.3). After PCR validation of these variants and testing of additional pedigree members, all of these with the exception of one were found to be real and co-segregate in all available affected members (Table 2.7). The variant that could not be confirmed due to poor sequencing traces, an insertion of one amino acid at position 18, was however detected by NGS in the two sequenced cases in that pedigree. Three variants in three different pedigrees were detected in the replication phase; however, two of these were found not to be real after PCR validation (see Table 2.7, these are the two variants that were successfully tested but that gave a negative result). During the integrative phase of this study, an additional rare variant in *SMG1* was found to co-segregate in all members of an Australian pedigree (Table 3.3). This variant was also confirmed by PCR. The PCR confirmation work and co-segregation assessment was performed by Dr. Mark Harland for Leeds samples and Lauren G. Aoude for Australian variants. All this information is summarised in Table 4.1.

We can appreciate, from visual inspection of the variants in a protein plot (generated by the tool discussed in Subsection 2.3.3), that all confirmed missense variants cluster in the C-terminus of the protein, outside any functional domains (Fig. 4.1). Two of these variants, S3047N and H3234Q, are found at a very low frequency in the set of exomes released by the NHLBI GO ESP [17] (1 out of 11,790 and 1 out of 12,146 alleles, respectively), but are found to co-segregate with melanoma in all three members of the pedigrees tested (Table 4.1). The two variants involving insertions or deletions, found

Table 4.1: **Summary of all variants found in *SMG1*.** The study phases in which each variant was identified is indicated. The number of individuals tested and that were found to carry the variant is shown in parentheses.

Study phase	Pedigree	Genomic position (GRCh37)	Protein	PCR	SIFT [417]	PolyPhen-2 [418]
Discovery	UF20 (4)	16:18840781, G/A	H3144Y	Real	Tolerated	Benign
Discovery	UF16 (3)	16:18841071, C/T	S3047N	Real	Tolerated	Benign
Discovery	UF17 (3)	16:18823187, C/A	V3602L, splice variant	Real	Tolerated	Probably damaging
Discovery	NF2 (2)	16:18937311, -/CGC	G18GG	Poor	-	-
Replication	RP1 (1)	16:18937327, CGT/-	S11-	Real	-	-
Replication	RP2	16:18880434, G/T	T942K	Not real	Tolerated	Probably damaging
Replication	RP3	16:18903640, G/A	S150L	Not real	Deleterious	Benign
Integrative	AF3 (3)	16:18839392, A/T	H3234Q	Real	Tolerated	Benign

in pedigrees NF2 and RP1, are in a repetitive region at the N-terminal portion of the protein that consists of two serines followed by six glycines.

4.1.2 Biological function of the SMG1 protein

SMG1 belongs to the PI3K-related kinase (PIKK) family of proteins, which includes ATM, ATR, MTOR and the DNA-dependent protein kinase, catalytic subunit (DNA-PK_{cs}, encoded by *PRKDC*). These proteins have important roles in the response to DNA damage, mitogenic signalling, and DNA break repair [426, 427]. Similarly to these proteins, SMG1 has been found to participate in DNA damage response, being activated upon UV or ionising radiation and participating in P53 phosphorylation and stabilisation [413]. It also participates in the nonsense-mediated mRNA decay (NMD) pathway, a process that monitors and destroys transcripts with premature termination codons (reviewed in [428]). In this pathway, it phosphorylates UPF1 regulator of nonsense transcripts homolog (UPF1), an essential helicase that is recruited to mRNA molecules

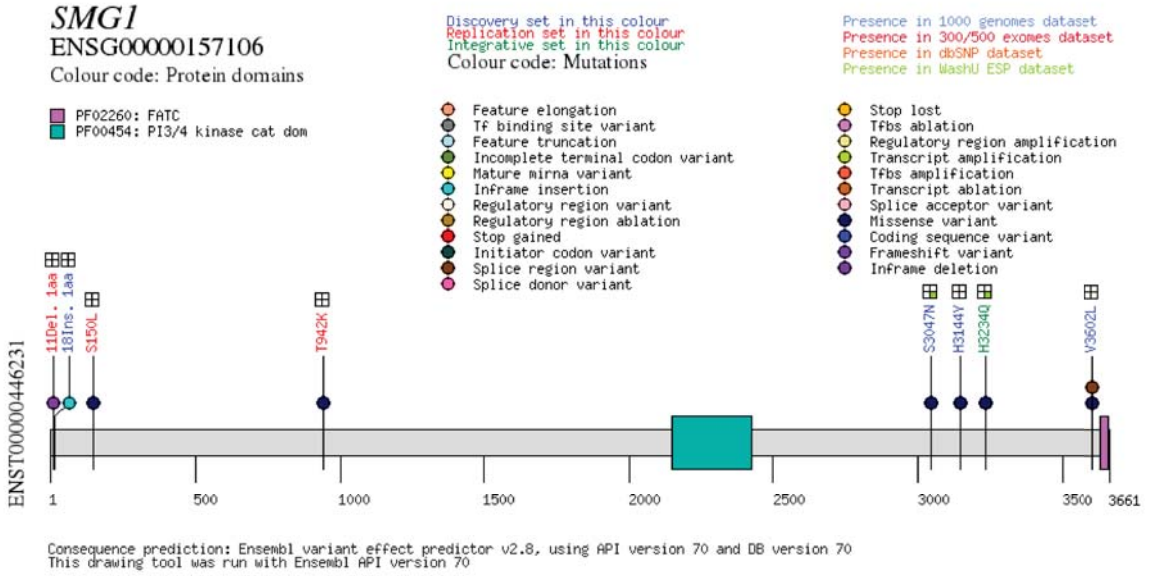

Figure 4.1: **Summary of variants detected in the SMG1 protein.** The variants we detected in familial melanoma pedigrees are shown in protein context, the transcript in the picture belongs to the consensus coding DNA sequence (CCDS45430). The four variants identified in the discovery part of the European phase are depicted in blue (Table 2.3), variants identified in the replication part are shown in red. The variant found co-segregating in an Australian pedigree is shown in green (Table 3.3). Subsequent PCR confirmation determined that two of the variants identified in the replication phase, S150L and T942K, were not real (see Table 2.7, these were the variants successfully tested but with a negative result). All other variants were confirmed as real with the exception of the 1 aminoacid insertion at position 18, as sequence in the trace was poor and could not be confirmed. However, this variant was found in two members of the same family by NGS. All variants co-segregate perfectly with melanoma in all members tested.

upon recognition of stop codons by the translation machinery (reviewed in [428]). This phosphorylation is central to the NMD pathway, as over-expression of a kinase-mutant form of SMG1 was found to suppress NMD, whereas that of wild-type SMG1 enhances it [429].

Additionally, SMG1 has been found to have roles in other biological processes. It has been implicated, for example, in the cellular response to low oxygen levels and in tumour necrosis factor (TNF)-α-induced apoptosis. In the first case, SMG1 was shown to be activated by hypoxia and its depletion augmented the activity of hypoxia inducible factor 1, alpha subunit (HIF1A), a master regulator of cellular responses to low oxygen levels [430]. In the second case, SMG1 depletion, but not that of other proteins part of

the PIKK family, was shown to rapidly increase the rate of apoptosis induced by TNFα activity [431].

SMG1 was recently found to have a role in adipogenesis, possibly via its participation in another mRNA decay pathway, mediated by staufen double-stranded RNA binding protein 1 (STAU1) [432]. This pathway targets higher-order structures in the 3' untranslated region (3'UTR) of mRNAs, and is more active than NMD during adipogenesis. SMG1 down-regulation was found to delay adipogenesis possibly by phosphorylating UPF1, an enzyme that participates in both NMD and STAU1-mediated mRNA decay (SMD) [432]. Additionally, SMG1 might have a role in the regulation of brain phosphorylated α-synuclein (p-syn) levels, which are involved in neurodegenerative disorders such as Parkinson's disease with dementia. SMG1 knockdown was found to significantly increase the levels of p-syn levels in the brain, and its low expression correlated with higher levels of this phosphoprotein [433].

More intriguingly, SMG1 has also been linked to telomere maintenance, and independently, to cancer predisposition. Depletion of SMG1 was shown to induce loss of entire telomere tracts, as well as to induce chromosome and chromatid breaks [434]. It can localise to telomeres *in vivo* and antagonise the association of telomeric repeat–containing RNA (TERRA), which is the product of telomere transcription, with chromatin. By preventing TERRA association with telomeres, it might aid efficient telomere capping and cell cycle progression [435]. Although *Smg1* is homozygous lethal in mice, animals heterozygous for a knockout allele of *Smg1* were found to be predisposed to different cancers, such as lymphomas and lung adenocarcinomas, as well as having chronic inflammation [436]. These mice did not have any defects in the NMD machinery, but displayed tissue oxidative damage and low-level inflammation prior to the development of tumours. These conclusions are supported by human data, as *SMG1* has been found to be significantly mutated in lung adenocarcinoma [437], and mutations in this gene have been found in more than 5% of all assessed samples of cervix, stomach and intestine cancers in the Catalogue of Somatic Mutations in Cancer (COSMIC) database [438]. However, it may have a different role in myeloma and acute myeloid leukaemia, in which it could contribute positively to cancer progression [439, 440].

In conclusion, SMG1 has been found to play a role in many diverse biological pathways, namely NMD, SMD, hypoxia response, adipogenesis, TNFα-induced apoptosis, inflammation, DNA damage response upon UV and ionising radiation and telomere maintenance. Perhaps it is the last two, together with the observation that *Smg1*$^{+/-}$ mice are predisposed to cancer, those which are most relevant for the melanoma phenotype

in these families with *SMG1* variants. In the next subsection, I describe the different experiments that were attempted in order to prove the role of *SMG1* and these variants in melanoma formation.

4.1.3 Description of biological assays

The experiments detailed in this section were performed by Drs. Jessamy C. Tiffen and James Hewinson, from the Experimental Cancer Genetics team at the Sanger, in the course of a year. The first experiment attempted was a colony formation assay in soft agar, using a melanocyte cell line, referred to as pmel*, that already carries the activating *BRAF* V600E mutation [259]. Cells that only have activated BRAF do not spontaneously transform in soft agar, possibly due to oncogene-induced senescence [441]. The rationale behind this experiment was that if *SMG1* functions as a tumour suppressor in melanocytes (as it may be assumed given the biological roles discussed above), then cells might transform and form colonies upon its knockdown. However, although a stable knockdown of *SMG1* was achieved via the use of lentiviral vectors, various technical problems prevented us from completing this experiment. Mainly, mycoplasma contamination before cells were shipped from the laboratory of origin meant that results obtained thus far had to be repeated, and spontaneous transformation of control cells (*i.e.*, those not treated with a *SMG1* knockdown vector) prevented us from deriving any conclusive results. Given these difficulties, a different strategy was attempted.

The second experiment that was tried was a knockdown or over-expression of *SMG1* in the melanoma cell lines A375 and Mel-ST [240, 442]. The first of these was derived from a 54-year-old female with malignant melanoma, and the second one is a human melanoma cell line that is immortalised but not transformed. The idea was to address whether *SMG1* expression changes had any effect on on apoptosis, cell cycle progression and proliferation upon UV or ionising radiation. Interestingly, Mel-ST cells that had been depleted of SMG1 via short hairpin (sh)-RNA interference showed an increase in the proportion of cells in the S and G2/M phases of the cell cycle upon UV irradiation when compared with irradiated controls (Fig. 4.2), in accordance with previous reports [413]. This result might indicate that *SMG1*-deficient cells have checkpoint signalling defects upon UV-induced DNA damage, but there might be another mechanism in place to arrest them at the S phase of the cell cycle (Fig. 4.2). This might indicate that cells depleted for *SMG1* need additional hits to progress towards malignancy. However, the behaviour of these cells could not be reproduced in A375 cells, as these did not show any changes in cell cycle profile upon UV irradiation.

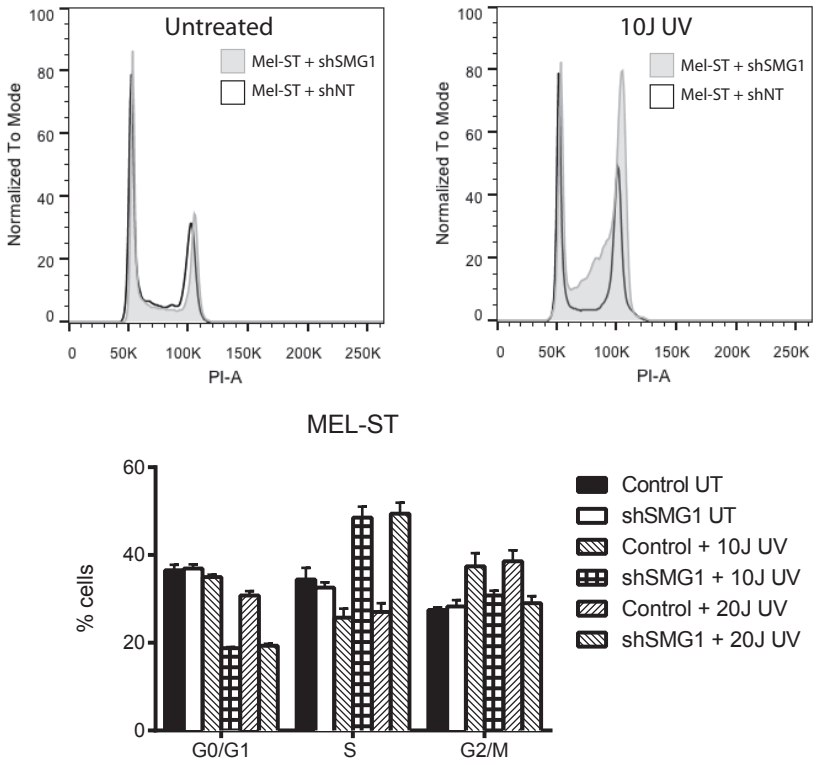

Figure 4.2: **Cell cycle profile for Mel-ST melanoma cells with a *SMG1* knockdown.** Top panel, cell cycle profiles. Gray line, Mel-ST with the anti-*SMG1* shRNA vector; black line, Mel-ST with a non-targeting vector. Two different UV intensities were assessed, note that only one is shown in the top panel. Bottom panel, bar chart showing the proportion of cells in each condition. PI: propidium iodide. I thank Drs. Jessamy C. Tiffen and Gabriel Balmus for this image.

Finally, site-directed mutagenesis is being attempted to assess whether the particular variants detected in melanoma pedigrees have any effect in SMG1 stability. So far, mutants have been generated in a cDNA expression vector (from ref. [429]), and have been confirmed to have the correct sequence. This experiment is on-going.

4.1.4 Conclusion

SMG1 is an attractive candidate for a melanoma susceptibility gene given the diverse biological roles it plays in the cell, among them, response to UV-induced DNA damage and P53 phosphorylation. We have attempted several different experimental methodologies to assess its role in melanocytes, namely *SMG1* knockdown, over-expression and site-directed mutagenesis in melanoma cell lines. However, studying *SMG1* has proven

difficult for several reasons, among them, cell contamination and irreproducibility of results between cell lines. We hope that we can address these concerns with further experiments.

4.2 Candidate genes from the integrative phase

During the integrative phase of this analysis, we prioritised a list of five genes, because they were found to have co-segregating variants with melanoma in two or more pedigrees for which we had sequences available for three or more affected individuals (Table 3.3). Of these, *POT1* and *ACD* were of special interest because they participate in telomere maintenance, and furthermore, interact with each other as part of the six-protein complex shelterin [423]. The location of the detected variants is also noteworthy, as all of the missense variants detected fall within functional domains in the protein structures, within the ssDNA-binding OB folds in the case of *POT1* and in the POT1-binding domain in the case of *ACD* (Fig. 3.4). The rest of the detected variants are predicted to be disruptive, introducing premature stop codons or affecting mRNA splicing. Additionally, we found variants in *POT1* at a statistically higher frequency in the melanoma pedigrees when compared with controls, even when co-segregation with melanoma was not considered, and genotyping of a case-control series identified a case that carried one of the familial *POT1* variants. This makes *POT1* and *ACD* attractive candidates for being melanoma susceptibility genes. This Section details the experiments carried out to test the involvement of the detected variants in melanoma susceptibility.

4.2.1 The biological role of *POT1*, *ACD* and the shelterin complex

Telomeres are structures at the ends of linear chromosomes that consist of double-stranded DNA repeats followed by a short ssDNA protrusion. They play an essential role in regulating genomic stability by allowing the cell to distinguish between chromosome ends and double-stand DNA breaks, and as such, they need to be replicated each cell cycle and protected from DNA-processing enzymes [443, 444]. The shelterin complex, a large macromolecular structure to which POT1 and ACD belong, binds telomeres and has a paramount role in their protection [423]. Shelterin displays a wide range of functions that not only include telomere length maintenance and protection from DNA repair mechanisms, but also the regulation of diverse signalling cascades from telomeres

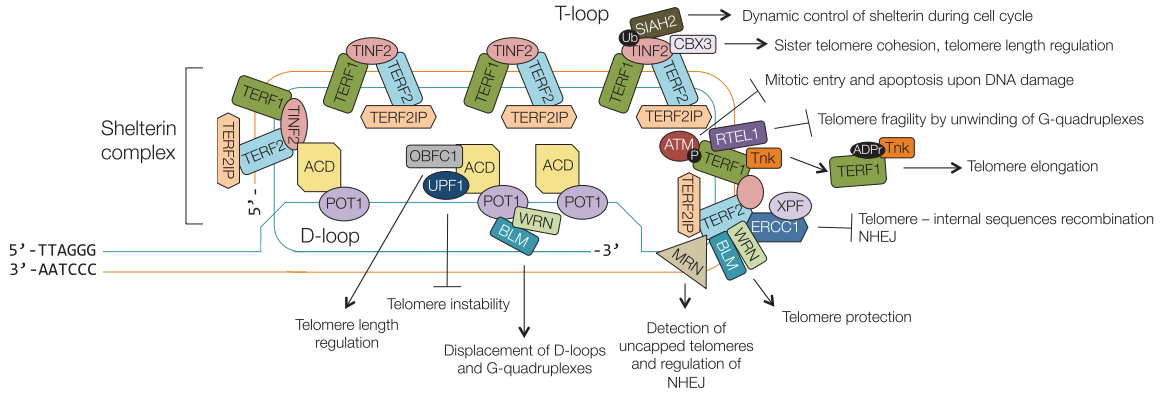

Figure 4.3: **The shelterin complex and its biological functions.** The shelterin complex, shown at the left of the figure, binds telomeres to protect them and regulate diverse signalling cascades. Chromosome ends can fold into T-loops and D-loops, which form when the 3' ssDNA protrusion invades the double-stranded telomere. Signalling cascades that shelterin regulates, and their biological effects, are shown at the right. SIAH2: Siah E3 ubiquitin protein ligase 2, CBX3: Chromobox homolog 3, RTEL1: Regulator of telomere elongation helicase 1, Tnk: Tankyrase, XPF: ERCC4, excision repair cross-complementation group 4, WRN: Werner syndrome, RecQ helicase-like, BLM: Bloom syndrome, RecQ helicase-like, MRN: MRE11-RAD50-NBN complex, OBFC1: oligonucleotide/oligosaccharide-binding fold containing 1, Ub: Ubiquitination, ADPr: ADP ribosylation, NHEJ: Non-homologous end joining.

[444] (Fig. 4.3).

The other components of shelterin are telomeric repeat-binding factors 1 and 2 (TERF1 and TERF2, also known as TRF1 and TRF2), TERF1-interacting protein 2 (TINF2, also known as TIN2) and TERF2-interacting protein 1 (TERF2IP, also known as RAP1) [423]. Although these proteins are fast-evolving, with the architecture of the complex being different in organisms such as ciliates and yeast when compared to mammals, the overall functionality of shelterin is highly conserved [445, 446].

TERF1 and TERF2 are double-stranded DNA-binding proteins that recognise telomeric repeats with high affinity upon homodimerisation, while POT1, the most evolutionarily conserved member of shelterin, can specifically recognise telomeric ssDNA [423, 447–449]. Therefore, the presence of several TTAGGG-binding domains in the complex gives shelterin its exquisite specificity for telomeric sequence. ACD binds to POT1, recruiting POT1 to telomeres and enhancing its recognition ability, whereas TERF2IP localises to telomeres via its interaction with TERF2 [423, 450]. TINF2 is able to bind TERF1, TERF2 and the ACD/POT1 sub-complex, therefore bringing all

the shelterin components together [451, 452].

The importance of the shelterin complex is evidenced by the fact that null mutations in the majority of its components result in embryonic lethality in mice [453–456]. One of the most important functions of shelterin is to protect chromosome ends from DNA repair nucleases, which it achieves by inhibiting six DNA damage signalling pathways: ATM- and ATR-signalling, classical non-homologous end joining (NHEJ), alternative NHEJ, homologous recombination, and resection [457]. Shelterin also has an important role in regulating telomere structure and length. TERF1 and TERF2 have DNA remodelling activities, being able to bend DNA and contributing to T-loop formation, whereas POT1 is able to regulate telomere length by contributing to the nucleolytic processing of the telomeric 5' end and control telomerase access to the end of chromosomes [423, 447, 458, 459]. The absence of functional shelterin therefore leads to polyploidization, fragile telomeres and sister chromatid exchanges, among other chromosomal aberrations [457].

In particular, POT1 can function both as a negative and a positive regulator of telomere length. In support of the former role, its knockdown has been shown to elicit telomere lengthening in HTC75 cells [424], and so does a mutant form lacking the DNA-binding domain [460], or the ACD-binding domain [461]. Additionally, point mutations affecting the ability of POT1 to bind to ssDNA in CLL cells were shown to lead to telomere lengthening, as well as various types of chromosomal aberrations [462]. In support of the latter role, POT1 has been found to enhance the processivity of telomerase when found in a complex with ACD [463], and its exogenous expression in telomerase-positive HT1080 cells led to telomere elongation, albeit with some variability [464]. In order to explain these seemingly contradictory results, Feng Wang and colleagues proposed a three-state model for the role of POT1 at telomeres [463]: First, when POT1 is bound to the 3' overhang as part of the shelterin complex, it blocks access to telomerase by sequestering the telomere, thus acting as a negative regulator of telomere length. Second, the POT1-ACD complex is removed from chromosome ends, possibly by disruption of the shelterin complex. Third, the POT1-ACD complex is recruited to telomeres to serve as a telomerase processivity factor. The cycle is restarted as the telomere is elongated, as it can bind shelterin complexes and the 3' overhang can be re-bound by the POT1-ACD complex. POT1 depends on ACD for its recruitment to telomeres, as *ACD* knockdown has been shown to reduce POT1 localisation to the ends of chromosomes and to promote telomere dysfunction [445].

Additionally, knockdown of *POT1* in tumour cells has been shown to elicit a transient DNA damage response and to alter the sequence of chromosome termini, thus also

determining telomere structure [459]. In this study, tumour cells depleted for POT1 displayed telomere dysfunction, yet they could proliferate as well as control cells without the *POT1* knockdown vector. In contrast, human primary fibroblasts responded to POT1 depletion by inducing senescence and reducing cell division, but proliferation was partially restored upon abrogation of the P53 and the INK4A/RB1 pathways. This observation is interesting, as it could mean that additional mutations are necessary in order for POT1-deficient cells to progress to tumourigenesis. Several studies have also shown that a single *POT1* allele that is unable to bind ssDNA is enough to exert a dominant-negative effect, indicating that abrogation of a single copy of the gene is sufficient to inhibit POT1 function [460, 462, 465]. This effect might be due to disrupted interactions at the telomere, as TERF2 has reduced association with its binding partners in the shelterin complex upon mutant POT1 expression [465].

In conclusion, POT1, ACD and the other shelterin components play an essential role in telomere length regulation and protection, and their malfunction might predispose cells to malignancy via telomere uncapping and length dysregulation, which then could lead to chromosomal abnormalities. In the next subsection, I describe the studies we performed to investigate the biological effect of the *POT1* and *ACD* variants detected in melanoma pedigrees.

4.2.2 The effect of *POT1* variants found in melanoma pedigrees

In total, we identified five pedigrees carrying different *POT1* variants, two predicted to affect mRNA splicing and three missense (Fig. 3.4, Table 3.4). We performed different analyses to study the function of these two types of mutations.

4.2.2.1 Investigation of the variants predicted to affect mRNA splicing

There were two variants predicted to affect *POT1* mRNA splicing detected in familial melanoma pedigrees, one falling in an intronic splice acceptor site and another one falling 6bp away from a splice donor site (Fig. 3.4, Table 3.4). In order to computationally predict whether these variants were likely to be deleterious, I used the MaxEntScan algorithm [466] to compare the sequences of the variant and wild-type splice sites. This algorithm was shown to be able to accurately discriminate real splice acceptor and donor sites from decoys, and works by supplying only a short sequence at the intron-exon boundary. It then assigns a score reflecting the likelihood that the sequence functions as a splice donor or acceptor. The scores for the region with the splice acceptor variant

were 5.53 (wild-type) and -3.32 (with variant), whereas for the splice donor region these were 7.96 (wild-type) and 4.7 (with variant).

To put these scores in context, I retrieved 10,000 sequences at random from the human genome, using the Ensembl API release 70, and obtained their MaxEntScan scores. Subsequently, I retrieved 10,000 real splice acceptor sites and 10,000 real splice donors from human genes chosen at random, but always choosing the second exon. I then obtained a distribution of all of their MaxEntScan scores. The splice acceptor variant lowered the score of the wild-type sequence from the 9.2th to the 0.57th percentile when compared to the score distribution of real splice acceptors, while the splice donor variant lowered it from the 33.67th to the 12.31th percentile when compared to the score distribution of real splice donors (Fig. 4.4). Therefore, the splice acceptor variant is predicted to be highly deleterious, whereas the splice donor variant still falls well within the scores of real splice sites.

We then decided to test whether these variants had any effect in *POT1* mRNA splicing. To do this, RNA extracted from the whole blood of two carriers of the splice acceptor variant (pedigree AF1) was converted to cDNA using SuperScript III Reverse Transcriptase (Invitrogen). RT-PCR was then performed to confirm that this variant was indeed disruptive to splicing. M13-tagged forward and reverse primers were designed to flank the spliced region. The product was visualised on a 3% NuSieve Agarose gel, and the sequence was verified using standard Sanger sequencing methods. This work was done by Lauren G. Aoude, at QIMR Berghofer. This experiment showed that this splice acceptor variant affects transcript splicing, as it leads to a frameshift and the introduction of a premature stop codon 11 amino acids downstream of the intron-exon boundary (Fig. 4.5).

As we failed to see any effect of the splice donor variant on transcript splicing by PCR, we decided to perform whole blood RNA sequencing from the carrier to see whether we could find traces of any aberrant *POT1* transcripts. However, we could not detect anything abnormal (data not shown). The RNA extraction was performed by Marcela Sjöberg and the transcriptome analysis was done by Martin del Castillo, at Sanger.

These analyses support that the splice acceptor variant carried by pedigree AF1 is disruptive to the *POT1* transcript, causing a frameshift and introducing a premature stop codon. However, we did not find evidence to support that the splice donor variant is deleterious to the carrier.

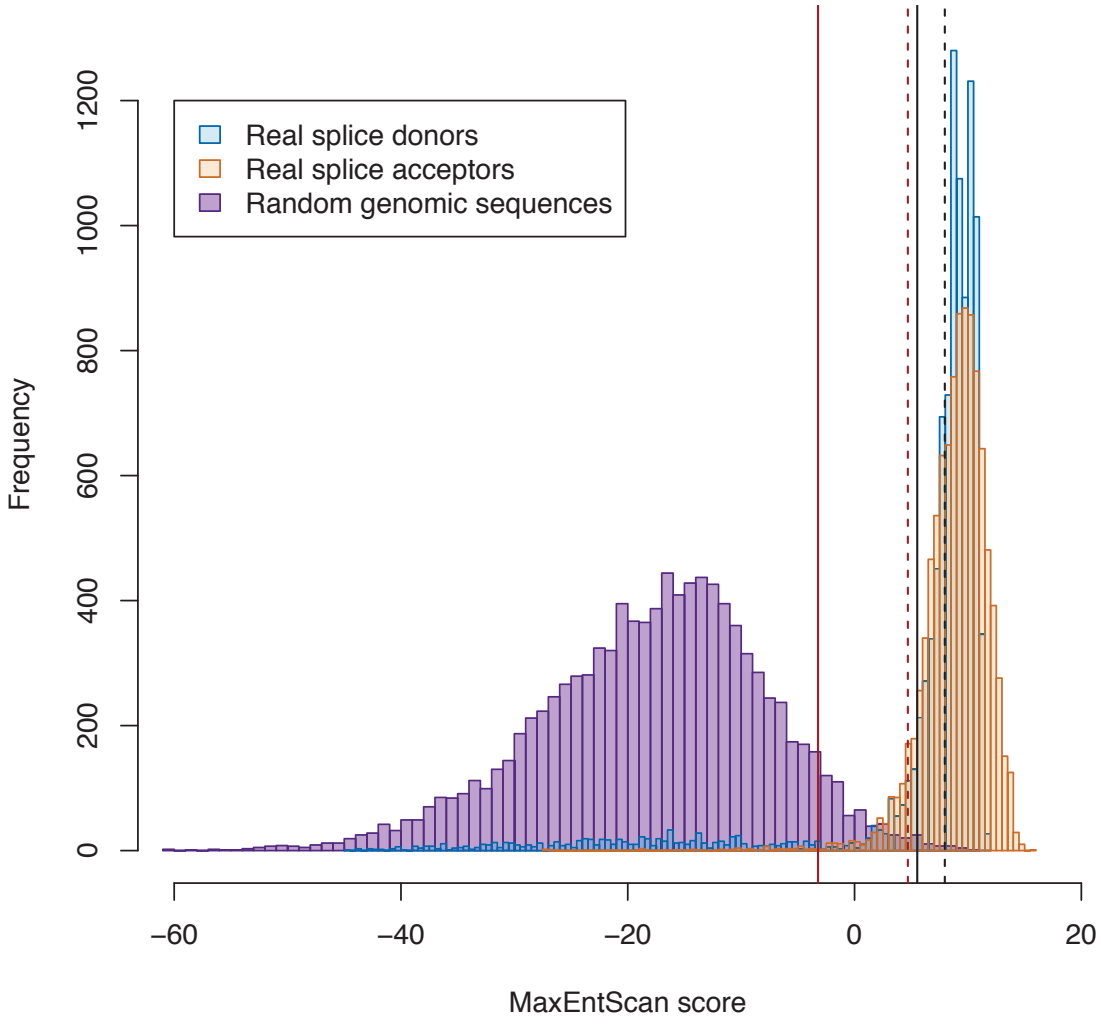

Figure 4.4: **MaxEntScan scores for the splice acceptor and splice donor variants detected in familial melanoma pedigrees.** The location of the scores for the wild-type (black) and variant (red) sequences for the splice acceptor (solid) and splice donor (dotted) sequences are shown against score distributions for real splice donors, real splice acceptors and random genomic sequences.

4.2.2.2 Investigation of the three missense variants detected in melanoma pedigrees

4.2.2.2.1 Evolutionary conservation of variant positions In order to assess the evolutionary conservation of the residues where we detected variants in POT1, I gathered amino acid sequences for the encoded protein in evolutionarily diverse species (*i.e.*, human, mouse, cow, armadillo, elephant, opossum, platypus, chicken, frog and zebrafish) from the National Centre for Biotechnology Information (NCBI) and aligned

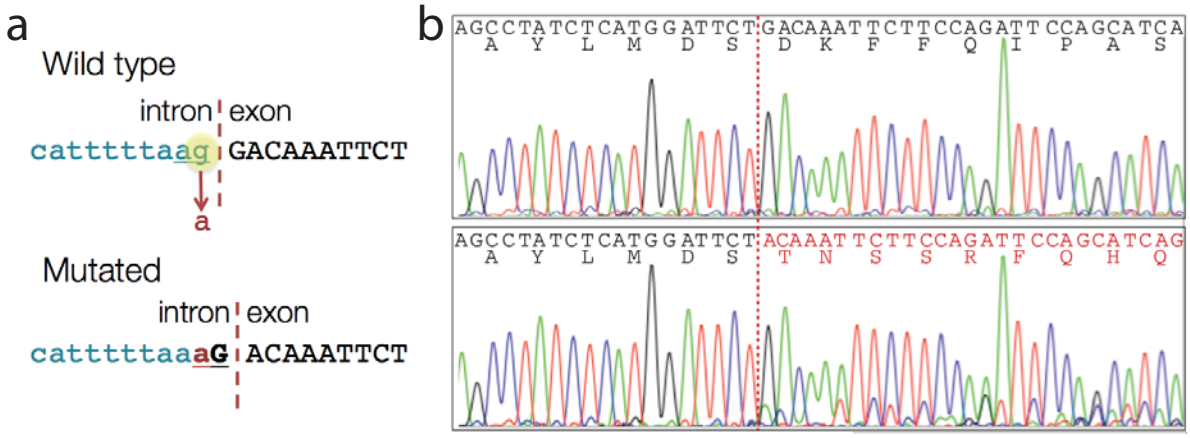

Figure 4.5: **The *POT1* splice acceptor variant affects transcript splicing.** a) Rationale for the effect of the splice acceptor variant. The splice acceptor signal sequence (AG) is underlined, and the variant is indicated. This variant is predicted to cause a one-base shift of the splice acceptor signal. b) 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. 3.2) (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.

them using Clustal Omega [467]. All of the missense variants we identified in *POT1* disrupt amino acids that are completely conserved through eutherians (Fig. 4.6).

To estimate the number of substitutions per site in this amino acid alignment, I used the ProtPars routine from the PHYLIP package of programmes [469]. This routine is able to infer an unrooted phylogeny from the sequences supplied by counting the number of mutations required to span all the amino acids observed at each site of the alignment, while being consistent with the genetic code. Synonymous variants are omitted from the final count, as it is assumed that they are not under selection. As such, it is capable of estimating the number of substitutions, at the DNA level, that occurred in the amino acid alignment shown in Fig. 4.6. 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 folds (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 we only consider sequences from eutherian organisms, then no substitutions have occurred in any of these three residues, compared to 0.8 substitutions per site in the OB folds and 1.39 substitutions

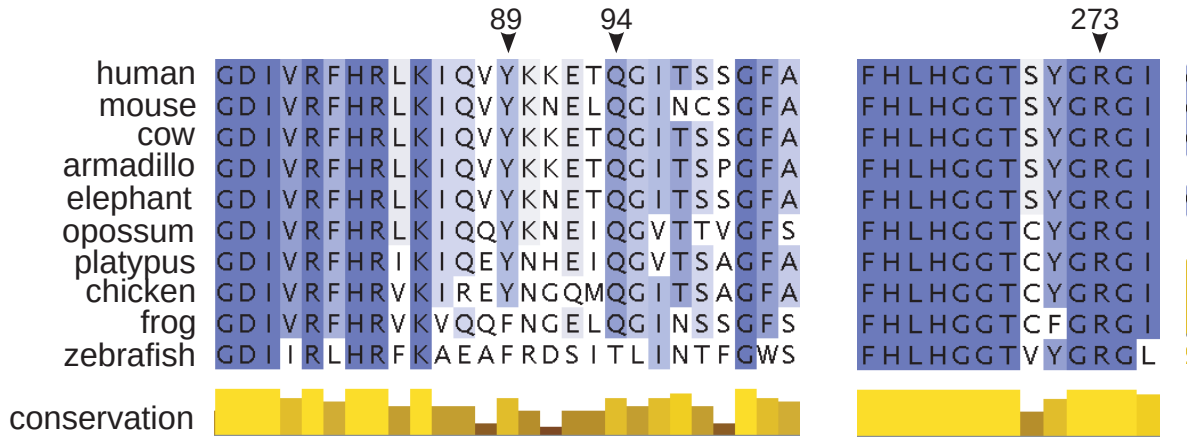

Figure 4.6: **Highly conserved residues of POT1 are altered in familial melanoma.** Shown are the positions of the missense variants on a multi-species amino acid alignment. More conserved amino acids through evolution are shown in a darker colour. The alignment is displayed using Jalview v2.7 [468].

per site in the whole protein. I defined the OB fold regions as amino acids 8-299 in the human sequence, according to the annotation in SUPERFAMILY [419].

4.2.2.2.2 Structural modelling and characterisation of *POT1* variants Having corroborated that the positions where these variants lie are highly conserved, we then proceeded to structurally characterise the residues in which we find germline variants. To do this, we examined the structure of the POT1 protein bound to a telomere-like polynucleotide (dTUdAdGdGdGdTdTdAdG), obtained from the Protein Data Bank (PDB), id: 3KJP [470, 471]. According to this model, all three altered residues (Y89, Q94 and R273) were among 24 residues located in close proximity (<3.5 Ångströms) to the telomeric polynucleotide [462] (Fig. 4.7a). R273 interacts with the oxygen at position 2 of telomeric deoxythymidine 7, whereas Q94 and Y89 both interact with the G deoxynucleotide at position 4. Remarkably, the POT1 codon for Q94 has been found to be a target for recurrent somatic alteration (Q94R) in CLL, where ~5% of cases carry POT1 mutations that cluster in the sequences encoding the OB folds [462]. Therefore, the POT1 variants we identified are expected to weaken or abolish the interaction of POT1 with telomeres.

The list of 24 residues located in close proximity to the telomeric polynucleotide, based on this crystal structure, was defined by Andrew J. Ramsay *et al.* in a previous publication [462]. These residues are 31, 33, 36, 39–42, 48, 60, 62, 87, 89, 94, 159,

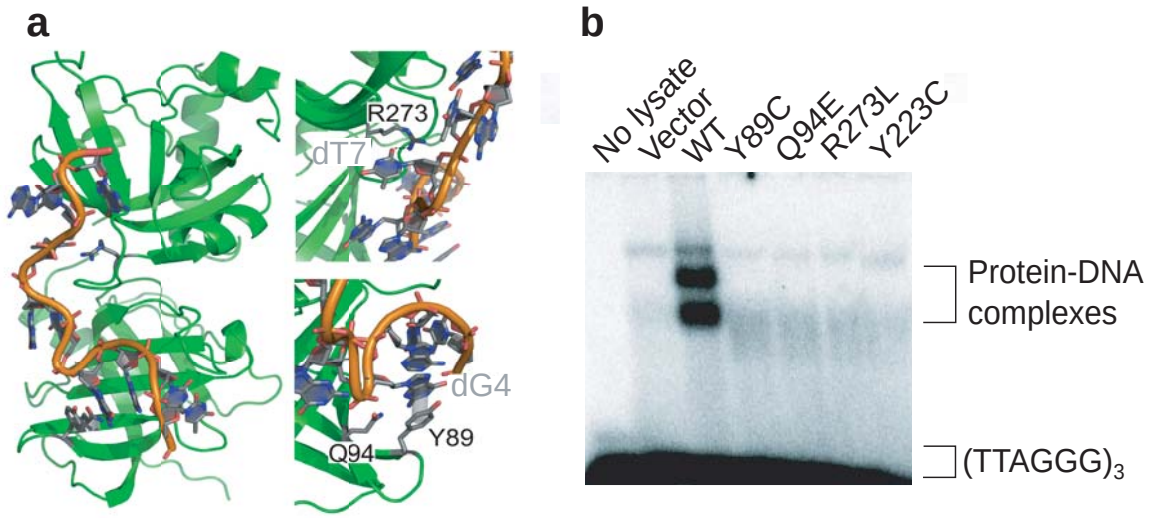

Figure 4.7: **Missense variants in *POT1* disrupt the interaction between POT1 and single-stranded DNA.** a) Shown are the locations of the POT1 Y89, Q94 and R273 residues in the two N-terminal OB folds (green). A telomere-like polynucleotide sequence is shown in orange. Interacting nucleotides in the telomeric sequence are labeled in gray. All three substitutions are predicted to disrupt the association of POT1 with telomeres. This image was generated by Víctor Quesada, at the University of Oviedo, Spain, and was rendered with PyMOL v0.99 [472]. b) Mutant Y89C, Q94E and R273L POT1 proteins are unable to bind telomeric $(TTAGGG)_3$ sequences as shown by an electrophoretic mobility shift assay. The Y223C POT1 mutant was used as a positive control representing a known disruptive alteration [462].

161, 223, 224, 243, 245, 266, 267, 270, 271 and 273. In order to assess the statistical significance of finding amino acid substitutions affecting these residues in the population, I searched 6,503 exomes released by the NHLBI GO ESP [17] 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 Table 4.2. In summary, a minimum of 6,498 exomes had all bases covered at a minimum average coverage of 59×. The variant encoding N224D was found at an overall allele frequency of 1 in 13,005. No other amino acid–changing variants were found. I then compared the number of variants found in these 24 OB domain residues in controls (1 in 6,498) to the number of variants found in all analysed pedigrees (3 in 105), obtaining a P-value of 1.54×10^{-5} using a two-tailed Fisher's exact test. This comparison indicates that our melanoma cohort is enriched for rare variants in DNA-interacting residues of POT1 when compared to the population.

The *in silico* analyses above point to these variants being deleterious, as not only are

Table 4.2: **Genomic location of bases encoding the 24 OB domain residues in close proximity to telomeres**

Amino acid	Genomic position (GRCh37)
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

they predicted to be so by bioinformatic algorithms (Table 3.4), but variants in amino acids that participate in protein-DNA interactions seem to be extremely rare. The positions at which these variants are found are also highly conserved through evolution, and the variants themselves have the potential to impair POT1 binding to telomeres. Therefore, we decided to formally test whether the variants we identified disrupt POT1 function.

4.2.2.2.3 *In vitro* translation and G strand binding assays To test the effect of these variants on POT1 function, we assessed the ability of *in vitro*-translated POT1 proteins carrying the Y89C, Q94E and R273L variants to bind to $(TTAGGG)_3$ sequences. Human *POT1* in a T7 expression vector (Origene) was mutated by site-

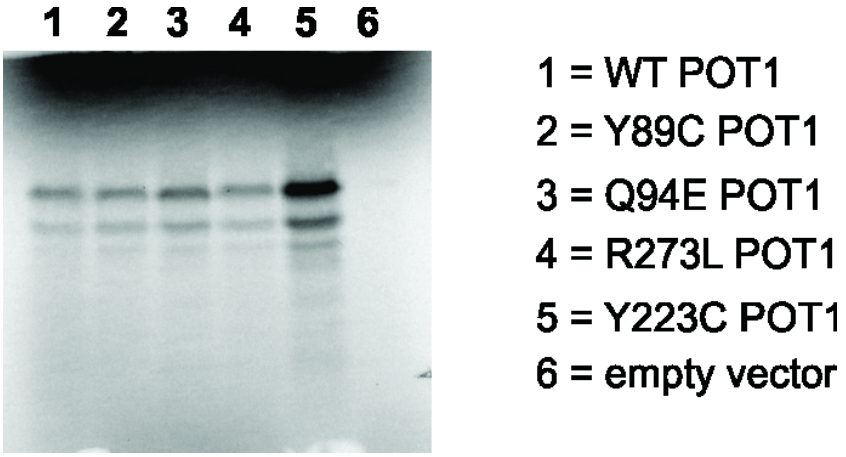

Figure 4.8: **^{35}S gel showing the *in vitro* translation products of *POT1* wild-type (WT) and OB domain mutants.** This gel confirms that each *in vitro* translation reaction successfully produced protein for the electrophoretic mobility shift assay shown in Fig. 4.7b. The Y223C variant was somatically acquired in CLL and has previously been shown to be unable to bind to telomeric DNA [462]. The DNA-protein complexes shown in 4.7b were visualised by [^{32}P]-labeling of $(TTAGGG)_3$ ssDNA.

directed mutagenesis to generate cDNAs encoding the POT1 Y89C, Q94E and R273L 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. A 5-µl fraction of each reaction was analysed by SDS-PAGE; proteins were visualised and relative amounts were quantified using the FLA 7000 phosphorimager system (Fujifilm) (Fig. 4.8). DNA binding assays were performed as described previously with minor modifications [473]. Protein-DNA complexes were analysed by electrophoresis on a 6% polyacrylamide Tris-borate-EDTA gel run at 80 V for 3 h. Gels were visualised by exposure to a phosphorimager screen (Fig. 4.7b). DNA mutagenesis, *in vitro* translation reactions and DNA binding assays were performed by Andrew J. Ramsay, at the University of Oviedo, Spain.

Electrophoretic mobility shift assays showed a complete abolition of POT1-DNA complex formation with mutant POT1 (Fig. 4.7b). Notably, other similar variants that abolish the binding of POT1 with telomeres have been described to be somatically acquired in CLL [462]. The variants found in CLL, in particular Y36N and Y223C, promote uncapping of telomeres, telomere length extension and chromosomal aberrations and thereby promote tumourigenesis.

4.2.2.2.4 Analysis of telomere length in *POT1* variant carriers Given these observations, and the important 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. Therefore, using exome sequence from the 41 cases that belong to the discovery phase (in which the three members of UF20 were sequenced), telomere length of each subject was estimated from NGS data. The algorithm used, called TelSeq and written by Zhihao Ding at the Sanger [474], estimates telomere length by counting TTAGGG sequences in unmapped reads in NGS data, and takes into account factors such as GC composition and read length in order to give an accurate estimation. The algorithm was shown to correlate with Southern blot measurements of telomere lengths, so it is a useful tool to compare telomere lengths in melanoma patients with and without *POT1* variants. Only the discovery phase cohort could be used because the exomes sequenced by BGI, to which the Q94 and R273 carriers belong, did not contain any unmapped reads and thus could not be assessed by TelSeq. The telomere length measurement was performed by Zhihao Ding, who was blinded to the *POT1* status of all samples.

After calculation of relative telomere length, I adjusted the 38 samples without germline *POT1* variants for age at blood draw and sex using a linear model (Fig. 4.9). I then estimated the corresponding values for *POT1* variant carriers on the basis of the same adjustment. I did these calculations with a custom R script, using the Hmisc library [475].

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*. I performed a Wilcoxon rank-sum test comparing the telomere lengths of the three Y89C cases to that for the 38 non-carrier controls (P-value <0.0002, Fig. 4.10a).

Because not all samples with missense variants could be included in the bioinformatic measurement, we decided to measure telomere lengths by PCR in all missense variant carriers, as well as other members of their families (both carriers and non carriers) and a panel of melanoma cases that did not carry *POT1* variants at these positions. These individuals had been genotyped for the detected variants before (see Subsection 3.4.1). Therefore, in this measurement, we included cases from the Leeds Melanoma case-control study, 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, Fig. 3.2). Relative mean telomere length was ascertained by SYBR Green RT-PCR using a version of the published

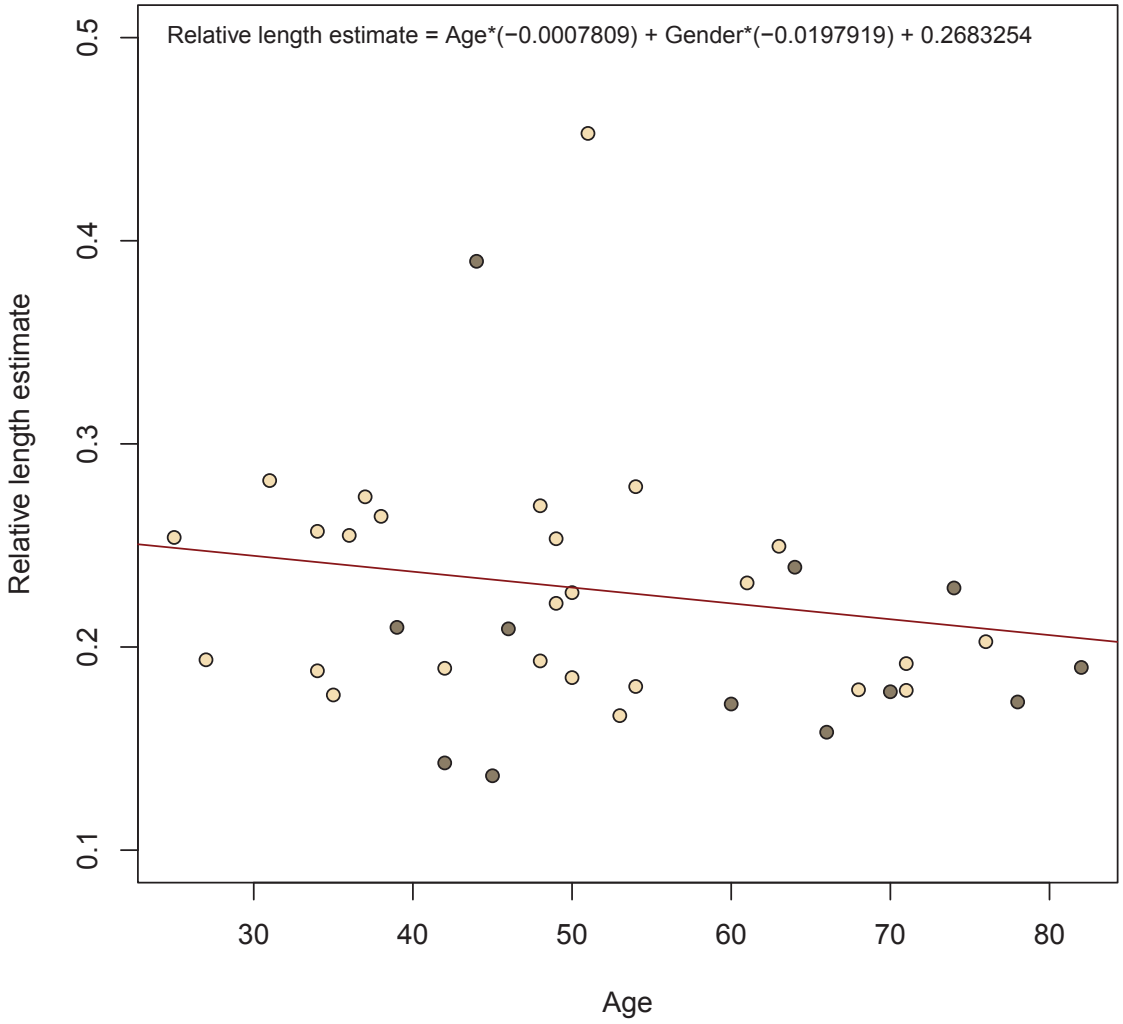

Figure 4.9: **Linear model used to adjust bioinformatically calculated telomere lengths for age at blood draw and sex.** Residuals were used as the adjusted relative telomere lengths. Dark circles represent male samples, and light circles represent females. The sex variable is coded as 0=female, 1=male. Note that only two dimensions (relative telomere length and age) are shown.

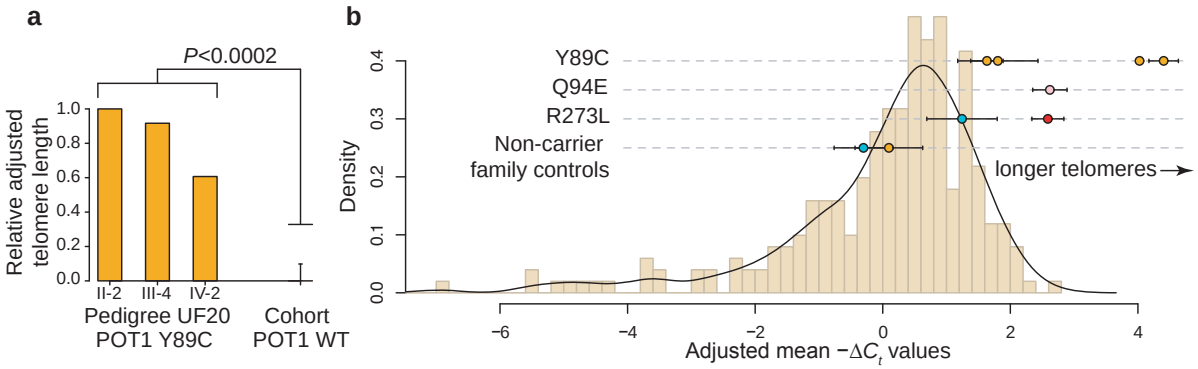

Figure 4.10: **Missense variants in *POT1* lead to elongated telomeres.** a) Calculation of telomere length from exome sequence data. Relative adjusted telomere lengths for the three sequenced members of pedigree UF20 are shown alongside the mean telomere length of 38 (all other) melanoma cases that were sequenced alongside them but were wild type for *POT1*. All values are shown relative to the largest sample measurement. Error bars, 1 s.d. b) PCR-based estimate of telomere length. Adjusted mean -ΔC_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. 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 [476] for bandwidth smoothing. Orange dots, members of pedigree UF20; pink dots, members of pedigree UF31; blue dots, members of pedigree UF23; red dots, individual from the Leeds Melanoma case control study carrying the R273L variant. 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, standard error of the mean.

quantitative PCR protocols [477, 478] that was modified as described previously [479]. 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, $\Delta C_t = C_t\text{TEL} - C_t\text{CON}$, was the measure of telomere length used in the analysis, as in other published data generated using this assay [479, 480].

For the analysis, samples with $C_t\text{CON} < 18$, $C_t\text{CON} > 27$ or $C_t\text{CON} > 2$ 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. The telomere length estimation by PCR was performed by Karen A. Pooley and Alison M. Dunning, at the University of Cambridge, UK, who were blinded to *POT1* status of all samples.

I then estimated mean ΔC_t values for each sample from all replicates. I then adjusted the estimated mean values of ΔC_t obtained from melanoma cases without germline *POT1* variants for age at blood draw and sex using a linear model (Fig. 4.11). The corresponding values for *POT1* variant carriers were estimated on the basis of the same adjustment. In Fig. 4.10b, adjusted mean ΔC_t values are plotted with the histogram showing the non-carrier melanoma cases compared to the missense variant carriers and the non-carrier family controls plotted above. I then performed 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, yielding a P-value of 3.62×10^{-5}.

Both the bioinformatic estimation and the PCR measurements of telomere lengths agree that all *POT1* variant carriers have significantly longer telomeres than individuals without these variants, even when compared to members of the same family (Fig. 4.10b). Thus, missense variants in the OB domains of POT1 not only abolish telomere binding, but are also associated with increased telomere length. Interestingly, the melanoma case in pedigree UF23, who does not carry the R273L variant, has shorter telomeres than her carrier relative, with length comparable to non-carrier cases or the disease-free individual from pedigree UF20. I elaborate on this observation, as well as the potential tumourigenesis mechanisms of *POT1* variants, in the Discussion.

The difference in telomere lengths for *POT1* missense variant carriers when compared to controls made us question whether there would be any effect in telomere length for the splice variant carriers. Unfortunately, both the bioinformatic algorithm and the PCR measurements are extremely sensitive to variations in DNA extraction methods, sequencing methodology and the plate where the sample resides (see Subsection A.1.10). These facts did not allow us to assess samples from the Australian cohort in this analysis. However, the sample carrying the *POT1* splice donor variant showed longer telomeres when compared to samples from the Leeds case-control study when analysed by PCR, suggesting that that this variant might have some effect on telomere regulation (Fig. 4.12).

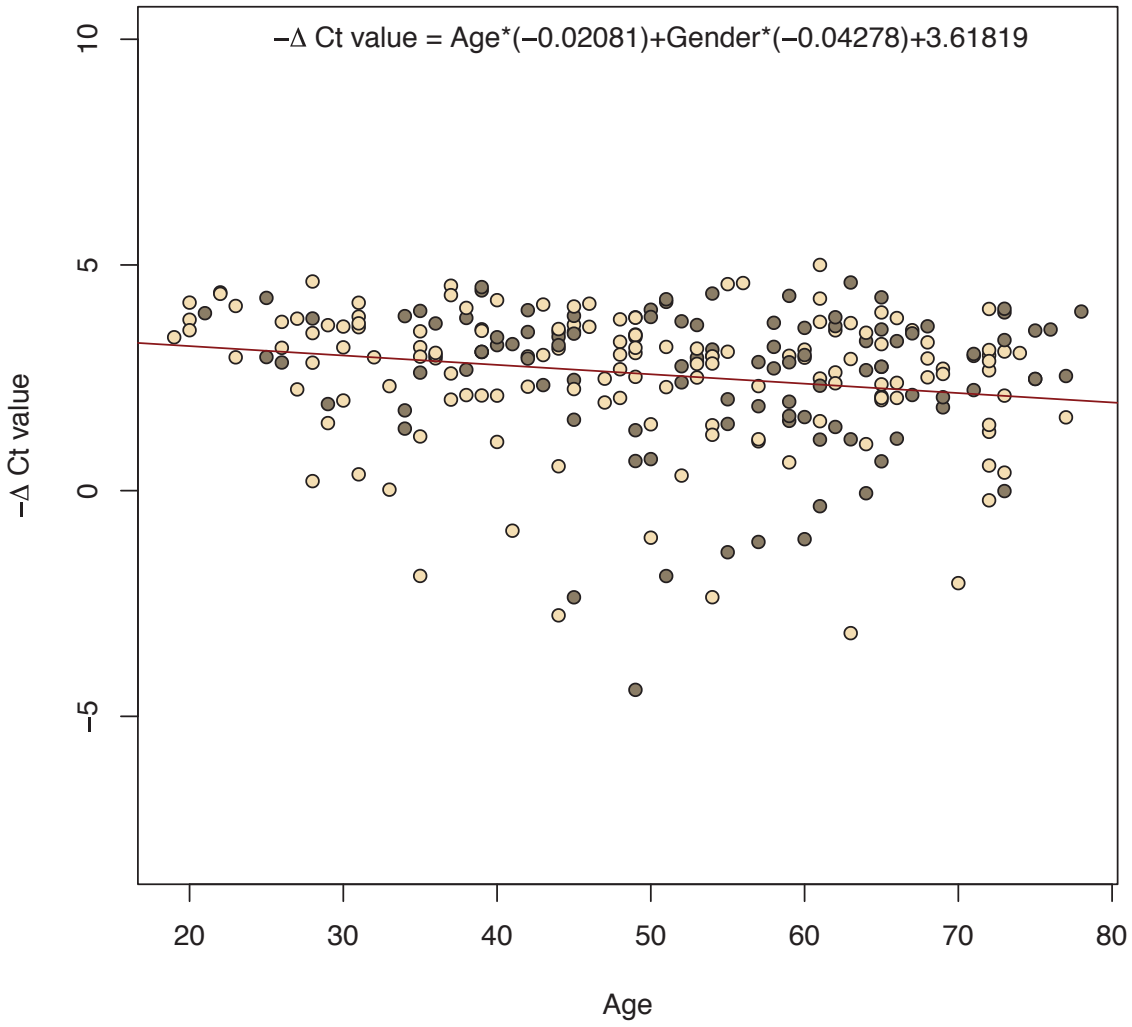

Figure 4.11: **Linear model used to adjust PCR mean ΔC_t values for age at blood draw and sex.** Residuals were used as the adjusted mean ΔC_t values. Dark circles represent male samples, and light circles represent females. The sex variable is coded as 0=female, 1=male. Note that only two dimensions (mean ΔC_t value and age) are shown.

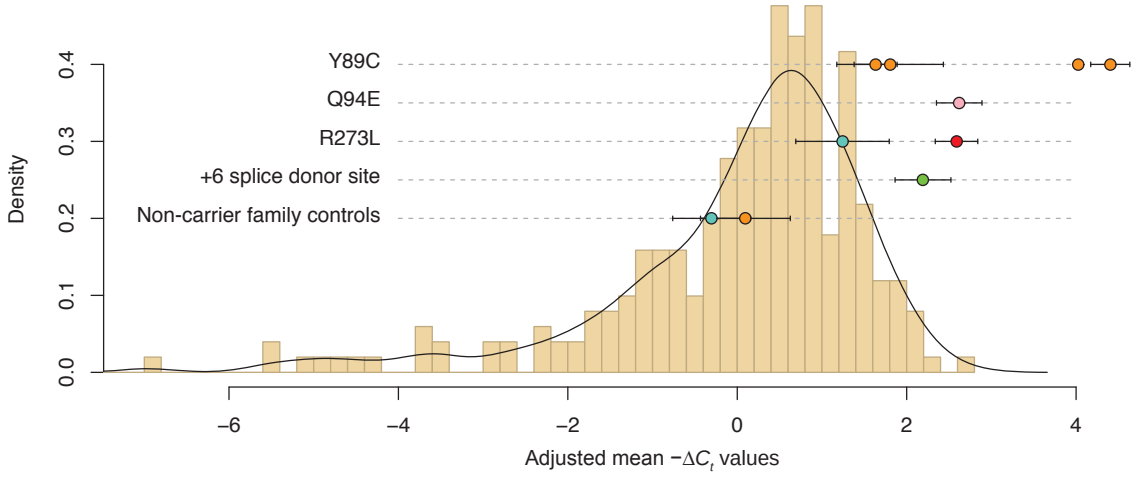

Figure 4.12: **PCR-based estimate of telomere length showing the carrier of a *POT1* intronic splice donor variant.** The carrier of the intronic splice donor variant is shown in the fourth row, in green. The rest of the figure is identical to Fig. 4.10b.

4.2.2.2.5 Investigating the role of *POT1* mutations in other cancers The identification of *POT1* mutations in CLL and the probable susceptibility of our *POT1*-mutated familial melanoma pedigrees to other tumour types (Fig. 3.2) suggest that *POT1* might have a more general role in tumourigenesis. To investigate this possibility, we examined pan-cancer data from the COSMIC [438] and IntOGen [481] databases (data from The Cancer Genome Atlas [TCGA] and the International Cancer Genome Consortium [ICGC]).

To statistically assess the mutational patterns affecting *POT1* in cancer, the COSMIC database, version 66, was mined 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. The total frequency of each reference/mutated base pair in the same COSMIC database was compiled. Finally, a Monte Carlo simulation with 100,000 groups of 35 mutations at random locations in the *POT1* ORF was performed. The probability of a given mutation from a reference base (*e.g.*, 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-value<0.03). To assess the clustering of mutations at sites encoding DNA-interacting residues (Table 4.2), only missense mutations were considered, as no selection would be expected for nonsense mutations. In the COSMIC database, 27 *POT1* missense mutations were found, 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-value<0.02). The Monte Carlo simulations were performed by Víctor Quesada, at the University of Oviedo, Spain.

Finally, to assess the functional impact bias of somatic mutations in *POT1*, I also considered all mutations in *POT1* that are present in the IntOGen database [481]. 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 [481], calculated with Oncodrive-fm [482], were combined to yield a P-value of 0.021, indicating that this gene is biased toward the accumulation of functional mutations.

Therefore, although mutations in *POT1* have not been found at a high frequency in the cancer studies deposited in COSMIC and IntOGen (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 (P-value<0.03), to alter residues in close proximity to DNA (P-value<0.02), and to have a higher functional impact bias (P-value<0.03) than expected by chance. These results suggest that, although rare, somatic *POT1* mutations may drive tumourigenesis across multiple histologies.

4.2.2.3 Conclusion

We identified five different variants in *POT1* in melanoma pedigrees, three that alter amino acid residues, one predicted to affect splicing and one of uncertain significance. In this part of the study, we investigated the mechanisms by which *POT1* variant carriers might be predisposed to the development of melanoma, and possibly other cancers (Fig. 3.2).

The missense variants are found in highly evolutionarily conserved positions, in the

OB folds of *POT1* (Figs. 3.4 and 4.6). We show here that these not only abolish the interaction of POT1 with telomeres *in vitro* (Fig. 4.7b) but also that variant carriers have longer telomeres than melanoma cases without *POT1* variants, and indeed that members of their own families without *POT1* variants (Fig. 4.10). These differences persist when measurements are adjusted by age at blood draw and sex. Additionally, *POT1* variants falling in amino acids in close proximity to DNA seem to be extremely rare, as only one variant allele was found when bases coding for these 24 amino acids were examined in more than 6,500 control exomes. Accordingly, somatically acquired variants in this gene in cancer databases show a tendency to be missense rather than silent, to affect amino acids in close proximity to DNA, and to have a higher functional impact bias than expected by chance.

We also show that one of the two splice variants affects *POT1* mRNA splicing, introducing a frameshift that leads to a premature stop codon (Fig. 4.5). Although we could not show this effect for the other variant, some impact on telomere length was suggested when we examined it by PCR (Fig. 4.12).

Given the essential role of POT1 in telomere protection and length regulation, the effect that these variants have on carriers, and the dominant-negative effect that has been described for similar mutations, it is not surprising that individuals with one non-functional allele are predisposed to malignancy. I cover the biological mechanisms by which this might happen in the Discussion.

4.2.3 The effect of *ACD* variants found in melanoma pedigrees

We identified two pedigrees with variants in *ACD*, one introducing a premature stop codon and another affecting an amino acid in the POT1-binding domain (Fig. 3.4 and Table 3.4). The QIMR Berghofer team led an investigation into the effect of these variants in protein associations within the shelterin complex. Because ACD does not bind DNA, an electrophoretic mobility shift assay could not be used. Therefore, in order to analyse protein-protein interactions, our collaborators used Octet RED biolayer interferometry technology, which is able to detect changes in biomolecular interactions between an immobilised receptor attached to a biosensor surface and analytes in solution [483]. Therefore, this method allows the comparison of binding wild-type and mutated proteins to an immobilised binding partner.

Their preliminary data has shown that the ACD Q320X variant results in altered binding to POT1, compared to ACD wild-type protein (Fig. 4.13). They are in the process of generating the ACD N249S protein for testing in a similar manner. This

work is on-going and the kinetics of the interactions, assessed by the same technology, will be performed shortly. The protein-protein interaction assays were performed by Dr. Antonia L. Pritchard, at QIMR Berghofer, whom I thank for supplying methods and figure.

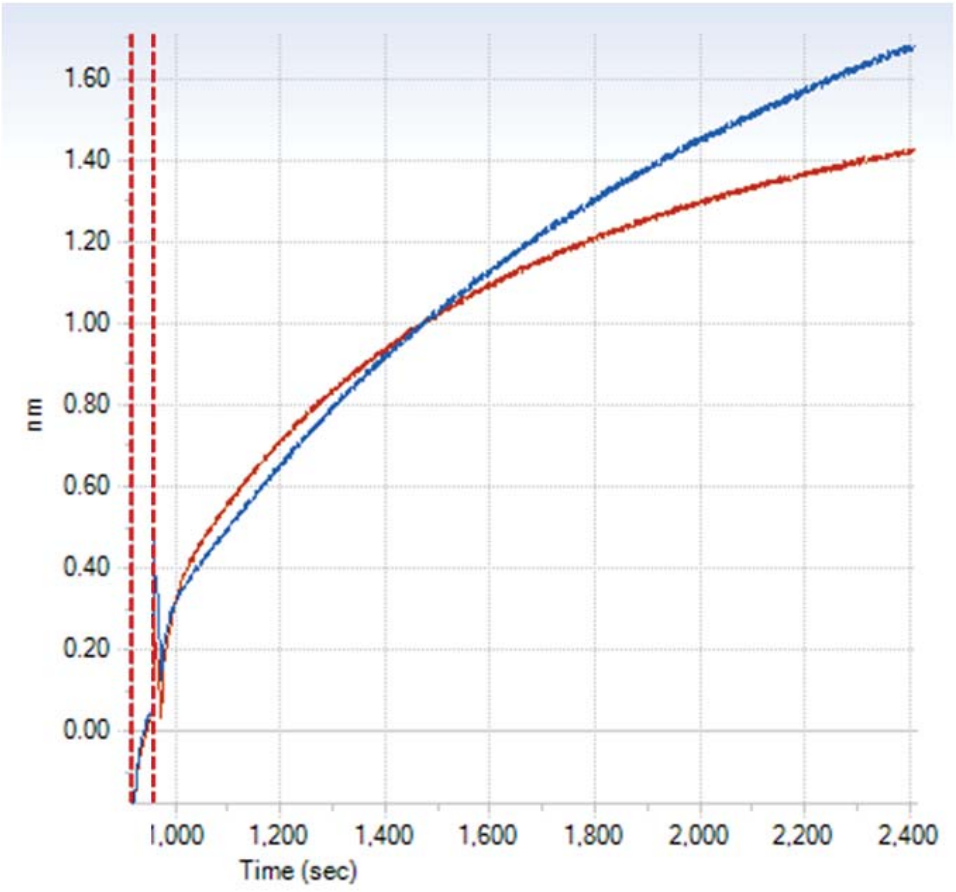

Figure 4.13: **Biolayer interferometry investigation of wild-type and mutated ACD interactions with POT1.** The POT1-ACD binding curve for the ACD wild-type protein is shown in red, whereas the one for the ACD Q320X protein is shown in blue. A lower end point indicates a lower total association, a similar slope and end point indicates a similar association between wild-type and mutated proteins, a higher end point indicates a potentially higher degree of association, while a steeper slope indicates a faster association rate. I thank Dr. Antonia L. Pritchard for this image.

Measurements for wild-type and mutant ACD when binding POT1, although they vary, were not significantly different. Although a cumulative binding effect over a lifetime cannot be discarded, we were unable to draw any conclusions from this experiment. Telomere length measurements were also performed by PCR on the ACD variant carriers, along with additional variants identified in other members in the shelterin complex (see

Subsection 4.2.4), but telomere lengths were not found to differ significantly (data not shown).

4.2.3.1 Conclusion

We identified two variants in *ACD* co-segregating with melanoma in two different pedigrees, one missense and another introducing a premature stop codon. The QIMR Berghofer team performed protein binding assays and telomere length measurements, but could not find any functional evidence for the mechanism by which these variants might be predisposing to melanoma, if indeed they are. However, it is likely that the stop codon is deleterious, as there are no stop codons found throughout the length of the protein in dbSNP release 138 [381] or the 6,500 exomes released by NHLBI GO ESP [17], but there are examples of somatically acquired stop codons in *ACD* in the COSMIC database [438]. This observation evidences that these mutational events are extremely rare in the germline, and, added to the perfect co-segregation with melanoma found in pedigree AF19, might indicate that this variant plays a role in the melanoma predisposition seen in this pedigree. However, more biological studies should be carried out to address the mechanism definitively, and to determine whether these *ACD* variants have any effect on telomere protection.

4.2.4 Other genes in the shelterin complex

Given their biological role in telomere length and protection, the rest of the members in the shelterin complex represent attractive candidate genes in melanoma susceptibility. The QIMR Berghofer team screened 601 individuals belonging to 510 families, from Australia, UK, The Netherlands, Denmark and Sweden, for germline variants in any of the members of the shelterin complex apart from *POT1*. The families tested did not have variants in *CDKN2A*, *BAP1*, *POT1*, *BRCA2*, *CDK4* and the promoter of *TERT*. We found, in addition to the two variants in *ACD* described above, a nonsense variant in *TERF2IP* co-segregating perfectly with melanoma in three members of pedigree UF3, of which one member had been sequenced as part of the discovery phase (Fig. A.1.1 and Table A.1.8).

Other variants they found co-segregating perfectly with melanoma in members of the shelterin complex are a V272M substitution in *ACD* in two members of one pedigree, a Q191R variant in *TERF2IP* in two members of another pedigree, and a M5I substitution in the same gene in one evaluated member of a pedigree, all in Australian families.

Also, they found other variants not co-segregating perfectly in these genes, as well as other genes in the shelterin complex, but these were not convincingly associated with melanoma.

4.3 Summary and conclusion

In this Chapter, I described the studies we performed to assess the biological function by which the variants detected in prioritised genes might be predisposing carriers to melanomagenesis. We were able to establish a role for variants in the shelterin complex member *POT1* in telomere length maintenance, with carriers of the variants having longer telomeres than both other non-carrier members in their families and other melanoma patients without these variants. I discuss the mechanism by which longer, and potentially unprotected, telomeres predispose to malignancy in the next Chapter.

We could not conclusively define the biological mechanism, if any, by which *SMG1* and *ACD* variants might be contributing to tumourigenesis. In the case of *SMG1*, difficulties such as cell culture contamination or assay irreproducibility prevented us from drawing any conclusions from the experiments we performed. In the case of *ACD*, although we could not pinpoint the function that might be altered in variant carriers, its potential involvement in melanoma predisposition is supported by its biological role and the absence of stop codon variants in common genomic variation datasets from thousands of individuals. However, biological assays focusing in other aspects relevant to melanoma, such as the efficiency of DNA repair upon UV damage, the rate of acquisition of chromosomal aberrations, or cell replicative ability should be performed in order to investigate the potential role of these variants.

In the next and final Chapter, I discuss the relevance of the findings presented in this dissertation and I cover the biological mechanisms that might be behind the variants identified in familial melanoma pedigrees.