Chapter 3

Familial melanoma sequencing: Integrative phase

Methods and results of this chapter have been published or are accepted for publication (refs. [376] and [414]). Some parts of the text have been reproduced from these references; I confirm I have ownership of copyright for reproduction in this work.

While I was studying the European samples discussed in Chapter 2, a large dataset of multi-case Australian samples became available for analysis, which more than doubled the initial dataset. This Chapter describes the clinical characteristics of the pedigrees and samples in this new dataset and explores the integrative analysis rationale we followed, as well as the potential melanoma susceptibility candidates that were uncovered. An overview of the steps followed in this phase is depicted in Fig. 3.1.

3.1 Patient selection

Australia is the country with the highest melanoma incidence in the world [369]. For this reason, criteria for the collection of melanoma families in the UK and Australia vary, as melanoma risk factors such as the presence of atypical naevi are much more common in the Australian population [415]. This suggests that sunlight-induced phenocopies occurring in melanoma-predisposed families might confound the search for high-penetrance susceptibility genes.

To account for this factor, the great majority of pedigrees sequenced for this phase had five or more melanoma cases, which reduces the probability that the aggregation of cases observed within families was due to clustering of sporadic cases. The additional

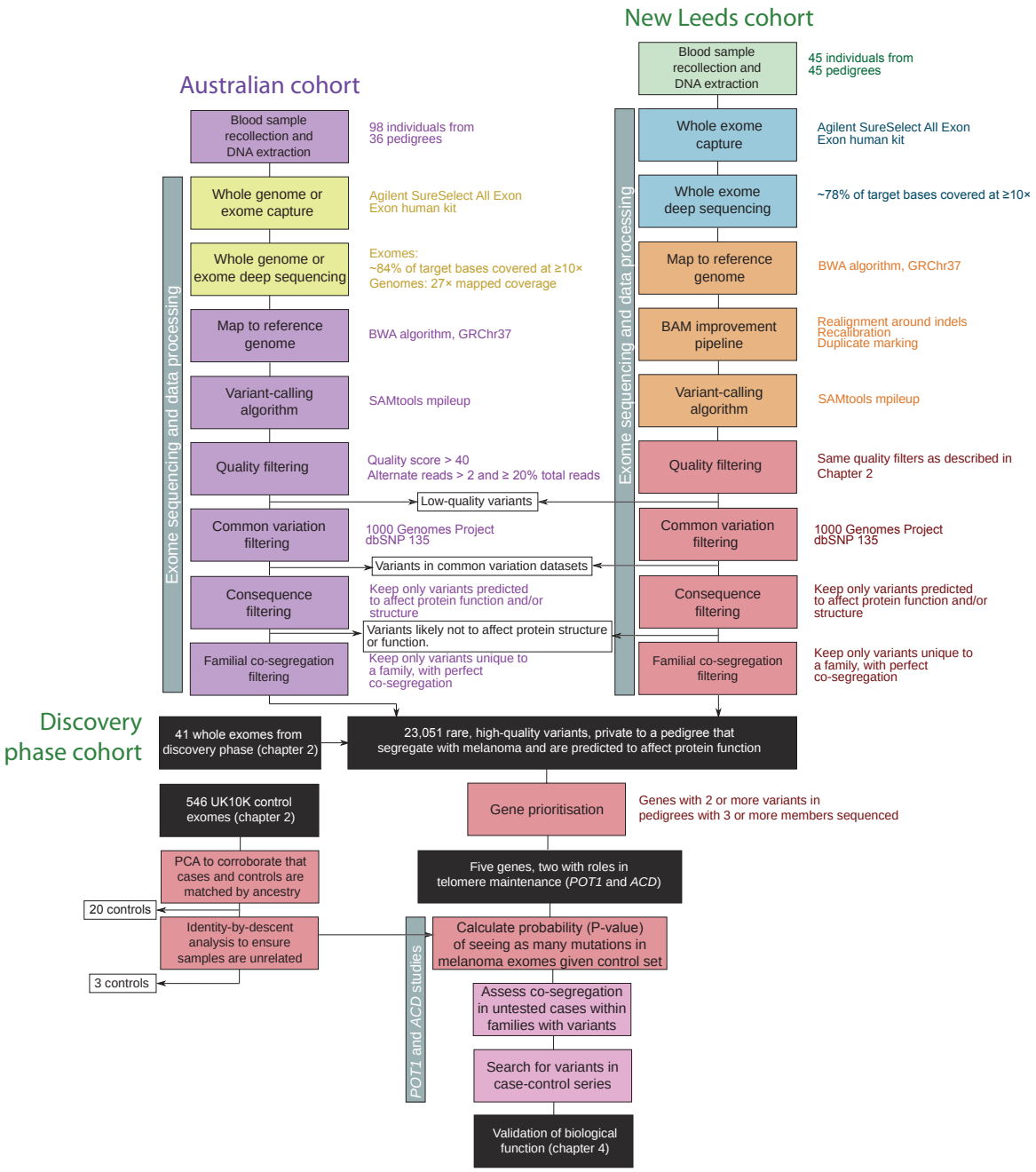

Figure 3.1: **Flowchart of analysis steps followed in the search for melanoma susceptibility genes, integrative phase.** Steps are colour-coded depending on the place where these were done, purple: QIMR Berghofer, yellow: Macrogen, Inc., green: Leeds or Leiden, blue: Beijing Genomics Institute, orange: Sanger Vertebrate Resequencing team, red: Sanger Experimental Cancer Genetics team, pink: Leeds and QIMR Berghofer. Black indicates a ready dataset. Arrows indicate datasets entering or exiting the pipeline. Details of each step are annotated at their right.

samples included in this study were recruited to the Queensland Familial Melanoma Project (QFMP) [416], and informed consent was obtained from the Human Research Ethics Committee of the QIMR Berghofer Medical Research Institute. Genomic DNA was extracted from peripheral blood using standard methods. This work was done by Lauren G. Aoude, Antonia L. Pritchard, Jane M. Palmer, Judith Symmons and Prof. Nicholas K. Hayward at QIMR Berghofer, Queensland, Australia.

In addition to this dataset, we decided to sequence the whole exomes of 45 additional samples from Leeds: 31 that had a family history of the disease and 14 single cases. Twenty-one of the samples with a family history had been already included in the replication set, but we only had sequence for 701 genes in these samples. Single cases were selected for sequencing if they presented with either MPM or an early age of onset ($\leq$40 years of age), which strongly suggests a genetic component to their disease.

Therefore, our new, integral dataset comprises 184 samples in which the whole protein-coding genome has been sequenced: 41 from the discovery set, 98 Australian samples, and 45 additional Leeds samples, belonging to a total of 105 pedigrees. These samples were processed in different institutes and sequencing centres (Tables 3.1 and A.1.8).

3.2 Exome sequencing and data processing

Sixteen samples from the Australian cohort underwent whole genome sequencing, and the rest were whole exome-sequenced (Table A.1.8). In both cases, DNA libraries were prepared from 5µg of genomic DNA. For whole exomes, exonic regions were captured with the Agilent SureSelect Target Enrichment System, 50 Mb Human All Exon kit, and for whole genomes, libraries were prepared using the standard Illumina library preparation protocol. Then, 100bp paired-end reads were generated on the Illumina HiSeq2000 platform. These samples were processed at Macrogen, Inc. The additional samples from the Leeds cohort were sequenced at the Beijing Genomics Institute (BGI), and were also captured with the Agilent SureSelect Target Enrichment System, 50 Mb Human All Exon kit and sequenced in the Illumina HiSeq2000 platform, generating 90bp paired-end reads. Finally, both sample sets were mapped to the reference GRCh37/hg19 human genome assembly using the Burrows-Wheeler Aligner (BWA) [377], and Leeds samples were further recalibrated and realigned around indels using the GATK package [379].

For Leeds samples, exome capture and sequencing resulted in an average of almost

Table 3.1: **Summary of pedigrees sequenced by Institute and sequencing centre, integrative phase.** Cases marked with an asterisk were selected for an absence of phenotypic risk markers (self-reported sun sensitivity and/or low mole count). The case marked with two asterisks was selected because they presented with three different primary cancers, one of which was melanoma. BGI: Beijing Genomics Institute.

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

78% of target bases being covered by $\geq$10$\times$ across the autosomes and sex chromosomes, rising to 82% when the whole set of exomes, including the Australian samples, were considered. Whole genomes were sequenced to at least 27$\times$ mapped coverage. Variants were then called using SAMtools mpileup [380] and filtered for quality. Then, we removed common variants found in the 1000 Genomes Project, October 2011 release [15] and the dbSNP 135 release [381] (as described in Subsection 2.1.3).

We also decided, as before, to take forward only protein-changing consequences. However, as we used an updated version of the Ensembl database (70) and the VEP (2.8), we also updated the consequences kept for further analyses (Table 3.2). We also kept only variants co-segregating in all affected members of a pedigree, while considering all variants called in pedigrees for which a single individual was sequenced. The processing of the Australian samples was done by Peter Johansson and Mitchell S. Stark, based

at the Oncogenomic Laboratory in QIMR Berghofer. Leeds samples were aligned, improved and variant-called by pipelines written by the Vertebrate Resequencing group at Sanger, while I performed all the variant filtering steps as described in Subsection 2.1.3. After these filtering steps, 23,051 non-polymorphic variants private to a single pedigree remained for downstream analysis, which were filtered for quality, co-segregation in all affected members, and were predicted to affect protein structure or function.

3.3 Gene prioritisation strategy

With the addition of the Australian families, we now had access to a much more extensive set of pedigrees in which multiple members have been sequenced (Table A.1.8). Therefore, we decided to search for genes that had variants, passing the above filtering criteria, co-segregating with melanoma in pedigrees for which we had sequence information for 3 or more members. We found 320 such genes (Table A.1.9); however, we only found 5 genes that had variants co-segregating with melanoma in more than one of these pedigrees (Table 3.3).

It would seem that there is a discrepancy between this list and the one presented in Table 2.7, as genes *RNF213*, *KLHDC8A* and *C6orf25* were found to harbour co-segregating variants in two or more pedigrees for which information was available for 3 or more members in the European phase of this study. However, the list compiled in Table 2.7 included capillary sequencing of additional members that were not exome-sequenced. In particular, *RNF213* had co-segregating variants in UF10, UF14, UF21 and NF2, of which only UF10 had three members sequenced by NGS (Fig. A.1.1). Similarly, *C6orf25* had co-segregating variants in UF10 and one individual from the replication cohort, for which two additional family members were available for testing. As such, these genes are included in Table A.1.9. *KLHDC8A*, however, was found with co-segregating variants in two pedigrees that had only two members sequenced by NGS, with an additional member per pedigree available for PCR testing (UF15 and UF16) (Fig. A.1.1). A novel variant in *SMG1* was found to co-segregate with melanoma in an Australian pedigree for which 3 members were sequenced (discussed in Chapter 4).

This prioritisation strategy, which adds additional weight to variants co-segregating in multiple pedigrees for which we have more information, identified two genes with similar biological roles (as indicated by GO terms) in telomere maintenance, protection of telomeres (*POT1*) and adrenocortical dysplasia homolog (*ACD*) (also known as *TPP1*, *TINT1*, *PTOP* and *PIP1*) (Figs. 3.2a,b and 3.3, and Table 3.3). We then extended our

Table 3.2: **Consequences of variants kept for further analyses, integrative phase.** Table reproduced from refs. [389, 390]

Sequence Ontology term	Sequence Ontology description
transcript_ablation	A feature ablation whereby the deleted region includes a transcript feature
splice_donor_variant	A splice variant that changes the 2 base region at the 5' end of an intron
splice_acceptor_variant	A splice variant that changes the 2 base region at the 3' end of an intron
stop_gained	A sequence variant whereby at least one base of a codon is changed, resulting in a premature stop codon, leading to a shortened transcript
frameshift_variant	A sequence variant which causes a disruption of the translational reading frame, because the number of nucleotides inserted or deleted is not a multiple of three
stop_lost	A sequence variant where at least one base of the terminator codon (stop) is changed, resulting in an elongated transcript
initiator_ codon_variant	A codon variant that changes at least one base of the first codon of a transcript
inframe_insertion	An inframe non synonymous variant that inserts bases into in the coding sequence
inframe_deletion	An inframe non synonymous variant that deletes bases from the coding sequence
missense_variant	A sequence variant, that changes one or more bases, resulting in a different amino acid sequence but where the length is preserved
transcript_amplification	A feature amplification of a region containing a transcript
splice_region_variant	A sequence variant in which a change has occurred within the region of the splice site, either within 1-3 bases of the exon or 3-8 bases of the intron
incomplete_terminal_ codon_variant	A sequence variant where at least one base of the final codon of an incompletely annotated transcript is changed
mature_miRNA_variant	A transcript variant located with the sequence of the mature miRNA
TFBS_ablation	A feature ablation whereby the deleted region includes a transcription factor binding site
TFBS_amplification	A feature amplification of a region containing a transcription factor binding site
TF_binding_site_variant	A sequence variant located within a transcription factor binding site
feature_elongation	A sequence variant that causes the extension of a genomic feature, with regard to the reference sequence
feature_truncation	A sequence variant that causes the reduction of a genomic feature, with regard to the reference sequence

Table 3.3: **Genes with co-segregating variants in more than one pedigree with three members sequenced.** The number of co-segregating pedigrees is shown alongside their pedigree IDs and number of sequenced members. GO terms are extracted from Ensembl release 70.

Gene name	Pedigrees with co-segregating variants	GO terms
NEK10	2 (NF1: 3, AF11: 4)	Protein serine/threonine kinase activity, ATP binding, protein phosphorylation, transferring phosphorus-containing groups, metal ion binding
POT1	2 (UF20: 3, AF1: 3)	Telomere maintenance via telomerase, nuclear telomere cap complex, single-stranded DNA binding, positive regulation of DNA strand elongation, protein binding
ACD	2 (AF19: 3, AF10: 3)	Telomere maintenance, nuclear telomere cap complex, negative regulation of telomere maintenance via telomerase, positive regulation of single-stranded telomeric DNA binding
MPDZ	2 (AF19: 3, AF4: 3)	Tight junction, cell adhesion, protein C-terminus binding, postsynaptic density, apical plasma membrane, virus-host interaction, myelination
SMG1	2 (UF20: 3, AF3: 3)	DNA repair, response to stress, nucleotide binding, nuclear-transcribed mRNA catabolic process, nonsense-mediated decay, protein serine/threonine kinase activity, protein binding, ATP binding

search for other variants falling within these genes in the remaining pedigrees that were not considered in the prioritisation strategy. We identified an additional two pedigrees in which we had sequenced only one individual, UF23 and UF31, also harbouring missense variants in *POT1* (Fig. 3.2c,d), and a single case that presented with MPM carrying a predicted splice site variant (UN4) (Table A.1.8). We could not identify other co-segregating variants in *ACD* (Table 3.4). All of these variants were confirmed by capillary sequencing by Dr. Mark Harland, for Leeds samples, and Lauren G. Aoude, for Australian samples.

In order to obtain a bioinformatic estimation for the pathogenicity of the variants

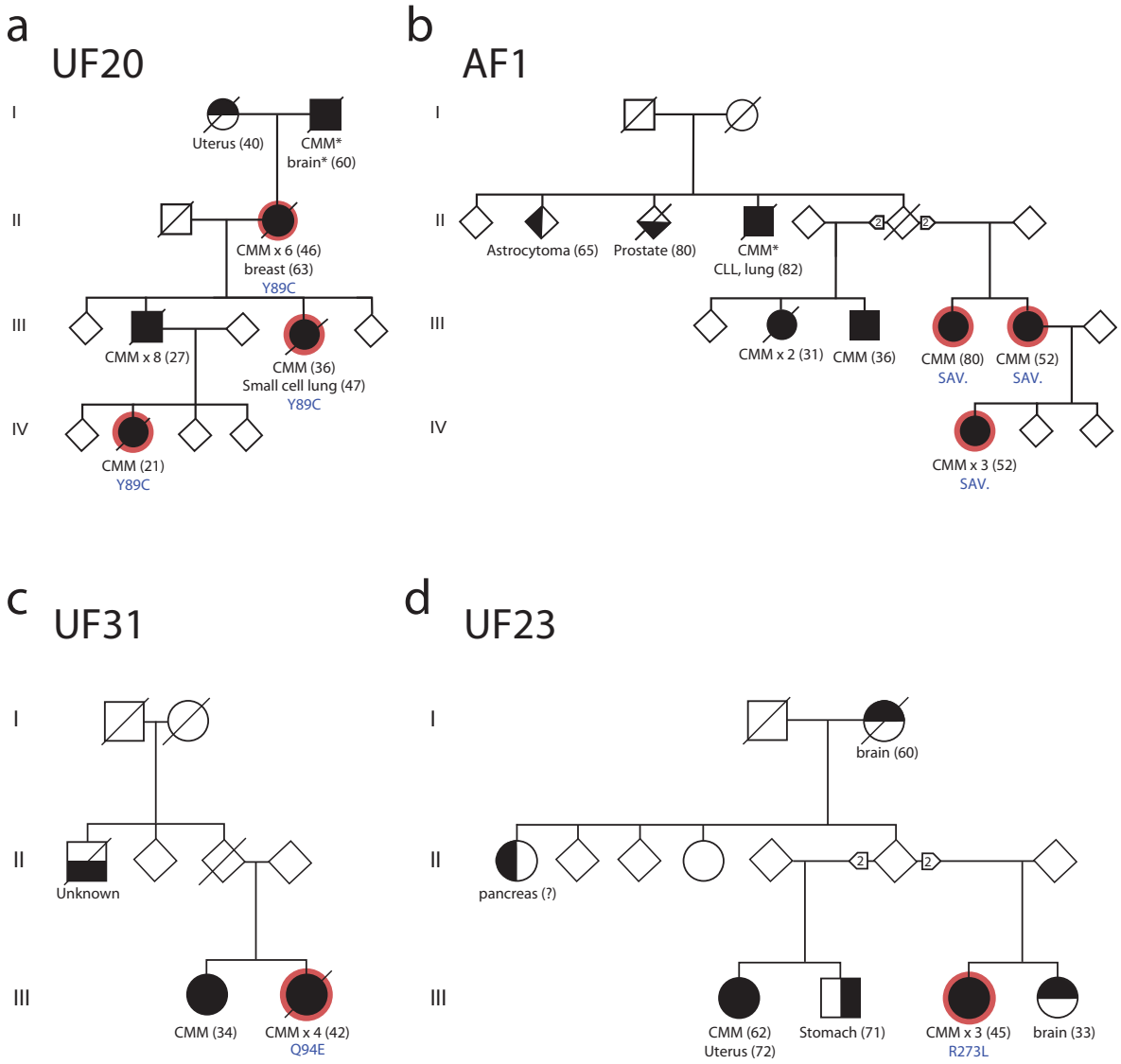

Figure 3.2: **Pedigree structure for families with *POT1* variants.** These pedigrees were sequenced as part of the European phase of this study (UF20), or were additional Australian (AF1) or Leeds (UF23 and UF31) pedigrees included in the integrative phase of the study. Individuals that had their whole exome sequenced are shown with a red outline. Types of cancers are indicated under each symbol, along with the number of primaries and the age of onset (if known) in parenthesis. Circles represent females, squares represent males, diamonds represent individuals of undisclosed sex. A line through the symbol means the individual is deceased. Filled symbols represent individuals with cutaneous malignant melanoma (CMM), and other cancers are indicated by half-filled symbols. All melanomas were confirmed by histological analysis, with the exception of two cases (marked by asterisks). Note that pedigrees have been adjusted to protect the identity of the families without loss of scientific integrity. a) Pedigree UF20, carrying the Y89C variant. b) Pedigree AF1, that was found to carry the splice acceptor variant. c) Pedigree UF31, that was found to be a carrier of the Q94E variant. d) Pedigree UF23, which carries the R273L variant.

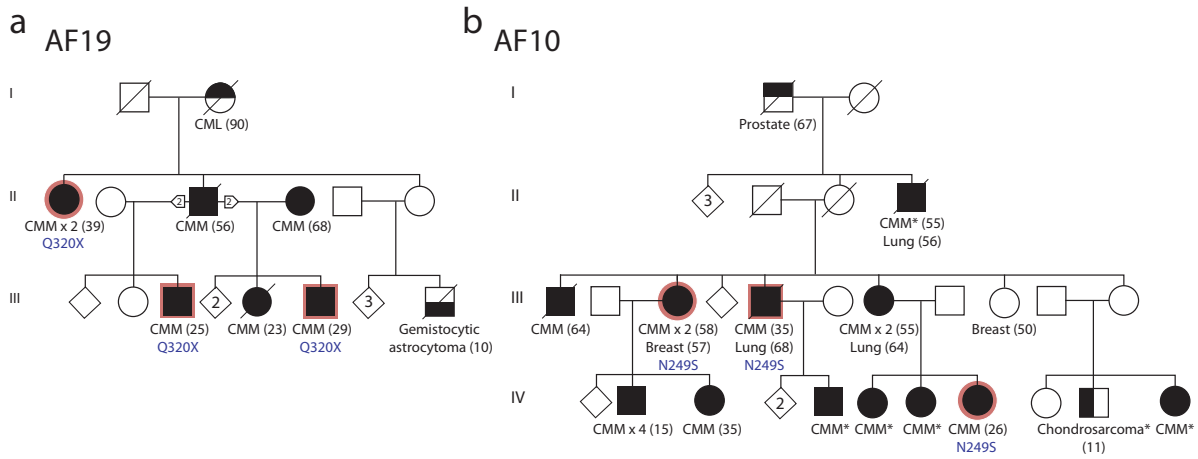

Figure 3.3: **Pedigree structure for families with *ACD* variants.** These pedigrees were additional Australian pedigrees included in the integrative phase of the study. Individuals that had their whole exome sequenced are shown with a red outline. Types of cancers are indicated under each symbol, along with the number of primaries and the age of onset (if known) in parenthesis. Circles represent females, squares represent males, diamonds represent individuals of undisclosed sex. A line through the symbol means the individual is deceased. Filled symbols represent individuals with cutaneous malignant melanoma (CMM), and other cancers are indicated by half-filled symbols. All melanomas were confirmed by histological analysis, with the exception of the cases marked by asterisks. Unaffected siblings are indicated by a diamond with the number of siblings shown in the centre of the symbol. Note that pedigrees have been adjusted to protect the identity of the families without loss of scientific integrity. CML: Chronic myeloid leukaemia. a) Pedigree AF19, that was found to be a carrier of the Q320X variant. b) Pedigree AF10, which carries the N249S variant.

Table 3.4: **Variants in *POT1* and *ACD* identified in melanoma pedigrees.** Pedigrees are shown with the number of individuals that were sequenced. The reference transcripts are ENST00000357628 for POT1 and ENST00000393919 for ACD, which are the only ones per gene for which both the Ensembl automated annotation and the manual Havana team annotation agree.

Gene	Pedigree	Genomic position	Consequence prediction	SIFT [417]	PolyPhen-2 [418]
POT1	UF20 (3)	7:124503684, T/C	Y89C	Deleterious	Probably damaging
POT1	AF1 (3)	7:124465412, C/T	Splice acceptor variant	-	-
POT1	UF23 (1)	7:124493077, C/A	R273L	Deleterious	Probably damaging
POT1	UF31 (1)	7:124503670, G/C	Q94E	Tolerated	Probably damaging
POT1	UN4 (1)	7:124467262, A/C	Splice variant	-	-
ACD	AF19 (3)	16:67692665, G/A	Q320X	-	-
ACD	AF10 (3)	16:67693137, T/C	N249S	Tolerated	Benign

identified in this study, I used the Sorting-Intolerant-From-Tolerant (SIFT) [417] and PolyPhen-2 [418] algorithms. These tools utilise sequence conservation in protein families and multiple-sequence alignments to determine the functional impact that a given substitution is likely to have. Almost all the variants identified in these genes are either obviously disruptive or are predicted to have a high functional impact (Table 3.4).

Using the visualisation tool described in Subsection 2.3.3, I could analyse the positions where these variants lie in the protein structures (Fig. 3.4). None of the variants detected in *POT1* are found in the common variation datasets that I used in this study, and are also not found in the set of ~6,500 exomes released by NHLBI GO ESP [17]. Of the two variants predicted to affect splicing, one falls in a splice acceptor site and the other one affects a base in close proximity to a splice donor site, both in the intron between exons 17 and 18, between amino acids 562 and 563 (Fig. 3.4). Interestingly, all missense variants in *POT1* lie within two repeats of a functional domain annotated in the SUPERFAMILY database [419] as a nucleic-acid binding, oligonucleotide/oligosaccharide-binding (OB) fold (Fig. 3.4).

Of the variants identified in *ACD*, one is obviously disruptive, introducing a premature stop codon, and the other one affects an amino acid in a functional domain. This second variant, N249S, was found at an allelic frequency of 0.0012 in the 500 Exome Project from the Metabolic Disease Group at Sanger, and at an overall frequency of

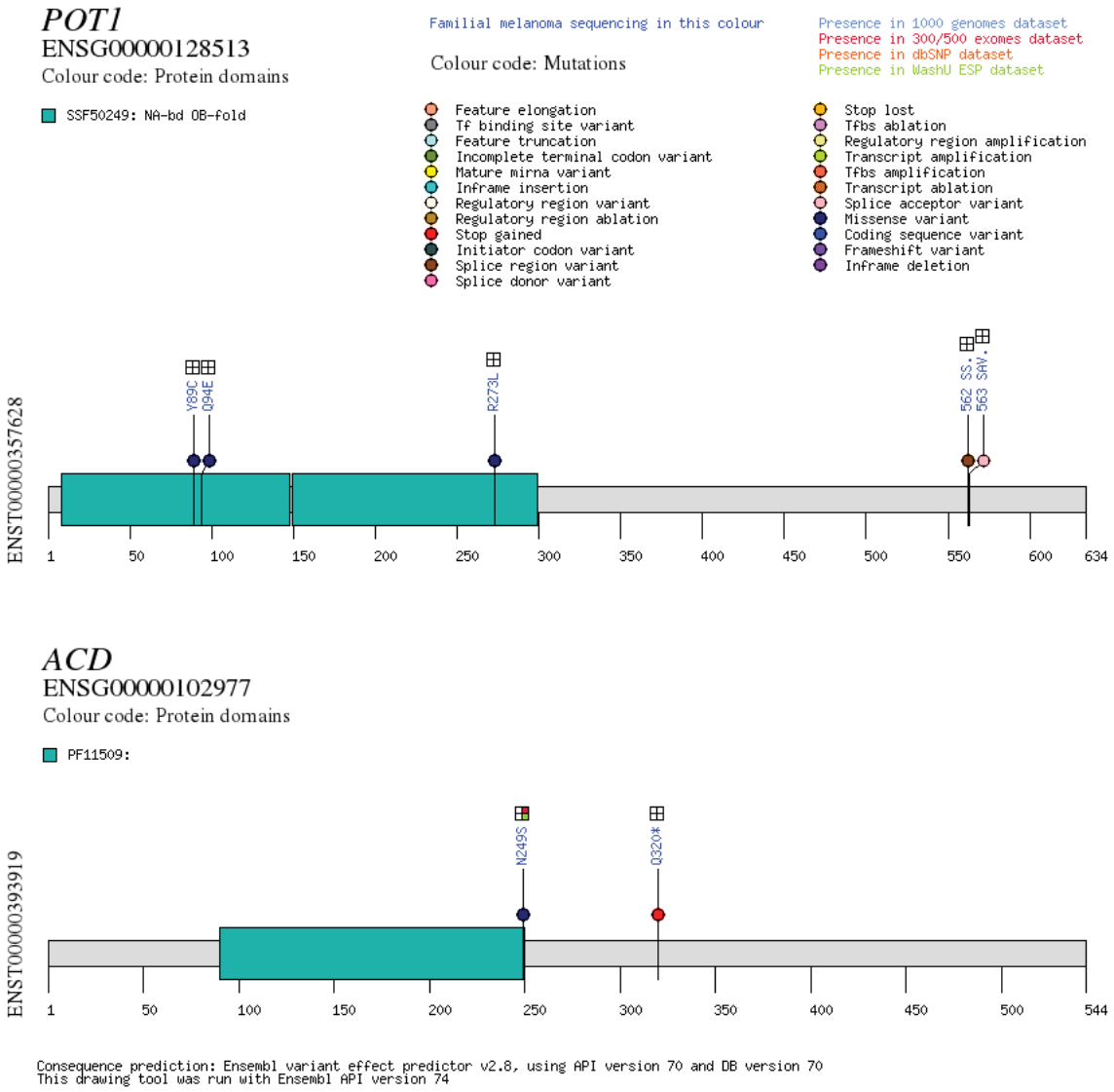

Figure 3.4: **Plot of the *POT1* and *ACD* variants found in familial melanoma pedigrees.** This image shows the variants detected in melanoma pedigrees in protein context. The programme was amended to display Sequence Ontology terms, as per the new terminology of Ensembl release 70. The "300/500 exome dataset" label refers to the 805 in-house exomes used in filtering steps during the European phase of the study. "WashU ESP dataset" refers to 6,500 exomes released by the NHLBI GO ESP [17]. Domains in POT1 were retrieved from the SUPERFAMILY database [419] and those in ACD from Pfam [402]. SS: Splice site, SAV: Splice acceptor variant.

0.0002 in the set of exomes released by NHLBI GO ESP [17]. The functional domain, annotated in the Pfam database [402], is described as having POT1-binding function [420, Note: The Pfam entry for PF11509, which is the domain in the figure, was recently merged with PF10341.].

3.4 Frequency of *POT1* and *ACD* variants in a control dataset

In order to gather further evidence for an association between variants in *POT1* and *ACD* and familial melanoma, I compared the representation between variants in our familial melanoma cases with variants in controls. The controls chosen for this comparison were the same three neurodevelopmental cohorts from the UK10K Project [399] described in Subsection 2.3.1.1 (546 exomes) (Table 2.5). Previous to making this comparison, however, it was necessary to ensure that cases and controls are matched by ancestry, as was performed in Subsection 2.3.1.2. It is necessary to perform this step again, because we have now both an extended set of cases and increased resolution to determine ancestry (with whole genomes instead of a small targeted capture).

To ensure that the controls were matched by ancestry to the melanoma cohort, I performed a PCA using 1,092 individuals across 14 populations from the 1000 Genomes Project, October 2011 release dataset [15]. I took forward for analysis 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. I also excluded SNPs with a minor allele frequency <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 $<1 \times 10^{-5}$ in the UK10K controls. After filtering, 7,196 SNPs remained that were spread across all autosomes. Then, using the R package SNPRelate [421], I estimated the first ten principal components using the 1000 Genomes Project individuals and then projected them onto the melanoma cohort samples and UK10K controls. I then removed controls lying greater than 2 standard deviations (s.d.) from the mean scores for principal component 1 or 2, calculated using only European individuals in the 1000 Genomes Project data set ($n = 20$). This left a total of 523 individuals in the control set, and the result of this analysis is shown in Fig. 3.5. I thank Jimmy Z. Liu, from the Sanger, for his time and valuable advice on

performing this analysis.

Finally, to ensure that individuals within the UK10K cohort were not related, an IBD analysis was performed using the PLINK toolset [422] and the same set of variants that were used for the PCA. For each pair of individuals with an estimated IBD >0.2, one individual was removed at random ($n = 3$). This filtering left 520 exomes for comparison against the melanoma cohort. The IBD analysis was done by Jimmy Z. Liu, from the Sanger.

I then filtered variants in this collection of 520 UK10K control exomes as described above (keeping positions within exonic regions and removing all variants in Phase 1 of the 1000 Genomes Project, October 2011 release [15] and the dbSNP 138 release [381]). I then predicted and filtered consequences as described above.

3.4.1 Variants in *POT1*

Because I used an updated version of dbSNP (138) for the UK10K filtering step, I assessed whether the *POT1* variants found in this phase of the study passed this filter. After ensuring that they did, I counted variants that passed our filters in the UK10K cohort. I found three variants in *POT1* in different individuals of the UK10K cohort: two predicted to affect splice regions falling in introns and one rare, missense variant outside the OB domains (Q539H).

To estimate the probability of observing as many variants as I found in the melanoma cohort, I performed a two-tailed Fisher's exact test comparing the 4 out of 105 families with melanoma, excluding a discovery pedigree, to three individuals out of 520 controls carrying rare variants in *POT1*, yielding a *P*-value of 0.01703. The detected variants in *POT1* are also not found in the 805 in-house control exomes (Fig. 3.4).

The low *P*-value obtained by this comparison suggests that we find rare *POT1* variants in our melanoma cohort at a higher-than-expected frequency when compared to a control population matched by ancestry. Additionally, this *P*-value is likely to be an underestimate, as it does not consider the fact that some of these variants are shared across multiple individuals within a pedigree.

We also decided to search for these variants in 2,402 population-matched controls belonging to the Leeds Melanoma Case-Control study. This control set includes 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. Moreover, when we genotyped 1,739 population-based melanoma cases that were

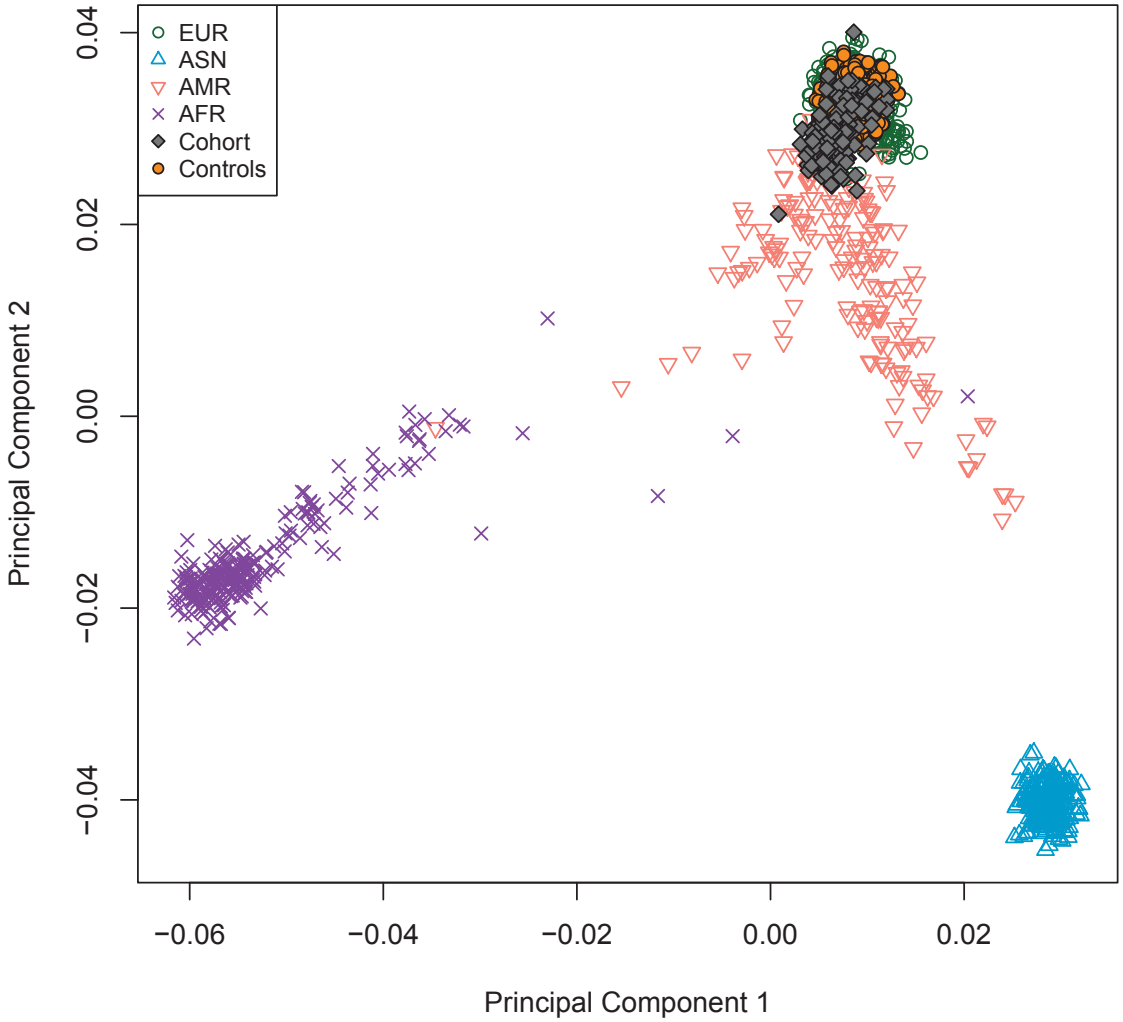

Figure 3.5: **Principal component analysis plot showing that cases and controls are matched by ancestry, integrative phase.** Plot showing the first and second principal components. Ancestry was estimated using the 1000 Genomes Project individuals and then projected onto the melanoma (gray) and UK10K control (orange) cohorts. Note that controls lying greater than 2 s.d. from the mean principal component 1 or 2 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. I did not depict three individuals from the QFMP cohort for whom the zygosity of the called genotypes was not available.

recruited from across Yorkshire, UK, as part of the same study, we found one case who carried the R273L variant. This individual presented with MPM with early onset (at 48 years of age), similar to the phenotype presented by the familial cases. This variant was confirmed by PCR sequencing. The genotyping work I describe here was performed by Dr. Mark Harland at the University of Leeds.

DNA was available from two individuals that were not whole-exome sequenced, one from pedigree UF20 (predicted to be an obligate carrier, individual III-2, Fig. 3.2a) and another one from pedigree UF23 (individual III-1, Fig. 3.2d). Unsurprisingly, the obligate carrier tested positive for the presence of the variant, whereas the additional member from pedigree UF23 did not. I explore the implications of this fact in Chapter 4 and in the Discussion.

Overall, we found that all nine *POT1* variant carrier individuals identified through whole-exome sequencing or by targeted PCR resequencing of additional family members (4 from UF20, 3 from AF1, one from UF23 and one from UF31) had developed melanoma, presenting from one primary (four cases) to eight melanomas at 21 to 80 years of age (Fig. 3.2). One variant carrier from these familial cases also developed breast cancer at 63 years of age, and another developed small cell lung cancer at 47 (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 tumours (pedigrees UF20 and UF23). I cover follow-up experiments on these variants, as well as the role of *POT1* in familial cancer predisposition, in the next Chapter, as well as in the Discussion.

3.4.2 Variants in *ACD*

Because I used an updated version of dbSNP for the UK10K filtering step, I assessed whether the *ACD* variants found in this study passed this filter. The N249S variant is annotated in dbSNP 138, having been submitted by the NHLBI GO ESP [17], in which it is found at an overall allele frequency of 0.0002 (Fig. 3.4). The Q320X variant passed this filter. I then counted the number of individuals in the UK10K control set that carried rare variants in *ACD* and found a single, missense variant, resulting in an amino acid substitution (A72E) that does not occur in any functional domains. A Fisher test would not be informative in this case, as just one variant has been found in each cohort.

The QIMR Berghofer team, headed by Prof. Nicholas K. Hayward, led an investigation on the prevalence and haplotype of the *ACD* N249S variant. Briefly, they genotyped an additional 4 cases or obligate carriers of pedigree AF10, and found that all of them

harboured the variant (7 out of 7 tested cases or obligate carriers) (individuals III-3, III-5, III-7, IV-3, IV-6, IV-7 and IV-8, Fig. 3.3b). Other carriers are individuals III-9, III-11, IV-1 and IV-9. Individual III-4 tested negative for the variant. They also studied the whole exomes of 7 families from Copenhagen, Denmark (recruited by the Danish project of hereditary malignant melanoma) and discovered one family harbouring the *ACD* N249S variant, segregating in 3 out of 4 members tested. This family was found to share the *ACD* haplotype with the Australian family. They also performed linkage analysis of both families and calculated a logarithm (base 10) of odds (LOD) score of 1.14.

They also searched for the N249S variant in Australian population-based case-control panel, but did not find it in 1,669 cases or 1,590 controls. The variant was also absent from 1,000 Danish diabetes cases and 1,000 metabolically healthy controls.

Overall, we found that *ACD* variant carriers presented with zero, one or two primaries, at 26 to 58 years of age (Fig. 3.3). Other malignancies, such as lung and breast cancer, are present in *ACD* variant carrier individuals, and brain cancer and chondrosarcoma are also present in untested first- or second-degree relatives of variant carriers (Fig. 3.3). I cover the implications of these results in the Discussion.

3.5 Summary and conclusion

In this phase of the study, I had access to exome sequences from a much larger set of familial melanoma pedigrees, recruited in Australia as part of the QFMP [416], totalling 98 exomes from 36 pedigrees. We also obtained whole exome sequences from an additional 45 Leeds patients, that were either familial cases (31 patients) or were single cases that presented with MPM or an early age of onset (14 patients) (Table 3.1). With this extended set of patients, an analysis of genes mutated in more than one pedigree with multiple members was more informative than when we performed a similar analysis in the European phase (shown in Table 2.3). We observed only five genes with co-segregating variants in more than one pedigree for which we had sequence for 3 or more members, and while all of them represent plausible candidates, we found it interesting that two of these genes seem to have very similar biological functions (telomere maintenance). These genes are *POT1* and *ACD*. Therefore, we decided to pursue these candidates and attempt to establish their role, if any, in melanoma predisposition.

Interestingly, *POT1* and *ACD* directly interact with each other, as ACD recruits POT1 to telomeres, where they belong to the six-protein complex shelterin [423, 424].

Shelterin has a paramount role in protecting telomere ends from the DNA damage response, as well as controlling telomerase access to telomeres [423]. The location of the mutations is also interesting, as missense variants fall within the OB folds of POT1, which allows it to bind to single-stranded (ss) DNA, or the POT1-interacting domain of ACD [423] (Fig. 3.4). The rest of the variants found in these genes are predicted to be disruptive, introducing a premature stop codon or affecting mRNA splicing.

A search for variants in these genes in a set of control exomes matched by ancestry to the melanoma cases suggested that germline mutations in *POT1* are rare, whereas results were inconclusive for variants in *ACD*. These hypotheses were further supported by genotyping of population-matched case control series for the particular variants found in this study. Therefore, we decided to further investigate the molecular mechanism by which these variants might increase melanoma risk in carriers, as well as their biological consequences.

I describe experiments investigating the function of these variants, as well as other genes detected in the European phase of this study, in the next Chapter. In the Discussion, I address the involvement of the protein complex to which POT1 and ACD belong, shelterin, in melanoma and generally in cancer predisposition.