POT1 loss-of-function variants predispose to familial melanoma

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

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

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

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

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

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

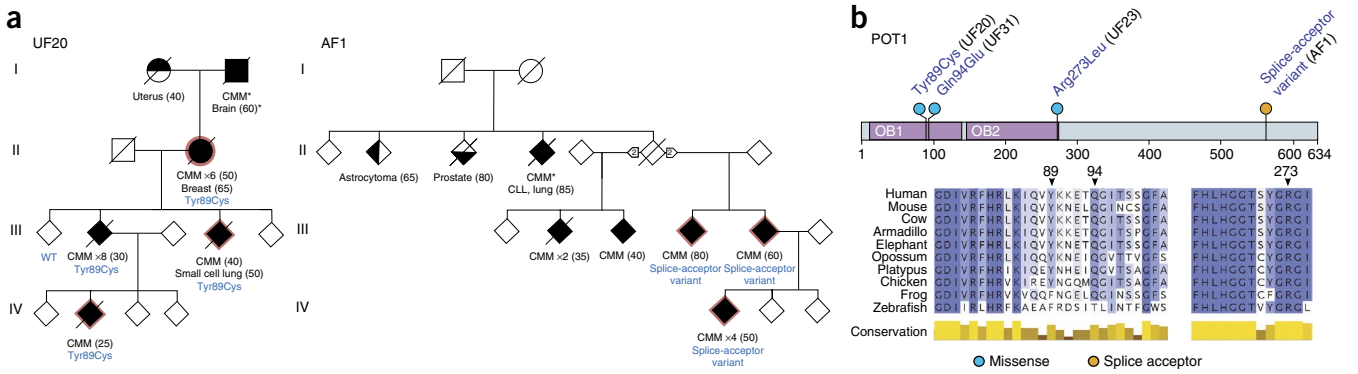

Figure 1 Rare variants in *POT1* found in familial melanoma pedigrees. (**a**) We identified four pedigrees carrying deleterious variants in *POT1*. Shown are a five-case pedigree (UF20) and a six-case pedigree (AF1) carrying the disruptive p.Tyr89Cys OB domain variant and a splice-acceptor variant,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

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

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

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

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

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

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

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

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

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

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

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

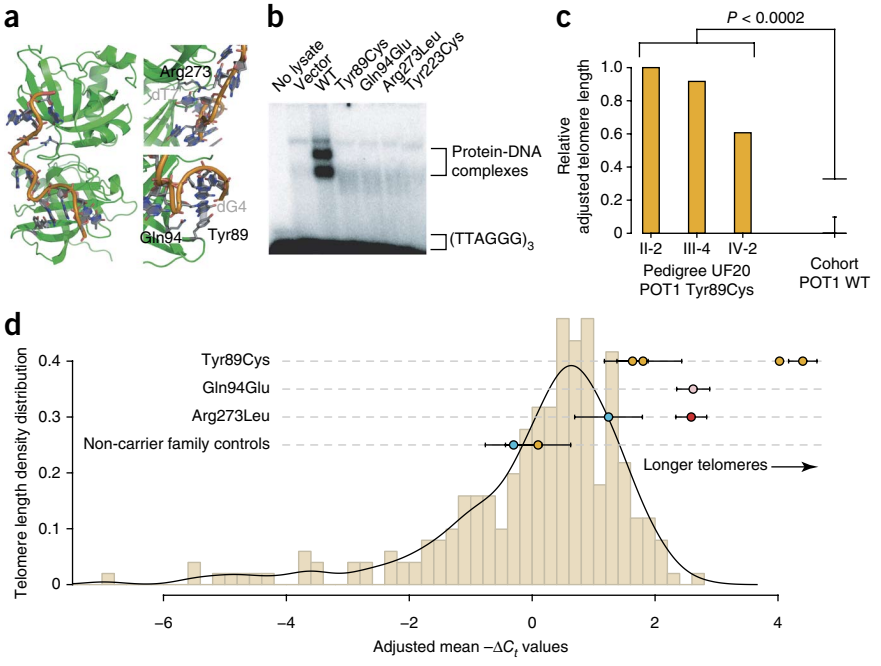

Figure 2 Missense variants in *POT1* disrupt the interaction between POT1 and single-stranded DNA and lead to elongated telomeres. (**a**) Shown are the locations of the POT1 Tyr89, Gln94 and Arg273 residues in the two N-terminal OB domains (green). A telomere-like polynucleotide sequence is shown in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

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

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

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

METHODS

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

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

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

ACKNOWLEDGMENTS

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

AUTHOR CONTRIBUTIONS

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

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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

ONLINE METHODS

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

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

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

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

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

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

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

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

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

doi:10.1038/ng.2947

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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