nature genetics

Rare missense variants in *POT1* predispose to familial cutaneous malignant melanoma

Jianxin Shi[1,25], Xiaohong R Yang[1,25], Bari Ballew[1], Melissa Rotunno[1], Donato Calista[2], Maria Concetta Fargnoli[3], Paola Ghiorzo[4,5], Brigitte Bressac-de Paillerets[6], Eduardo Nagore[7,8], Marie Francoise Avril[9], Neil E Caporaso[1], Mary L McMaster[1], Michael Cullen[1,10], Zhaoming Wang[1,10], Xijun Zhang[1,10], NCI DCEG Cancer Sequencing Working Group[11], NCI DCEG Cancer Genomics Research Laboratory[11], French Familial Melanoma Study Group[11], William Bruno[4,5], Lorenza Pastorino[4,5], Paola Queirolo[5], Jose Banuls-Roca[12], Zaida Garcia-Casado[13], Amaury Vaysse[14,15], Hamida Mohamdi[14,15], Yasser Riazalhosseini[16,17], Mario Foglio[18], Fanélie Jouenne[6], Xing Hua[1], Paula L Hyland[1], Jinhu Yin[19], Haritha Vallabhaneni[19], Weihang Chai[20], Paola Minghetti[2], Cristina Pellegrini[21], Sarangan Ravichandran[22], Alexander Eggermont[6,23], Mark Lathrop[16,17,24], Ketty Peris[3], Giovanna Bianchi Scarra[4], Giorgio Landi[2], Sharon A Savage[1], Joshua N Sampson[1], Ji He[1,10], Meredith Yeager[1,10], Lynn R Goldin[1], Florence Demenais[14,15], Stephen J Chanock[1], Margaret A Tucker[1], Alisa M Goldstein[1], Yie Liu[19] & Maria Teresa Landi[1]

Although *CDKN2A* is the most frequent high-risk melanoma susceptibility gene, the underlying genetic factors for most melanoma-prone families remain unknown. Using whole-exome sequencing, we identified a rare variant that arose as a founder mutation in the telomere shelterin gene *POT1* (chromosome 7, g.124493086C>T; p.Ser270Asn) in five unrelated melanoma-prone families from Romagna, Italy. Carriers of this variant had increased telomere lengths and numbers of fragile telomeres, suggesting that this variant perturbs telomere maintenance. Two additional rare *POT1* variants were identified in all cases sequenced in two separate Italian families, one variant per family, yielding a frequency for *POT1* variants comparable to that for *CDKN2A* mutations in this population. These variants were not found in public databases or in 2,038 genotyped Italian controls. We also identified two rare recurrent *POT1* variants in US and French familial melanoma cases. Our findings suggest that *POT1* is a major susceptibility gene for familial melanoma in several populations.

Approximately 10% of cutaneous malignant melanoma (CMM) cases occur in a familial setting[1]. Established high-penetrance melanoma susceptibility genes include *CDKN2A* on chromosome 9p21 and *CDK4* on chromosome 12q14. Germline mutations of the *CDKN2A* gene have been described in approximately 20–40% of familial melanoma kindreds[2–6], whereas mutations of *CDK4* are quite rare, with only 17 published families worldwide[7,8]. Recently, germline mutations in *BAP1* (located at 3p21) were reported to predispose to melanocytic tumors, including uveal and cutaneous melanomas[9]. In addition, a germline mutation in the promoter of the telomerase reverse transcriptase gene (*TERT*), which encodes the catalytic subunit of telomerase, demonstrated cosegregation with disease in an informative melanoma-prone family[10], suggesting that this gene may also be a rare high-penetrance CMM susceptibility gene. Together, these genes account for melanoma susceptibility in a small proportion of melanoma-prone families.

To identify additional high-penetrance susceptibility genes for familial CMM, we performed whole-exome sequencing in 101 CMM

[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, US National Institutes of Health, US Department of Health and Human Services, Bethesda, Maryland, USA. [2]Department of Dermatology, Maurizio Bufalini Hospital, Cesena, Italy. [3]Department of Dermatology, University of L'Aquila, L'Aquila, Italy. [4]Department of Internal Medicine and Medical Specialties, University of Genoa, Genoa, Italy. [5]Genetics of Rare Hereditary Cancers, Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) San Martino–IST Istituto Nazionale per la Ricerca sul Cancro, Genoa, Italy. [6]Service de Génétique, Gustave Roussy, Villejuif, France. [7]Department of Dermatology, Instituto Valenciano de Oncología, Valencia, Spain. [8]Department of Dermatology, Universidad Católica de Valencia, Valencia, Spain. [9]Université Paris Descartes, Assistance Publique–Hôpitaux de Paris (AP-HP), Hôpital Cochin, Paris, France. [10]Cancer Genomics Research Laboratory, NCI-Frederick, SAIC-Frederick, Inc., Frederick, Maryland, USA. [11]A full list of members appears in the **Supplementary Note**. [12]Department of Dermatology, Hospital General Universitario de Alicante, Alicante, Spain. [13]Laboratory of Molecular Biology, Instituto Valenciano de Oncología, Valencia, Spain. [14]INSERM, UMR 946, Genetic Variation and Human Diseases Unit, Paris, France. [15]Université Paris Diderot, Sorbonne Paris Cité, Institut Universitaire d'Hématologie, Paris, France. [16]McGill University and Génome Québec Innovation Centre, Montreal, Quebec, Canada. [17]Department of Human Genetics, McGill University, Montreal, Quebec, Canada. [18]SAS Quantome, Paris, France. [19]Laboratory of Molecular Gerontology, National Institute on Aging, US National Institutes of Health, US Department of Health and Human Services, Baltimore, Maryland, USA. [20]Section of Medical Sciences, School of Molecular Biosciences, Washington State University, Spokane, Washington, USA. [21]Department of Biotechnological and Applied Clinical Sciences, University of L'Aquila, L'Aquila, Italy. [22]SAIC-Frederick, Inc., Frederick National Laboratory for Cancer Research, Simulation, Analysis and Mathematical Modeling Group, Advanced Biomedical Computing Center, Frederick, Maryland, USA. [23]Université Paris–Sud, Kremlin Bicêtre France, Gustave Roussy, Villejuif, France. [24]Fondation Jean Dausset–Centre d'Etude du Polymorphisme Humain (CEPH), Paris, France. [25]These authors contributed equally to the work. Correspondence should be addressed to M.T.L. (landim@mail.nih.gov).

Received 22 September 2013; accepted 7 March 2014; published online 30 March 2014; doi:10.1038/ng.2941

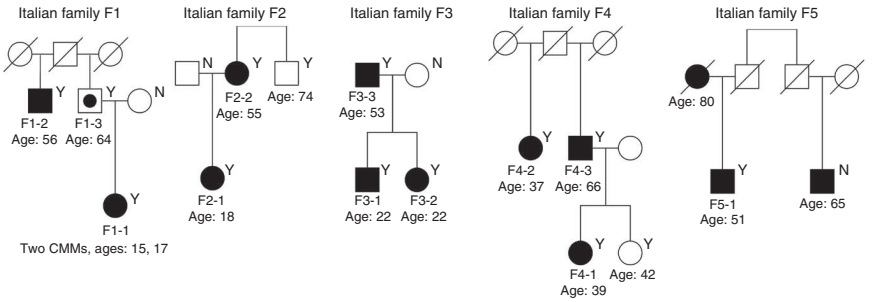

Figure 1 Pedigrees of CMM-prone families with the *POT1* variant g.124493086C>T encoding p.Ser270Asn. Filled symbols, CMM cases; square with a dot in the center, obligate variant carrier. "Y" indicates a variant carrier, and "N" indicates a non-carrier. Age (in years) is the age at diagnosis for CMM cases and the age at examination for unaffected family members.

cases or obligate carriers (assuming dominant inheritance) in 56 unrelated melanoma-prone families without *CDKN2A* or *CDK4* mutations recruited from the Romagna area in Italy. After filtering out common and non-segregating variants (Online Methods), the cosegregating rare variant that was present at the highest frequency in multiple families was a missense rare variant in the protection of telomeres 1 gene (*POT1*; g.124493086C>T; p.Ser270Asn, NP_056265); this variant was observed in all CMM cases or obligate carriers ($n = 11$) from 4 Italian families (F1–F4; **Fig. 1** and **Supplementary Table 1**). We also detected this variant in one of two cases examined in a bilineal Italian family (F5; **Fig. 1**). One affected individual without the *POT1* variant may have inherited the melanoma risk allele from the other side of the family, but the parental origin of this variant could not be determined. Targeted sequencing performed on all available subjects in the five families with the p.Ser270Asn variant confirmed its presence in CMM cases. The mode of transmission was consistent with autosomal dominant inheritance with incomplete penetrance. We performed genotyping for the variant in 2,038 Italian controls, and none carried the variant, resulting in a significantly different frequency of this variant in familial CMM cases compared to controls ($P = 1.2 \times 10^{-11}$ after accounting for the relatedness among familial cases using a chromosome-based exact test; Online Methods). We further genotyped for this variant in 1,824 Italian CMM cases (Online Methods), and this variant was only found in a single sporadic melanoma case from Romagna (**Supplementary Table 1**). This variant was not present in dbSNP, the 1000 Genomes Project (1,092 subjects), the National Heart, Lung, and Blood Institute (NHLBI) Grand Opportunity (GO) Exome Sequencing Project (ESP; including exomes from up to 4,300 subjects of European ancestry and 2,200 subjects of African-American ancestry), the National Cancer Institute (NCI) Division of Cancer Epidemiology and Genetics (DCEG) Cancer Sequencing Working Group database (290 and 200 individuals of European and Asian ancestry, respectively, from families with high risk of non-melanoma malignancies), an in-house McGill exome database (1,307 individuals of European ancestry with no cancers; M.L., unpublished data), and 878 CMM cases and 1,451 controls from Spain (Online Methods).

Because all subjects carrying the p.Ser270Asn variant were from Romagna, we verified the relatedness among p.Ser270Asn carriers using genome-wide SNP genotyping data (Online Methods). The kinship analysis did not detect cryptic relatedness across families at a threshold of 0.05 (**Supplementary Fig. 1**). Interestingly, the analysis identified a haplotype encompassing the *POT1* region encoding p.Ser270Asn shared by all carriers, suggesting that the p.Ser270Asn variant was inherited from the same common ancestor. Using long-range phasing and modeling of the distribution of the lengths of the shared haplotypes encoding p.Ser270Asn, we estimated the age of the most recent common ancestor carrying the variant to be 9.5 (s.d. = 3.9) generations (Online Methods and **Supplementary Fig. 2**).

POT1 is a component of the telomeric shelterin complex that directly binds with high specificity to single-stranded telomeric repeats[11]. POT1 prevents inappropriate processing of exposed chromosome ends by DNA damage response pathways and regulates telomerase function, thereby having a critical role in maintaining telomere integrity and regulating telomere length[12]. POT1 contains two oligonucleotide/oligosaccharide-binding (OB) folds in the N terminus that bind to telomeric overhang. The first OB fold interacts with the first six nucleotides of the telomeric repeat, whereas the second OB fold binds and protects the 3′ single-stranded DNA (ssDNA)[13,14]. Most reported *POT1* somatic mutations[15], including those found in chronic lymphocytic leukemia (CLL)[16], affect the OB folds of POT1, indicating the importance of this region in human cancers. The previously unreported POT1 variant we identified, p.Ser270Asn, is located in the OB2 domain and appears to be in very close proximity to DNA-binding sites, on the basis of the crystal structure (**Fig. 2** and **Supplementary Fig. 3**)[13,14]. The affected residue is highly conserved among vertebrates, and the variant was predicted to be deleterious by most computational programs we evaluated (**Fig. 2** and **Supplementary Table 2**).

Telomere length influences the requirement for telomere maintenance in human tumorigenesis[17]. To determine the impact of the p.Ser270Asn variant on telomere length and integrity, we compared telomere lengths in peripheral blood mononuclear cell (PBMC) DNA from CMM cases carrying the variant to those in age-matched CMM cases without the variant (controls) by telomere restriction fragment analysis (**Fig. 3a**). In agreement with previous reports on *POT1* mutations[11,16,18], we found that, in four unrelated carriers of the p.Ser270Asn variant with melanoma from whom we had obtained PBMCs, telomeres appeared to have increased in length and heterogeneity (evident by the presence of telomeric signals at higher and lower molecular weights), indicating that this variant perturbs telomere length maintenance. Quantitative telomere FISH (Q-FISH) analysis of metaphase spreads from *ex vivo* stimulated PBMCs further confirmed this observation (**Fig. 3b**). Compared to the age-matched control, the carrier of the p.Ser270Asn variant also had a significant increase in the average number of fragile telomeres ($P < 0.0001$; **Fig. 3b**), the aberrant structures that are found in cells deficient in shelterin proteins and/or other proteins required for telomere replication[19–22]. The p.Ser270Asn variant did not lead to chromosome breaks or fusions or to chromosomes or chromatids without detectable telomere signals. Moreover, it did not alter the length of 3′ overhangs or telomere sister chromatid exchanges, affect telomerase activity or induce a DNA damage response at telomeres indicated by the presence of telomere dysfunction–induced foci[23] (data not shown). These findings indicate that this rare variant perturbs telomere maintenance rather than causing direct DNA damage.

Exome sequencing analysis of Italian families also identified two other rare missense substitutions in *POT1* (g.124464052C>G (p.Gln623His) and g.124503540C>T (p.Arg137His)) that were found in all cases sequenced in two additional families (F6 and F7, respectively; **Supplementary Fig. 4**). Both missense variants were absent from public databases and 3,489 genotyped controls (2,038 Italian and 1,451 Spanish individuals) and were predicted to

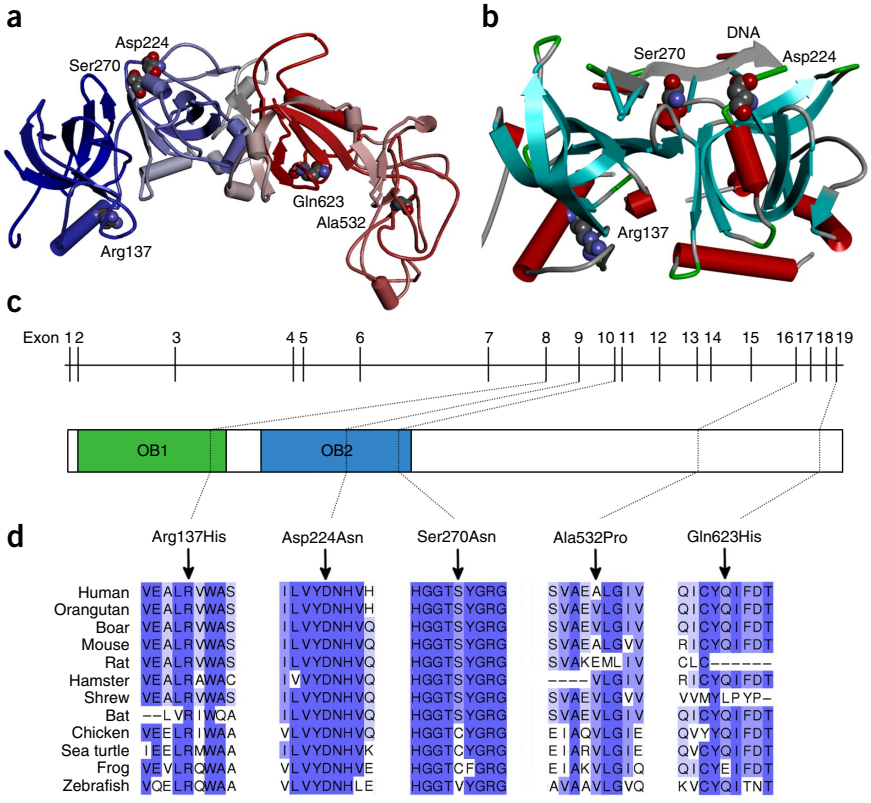

Figure 2 Structural impact of rare variants in POT1. (**a**) Homology-based three-dimensional model of human POT1 (UniProt, Q9NUX5, POTE1_HUMAN) constructed using Phyre2 (refs. 28,29). Secondary structural domains are shown schematically as cylinders (α helices), arrows (β sheets) and tubes (loop regions). The top templates from Protein Data Bank (PDB) used for modeling are 1XJV_A, 1PH4_A, 1JB7_A and 1K8G_C[13,28–30]. The N-terminal sequence end (blue) is the same as for the experimentally derived structure (1XJV_A), and the C-terminal end regions (red) are highly similar to the *Oxytricha nova* telomere-binding protein structures (1PH4 and 1JB7). (**b**) Schematic of the N-terminal region (OB1 and OB2 domains) of the human POT1 protein (Research Collaboratory for Structural Bioinformatics (RCSB) PDB, 1XJV). Helical motifs are shown as cylinders (red), and β sheets are shown as arrows pointing in the direction of the C terminus. The coil and turn region segments are displayed as tubes. The three residues in the OB1 and OB2 domains affected by missense variants are shown in Corey-Pauling-Koltun (CPK) mode. The two affected amino acids (Ser270Asn and Asp224Asn) are in close proximity to bound telomeric ssDNA decamer. Discovery Studio (v. 3.5, Accelrys) was used for visualization and display. (**c**) Schematic of the *POT1* genomic structure and conserved OB domains. *POT1* is composed of 19 exons spanning approximately 108,000 bases of genomic sequence on chromosome 7q31.33. (**d**) Amino acid conservation across POT1 homologs. Higher percentage identity at a given amino acid position is indicated by deeper blue color. The positions of the variants identified in this study are indicated relative to NP_056265.

be deleterious by most algorithms (**Supplementary Table 2**). The p.Arg137His substitution affects the α helix in the OB1 domain, and the affected residue could have a key role in the structural integrity of the protein fold, indirectly influencing DNA binding when altered (**Fig. 2** and **Supplementary Fig. 3**). The p.Gln623His substitution affects the C terminus of POT1, which contains the TPP1-binding region. TPP1 (encoded by *ACD*) is another shelterin component and forms a heterodimer with POT1. POT1-TPP1 binding increases the affinity of POT1 for ssDNA by tenfold and enables the recruitment of telomerase to the overhang[24,25]. Carriers of the p.Arg137His and p.Gln623His variants showed slightly but significantly increased telomere intensity signals (**Supplementary Fig. 5**) and telomere fragility (**Supplementary Fig. 6**) in PBMCs compared to age-matched controls ($P < 0.01$).

We also evaluated copy number variations (CNVs) using exome data in these families (Online Methods) and observed no CNVs disrupting *POT1*.

To assess the overall genetic burden due to rare variants in the *POT1* gene, we sequenced *POT1* exons in 768 CMM cases and 768 controls collected from a multicentric melanoma case-control study from Italy (Online Methods), and we found that CMM cases showed a significant increase in burden for all rare variants (<1% in ESP and 1000 Genomes Project databases) compared to controls (31 carriers among cases and 15 carriers among controls; odds ratio (OR) = 2.1, 95% confidence interval (CI) = 1.1–4.2; $P = 0.024$, Fisher's exact test; **Supplementary Table 3**). The difference was more significant when we restricted the analysis to exonic variants (16 carriers among cases and 3 carriers among controls; OR = 5.4, 95% CI = 1.5–29.2; $P = 0.0021$).

Subsequently, we used exome and targeted sequencing data obtained from familial CMM cases or sporadic cases with multiple primary melanomas (MPMs) in 3 other populations to further evaluate *POT1* variants, including 68 families (139 cases) from the United States, 234 families (267 cases) and 157 MPMs from France and 3 families (10 cases) from Spain. We identified two recurrent germline missense variants (g.124499043C>T (p.Asp224Asn) and g.124469308C>G (p.Ala532Pro). The p.Asp224Asn variant was observed in 4 of 5 CMM cases from a US family (**Supplementary Fig. 4** and **Supplementary Table 1**) and was found in only 1 of 6,500 subjects in ESP. This variant was also seen in a sporadic MPM case from Genoa, Italy. Like p.Ser270Asn, the p.Asp224Asn variant is located in the OB2 domain near DNA-binding sites (**Fig. 2**). It was predicted to be deleterious by all evaluated variant prediction programs (**Supplementary Table 2**). The p.Ala532Pro variant was identified in a familial case (the only case in the family with DNA available) from a three-case family from France and in one French case with five melanomas and one basal cell carcinoma (**Supplementary Fig. 4** and **Supplementary Table 1**) and was not reported in ESP. The p.Ala532Pro variant was predicted to be damaging only by SIFT and PolyPhen-2, but the underlying mutation is located near a splice junction and is predicted to disturb normal splicing. Both p.Asp224Asn and p.Ala532Pro were technically validated, and neither was observed in the NCI DCEG or McGill in-house sequencing databases. Other rare *POT1* variants identified in a single French familial or MPM case, all of which were confirmed by Sanger sequencing, are shown in **Supplementary Table 4**. No rare *POT1* variant cosegregating with disease was found in the Spanish families with CMM.

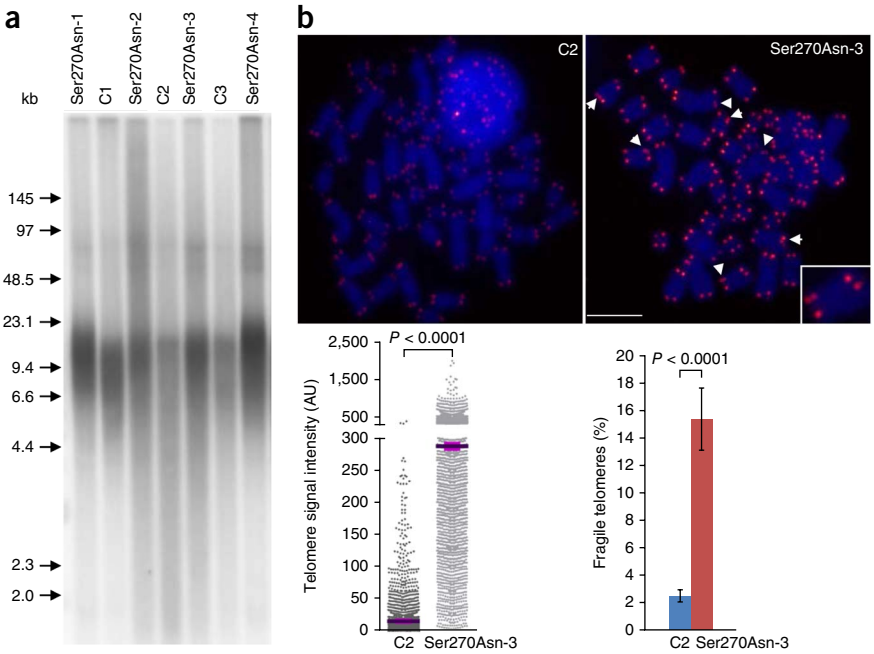

Figure 3 Telomere length in PBMCs from individuals with the *POT1* variant encoding p.Ser270Asn. (**a**) Telomere restriction fragment analysis of PBMCs. A representative blot of at least three experiments shows telomere restriction fragments in each *POT1* variant carrier ($n = 4$) and the corresponding age-matched control (C, melanoma cases without *POT1* variants; $n = 3$). A labeled molecular weight marker (shown in kilobases) is used to determine telomere length. DNA was separated in a CHEF DR-II pulsed-field apparatus at 5 V/cm at an angle of 120° with switching times ramped from 1 to 15 s for 15 h. Telomere length and heterogeneity are demonstrated by telomere signal intensities at higher and lower molecular weights. (**b**) Q-FISH analysis of *ex vivo* stimulated PBMCs. Top, representative metaphase spreads for an age-matched control and *POT1* variant carrier with 4′,6-diamidino-2-phenylindole (DAPI) staining (blue) and telomere fluorescence signals (red). Arrows indicate fragile telomeres (enlarged view in inset; 2× magnification). Scale bar, 5 mm. Bottom left, quantitative measurement of telomere signal intensity in a jitter plot displaying the complete distribution of telomeres with diverse signal intensities. Purple bars denote the mean telomere signal intensities in a single control (C2) and a single carrier of p.Ser270Asn (Ser270Asn-3). Additional variant carriers are described in **Supplementary Figure 6**. Telomere signal intensity is depicted in arbitrary units (AU). Bottom right, percentage of fragile telomeres in C2 and Ser270Asn-3. Thirty metaphase spreads per sample were counted. Error bars represent, s.d. *P* values were obtained from the Wilcoxon rank-sum test. The experiments were repeated three times on each sample.

Because *POT1* was reported to frequently undergo somatic mutations in CLL[16], we also examined *POT1* exome sequencing data obtained from 51 families with lymphoproliferative disorder (LPD) from the United States, including 171 LPD cases affected with CLL ($n = 70$), non-Hodgkin lymphoma (NHL; $n = 16$), Hodgkin lymphoma ($n = 46$) or Waldenström macroglobulinemia ($n = 39$) (Online Methods). No rare variants in *POT1* were identified in the subjects with LPD.

In summary, we have identified *POT1* as a susceptibility gene for familial melanoma in several populations. This finding is further supported by the discovery of *POT1* variants in an independent study of melanoma-prone families from the UK and Australia[26]. In particular, we have identified germline rare variants of *POT1* in Romagna, Italy, which showed a frequency comparable to that for *CDKN2A* mutations among melanoma-prone families in this population[27]. Together with the finding of *TERT* as a potential susceptibility gene in melanoma, our findings suggest that genes involved in telomere maintenance may have important roles in melanoma development.

URLs. PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/; UCSC Genome Browser, http://genome.ucsc.edu/; Kaviar, http://db.systemsbiology.net/kaviar/; dbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP/; 1000 Genomes Project, http://www.1000genomes.org/; CoNIFER, http://conifer.sourceforge.net/; RegulomeDB, http://regulome.stanford.edu/; PolyPhen-2, http://genetics.bwh.harvard.edu/pph2/; SIFT, http://sift.jcvi.org/; PROVEAN, http://provean.jcvi.org/index.php; MutationAssessor, http://mutationassessor.org/; MutationTaster, http://www.mutationtaster.org/; GERP, http://mendel.stanford.edu/SidowLab/downloads/gerp/; PhastCons, http://compgen.bscb.cornell.edu/phast/phastCons-HOWTO.html; PHAST, http://compgen.bscb.cornell.edu/phast/; ProPhylER, http://www.prophyler.org/; BEAGLE, http://faculty.washington.edu/browning/beagle/beagle.html; Genome Analysis Toolkit (GATK), http://www.broadinstitute.org/gatk/; NHLBI ESP, http://evs.gs.washington.edu/EVS/; VariantGPS, http://variantgps.nci.nih.gov/; Jalview, http://www.jalview.org/; Cancer Genomics Research (CGR), http://cgf.nci.nih.gov/; Novoalign, http://www.novocraft.com/; AmpliSeq primer design engine, https://www.ampliseq.com/; Ion Community, http://ioncommunity.lifetechnologies.com/.

METHODS

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

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

ACKNOWLEDGMENTS

This work was supported by the Intramural Research Program of the US National Institutes of Health (NIH), National Cancer Institute (NCI), Division of Cancer Epidemiology and National Institute on Aging (NIA), Division of Molecular Gerontology. The samples from the Instituto Valenciano de Oncología were retrieved from the Biobanco del Instituto Valenciano de Oncología. The Genoa collection was partly supported by Università degli Studi di Genova Progetti di Ricerca di Ateneo PRA 2012-2013 and IRCCS Azienda Ospedaliera Universitaria San Martino–IST Istituto Nazionale per la Ricerca sul Cancro, 5 per 1000 per la Ricerca Corrente. The French MELARISK collection was partly supported by Programme Hospitalier de Recherche Clinique (AOM-07-195) and Ligue Nationale Contre le Cancer (PRE 09/FD). We acknowledge the contribution of Institut Gustave-Roussy (IGR) Biobank and Fondation Jean Dausset–CEPH Biobank in providing samples for melanoma-prone families or individual's. The exome sequencing of French samples was supported by a grant from Génome Québec, le Ministère de l'Enseignement Supérieur, de la Recherche, de la Science et de la Technologie (MESRST) du Québec and McGill University. This project has also been funded in part with federal funds from the US NIH, NCI, under contract HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the US Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US government.

AUTHOR CONTRIBUTIONS

J.S., X.R.Y., M.R., X.H., A.M.G., L.R.G., J.N.S., F.D., A.V., H.M. and M.T.L. performed and interpreted genetics analyses. B.B., S.A.S. and S.R. performed

functional prediction analyses. D.C., M.C.F., P.G., B.B.-d.P., E.N., M.F.A., W.B., L.P., P.Q., J.B.-R., Z.G.-C., F.J., K.P., G.B.S., G.L., M.A.T. and M.T.L. directed the clinical work for the melanoma cases from Italy, Spain, the United States and France. N.E.C. and M.L.M. directed the clinical work for the LPD cases. M.Y., M.A.T. and S.J.C. directed all sequencing analyses at NCI DCEG CGR. J.H., M.C., Z.W., X.Z. and the NCI DCEG Cancer Genomics Research Laboratory conducted the whole-exome and targeted sequencing and the genotyping analyses for the Italian, US, Spanish and part of the French melanoma cases and controls. The NCI DCEG Cancer Sequencing Working Group examined and enrolled families without melanoma and LPD for whole-exome sequencing analyses. The French Familial Melanoma Study Group examined and enrolled melanoma cases and controls from France. M.L., Y.R. and M.F. conducted whole-exome sequencing and genotyping for the French melanoma cases and individuals with other non-cancer diseases. P.L.H., P.M., C.P. and A.E. oversaw sample preparation for the laboratory studies. J.Y., H.V., W.C. and Y.L. performed functional analyses of POT1. M.T.L. designed the overall study. X.R.Y., J.S., B.B., A.M.G. and M.T.L. drafted the manuscript. All authors contributed to the final manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Goldstein, A.M. & Tucker, M.A. Genetic epidemiology of cutaneous melanoma: a global perspective. *Arch. Dermatol.* **137**, 1493–1496 (2001).
2. Cannon-Albright, L.A. *et al.* Assignment of a locus for familial melanoma, MLM, to chromosome 9p13-p22. *Science* **258**, 1148–1152 (1992).
3. Hussussian, C.J. *et al.* Germline p16 mutations in familial melanoma. *Nat. Genet.* **8**, 15–21 (1994).
4. Gruis, N.A. *et al.* Homozygotes for *CDKN2* (p16) germline mutation in Dutch familial melanoma kindreds. *Nat. Genet.* **10**, 351–353 (1995).
5. Goldstein, A.M. Familial melanoma, pancreatic cancer and germline *CDKN2A* mutations. *Hum. Mutat.* **23**, 630 (2004).
6. Eliason, M.J. *et al.* Population-based prevalence of *CDKN2A* mutations in Utah melanoma families. *J. Invest. Dermatol.* **126**, 660–666 (2006).
7. Zuo, L. *et al.* Germline mutations in the p16INK4a binding domain of CDK4 in familial melanoma. *Nat. Genet.* **12**, 97–99 (1996).
8. Puntervoll, H.E. *et al.* Melanoma prone families with *CDK4* germline mutation: phenotypic profile and associations with *MC1R* variants. *J. Med. Genet.* **50**, 264–270 (2013).
9. Wiesner, T. *et al.* Germline mutations in *BAP1* predispose to melanocytic tumors. *Nat. Genet.* **43**, 1018–1021 (2011).
10. Horn, S. *et al. TERT* promoter mutations in familial and sporadic melanoma. *Science* **339**, 959–961 (2013).
11. Loayza, D. & De Lange, T. POT1 as a terminal transducer of TRF1 telomere length control. *Nature* **423**, 1013–1018 (2003).
12. Palm, W. & de Lange, T. How shelterin protects mammalian telomeres. *Annu. Rev. Genet.* **42**, 301–334 (2008).
13. Lei, M., Podell, E.R., Baumann, P. & Cech, T.R. DNA self-recognition in the structure of Pot1 bound to telomeric single-stranded DNA. *Nature* **426**, 198–203 (2003).
14. 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).
15. 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).
16. Ramsay, A.J. *et al. POT1* mutations cause telomere dysfunction in chronic lymphocytic leukemia. *Nat. Genet.* **45**, 526–530 (2013).
17. Taboski, M.A. *et al.* Long telomeres bypass the requirement for telomere maintenance in human tumorigenesis. *Cell Reports* **1**, 91–98 (2012).
18. Kendellen, M.F., Barrientos, K.S. & Counter, C.M. POT1 association with TRF2 regulates telomere length. *Mol. Cell. Biol.* **29**, 5611–5619 (2009).
19. Martínez, P. *et al.* Increased telomere fragility and fusions resulting from TRF1 deficiency lead to degenerative pathologies and increased cancer in mice. *Genes Dev.* **23**, 2060–2075 (2009).
20. Sfeir, A. *et al.* Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. *Cell* **138**, 90–103 (2009).
21. Badie, S. *et al.* BRCA2 acts as a RAD51 loader to facilitate telomere replication and capping. *Nat. Struct. Mol. Biol.* **17**, 1461–1469 (2010).
22. Gu, P. *et al.* CTC1 deletion results in defective telomere replication, leading to catastrophic telomere loss and stem cell exhaustion. *EMBO J.* **31**, 2309–2321 (2012).
23. Takai, H., Smogorzewska, A. & de Lange, T. DNA damage foci at dysfunctional telomeres. *Curr. Biol.* **13**, 1549–1556 (2003).
24. Wang, F. *et al.* The POT1-TPP1 telomere complex is a telomerase processivity factor. *Nature* **445**, 506–510 (2007).
25. Xin, H. *et al.* TPP1 is a homologue of ciliate TEBP-β and interacts with POT1 to recruit telomerase. *Nature* **445**, 559–562 (2007).
26. Robles-Espinoza, C.D. *et al. POT1* loss-of-function variants predispose to familial melanoma. *Nat. Genet.* doi:10.1038/ng.2947 (30 March 2014).
27. Landi, M.T. *et al.* Genetic susceptibility in familial melanoma from northeastern Italy. *J. Med. Genet.* **41**, 557–566 (2004).
28. Horvath, M.P. & Schultz, S.C. DNA G-quartets in a 1.86 Å resolution structure of an *Oxytricha nova* telomeric protein-DNA complex. *J. Mol. Biol.* **310**, 367–377 (2001).
29. Theobald, D.L. & Schultz, S.C. Nucleotide shuffling and ssDNA recognition in *Oxytricha nova* telomere end-binding protein complexes. *EMBO J.* **22**, 4314–4324 (2003).
30. Classen, S., Ruggles, J.A. & Schultz, S.C. Crystal structure of the N-terminal domain of *Oxytricha nova* telomere end-binding protein α subunit both uncomplexed and complexed with telomeric ssDNA. *J. Mol. Biol.* **314**, 1113–1125 (2001).

ONLINE METHODS

Subjects and families. Italian families with CMM, US families with CMM, Spanish families with CMM and families with LPD are participants in institutional review board (IRB)-approved studies at NCI. Detailed descriptions of these studies have been reported previously[27,31,32]. In brief, the study of Italian families with CMM involving exome sequencing and *POT1* sequencing or genotyping analyses included 101 CMM cases or obligate carriers and 198 unaffected individuals from 56 families negative for mutations in *CDKN2A* and *CDK4* with 2–5 affected relatives who were recruited at the Dermatology Unit of Maurizio Bufalini Hospital in Cesena, Italy. The study of US families with CMM comprised 68 families negative for mutations in *CDKN2A* and *CDK4* with 2–13 affected individuals who were ascertained through healthcare professionals or self-referral. All study participants were of European ancestry, and exome sequencing data were based on 139 CMM cases and 6 unaffected individuals. The families with LPD had two or more individuals affected with the same lymphoproliferative disorders, including CLL ($n = 70$), Waldenström macroglobulinemia ($n = 39$), Hodgkin lymphoma ($n = 46$) and NHL ($n = 16$), and were accrued through healthcare professionals or self-referral. Exome sequencing was based on 171 affected individuals or obligate carriers and 5 unaffected individuals from 51 families. The French families with melanoma and sporadic cases with MPMs were recruited through a national network of French Dermatology and Oncogenetic clinics that constitute the French Familial Melanoma Study Group and the MELARISK collection. This study was approved by IRBs and has been described in detail previously[33]. The French data investigated in the present study included two sets of cases negative for mutations in *CDKN2A* and *CDK4* that were sequenced independently. The first set contained 22 melanoma-prone families (45 cases) and 4 sporadic melanoma cases with MPMs with whole-exome sequencing conducted at McGill University. The second set included 212 families with melanoma (222 cases) and 153 MPMs, with *POT1* sequenced at the Cancer Genomics Research (CGR) Laboratory at NCI. The Spanish family study included three families with seven affected and ten unaffected individuals, with whole-exome sequencing conducted at CGR.

The 2,702 additional CMM cases included in *POT1* genotyping were derived from a multicentric case-control study from Italy ($n = 1,824$)[34] and a case-control study from Valencia, Spain ($n = 878$)[35]. Unaffected individuals were controls from the two melanoma case-control studies (511 Italian, 1,451 Spanish) and 1,527 individuals without a personal history of cancer who were recruited as controls for Environment And Genetics in Lung cancer Etiology (EAGLE), a population-based case-control study of lung cancer conducted in Italy[36,37]. A subset of these subjects, including 768 cases and 768 controls from Italy, was further sequenced for the coding region of *POT1*. Subjects were selected on the basis of their residence within or in the proximity of the Romagna area, where the rare variant encoding p.Ser270Asn that arose as a founder mutation was identified.

Of note, for all studies, familial CMM cases included individuals with single or multiple primary melanomas. Sporadic CMM cases with MPMs were counted separately and defined as MPMs to emphasize their potential genetic etiology.

All studies were approved by IRBs at NCI and/or local institutions. All subjects provided written informed consent.

Exome sequencing. Whole-exome sequencing for the Italian families with CMM, the US families with CMM, the Spanish families with CMM and the families with LPD was performed at CGR, DCEG, NCI. Detailed procedures are presented in the **Supplementary Note** and have been described previously[38]. Briefly, for each sample, 1.1 µg of genomic DNA extracted from blood was used for exome sequence capture, which was performed with SeqCap EZ Human Exome Library v2.0 or v3.0 (Roche NimbleGen). Pools of captured DNA then underwent paired-end sequencing using an Illumina HiSeq according to Illumina-provided protocols for 2 × 100-bp paired-end sequencing. Each exome was sequenced to high depth to achieve a minimum threshold of 80% of coding sequence covered with by least 15 reads, on the basis of the UCSC hg19 'known gene' transcripts. According to US NIH policy, the exome sequencing data used for discovery will be released to the database of Genotypes and Phenotypes (dbGaP).

Whole-exome sequencing of the French subjects was performed at the McGill University and Génome Québec Innovation Centre. Briefly, 3 µg of genomic DNA was used to perform exome capture with an Agilent SureSelect kit according to the manufacturer's instructions. We performed 100-bp paired-end sequencing on the Illumina HiSeq 2000 platform. We generated approximately 10 Gb of sequence for each sample such that at least 90% of the coding sequence was covered by at least ten reads.

Bioinformatic analysis. Details of the bioinformatic analysis pipeline are presented in the **Supplementary Note** and were previously described[38]. Briefly, sequencing reads were aligned to the hg19 reference genome using Novoalign software version 2.07.14. Variant discovery and genotype calling of multi-allelic substitutions, insertions and deletions was performed on all individuals globally using the UnifiedGenotyper module from the Genome Analysis Toolkit (GATK)[39] with the minimum call quality parameter set to 30. An in-house NCI DCEG database containing variants present in 490 exomes (200 Asians and 290 individuals of European ancestry) from subjects in families with melanoma-unrelated cancer sequenced in parallel with our families with CMM and LPD was used as an additional control data set. We excluded variants that were observed more than once in the 1000 Genomes Project or the NHLBI ESP databases or were present more than once in our in-house exome database. We applied a dominant genetic model that required the variant to be present in all affected individuals within a family (including the single case sequenced in the family) or in all but 1 affected subject in families with >3 cases sequenced to reduce the impact of phenocopies or sequencing errors.

Variant annotation. We used computational tools including PolyPhen-2, SIFT, Provean, MutationAssessor and MutationTaster to predict the potential impact of sequence variants on protein function. We also obtained conservation scores using GERP, PhastCons and PhyloP, as well as ProPhylER, to predict the impact of mutations on the basis of evolutionary constraint analyses.

Chromosome-based exact test. We developed a chromosome-based exact test to test the association between CMM and the p.Ser270Asn variant by appropriately accounting for the relatedness of familial CMM cases within the same families. The details of this approach are presented in the **Supplementary Note**. Briefly, we first conducted a kinship analysis, which did not identify cryptic relatedness across families using a kinship coefficient cutoff of 0.05. We then inferred identity by descent (IBD) at the region encoding p.Ser270Asn using SNP genotypes on the basis of exome sequencing data to identify independent chromosomes and the number of cases carrying each chromosome in each family at the locus. Under H_0 of no association, we randomly chose 5 chromosomes (from $n = 4,229$ total independent chromosomes), assigned them as p.Ser270Asn variant alleles and counted the number of cases carrying the variant allele (denoted M). The P value for the chromosome-based exact test is defined as the probability $P(M \geq M_0|H_0)$, where $M_0 = 11$, which is the observed number of familial CMM cases carrying the variant allele encoding p.Ser270Asn (the obligate carrier was not included in this analysis).

Estimation of kinship coefficients across Italian families with CMM. We estimated kinship coefficients to confirm the relationships of family members and to detect cryptic relatedness. In a first analysis, we merged the genotypes of Italian families with CMM and the EAGLE controls for 27,674 common autosomal SNPs overlapping between exomic sequencing (families with CMM) and the Illumina HumanHap550 BeadChip (EAGLE controls) and estimated the kinship coefficient using PLINK (v.1.07)[40]. In a second analysis, we genotyped the subjects carrying rare variants in *POT1* using Illumina OmniExpress arrays with higher density to refine the estimates of kinship coefficients for these *POT1* variant carriers using 306,684 common SNPs overlapping on the two Illumina genotyping platforms.

Estimation of the age of the most recent common ancestor carrying the variant encoding p.Ser270Asn. To investigate whether variants occurred independently or were inherited from a common ancestor, we performed haplotype analysis using BEAGLE[41], combining genotype data for all carriers of p.Ser270Asn and EAGLE controls to improve haplotyping accuracy.

We next estimated the age of the most recent common ancestor carrying the allele encoding p.Ser270Asn on the basis of the lengths of the haplotypes shared by carriers. Because close relatives are from the same lineage and do not provide additional information, we included only the proband from each of the five families with the p.Ser270Asn variant as well as the single sporadic case who also carried the p.Ser270Asn variant. For a subject with one parent genotyped, we resolved the phase for SNPs homozygous in the subject and heterozygous in the parent or vice versa. For subjects with no parent genotyped, we followed the 'surrogate parent' concept[42] to identify subjects who shared a long stretch of IBD and resolved the phase for this stretch. For a given pair of subjects, the length of shared IBD was determined using SNPs with resolved phase. This approach provided accurate estimates for the lengths of shared haplotypes, with error of typically less than 50 kb in the haplotype boundary, which has a minimal effect on the inference of the age of the most recent common ancestor. Given the haplotype sharing structure, we modified the likelihood framework for sparse microsatellite markers[43] to estimate the age of the most recent common ancestor and the corresponding s.d.

Copy number variation detection from exomic sequencing data. CoNIFER (v0.2.2)[44] was used to analyze the sequence data on chromosome 7 for subjects in the Italian families with exome sequencing data. Briefly, the reads per kilobase per million (RPKM) value for each exon was calculated for each subject, and this value was normalized across subjects to derive ZRPKM (a *z* score of RPKM with mean of 0 and variance of 1). Singular value decomposition (SVD) was then performed on ZRPKM values, and the top components (six in our analysis), which typically represent sequencing artifacts, were removed. Resulting SVD-ZRPKM values were used to detect CNVs. CNVs were called if the average SVD-ZRPKM value was greater than 1.5 (duplication) or less than –1.5 (deletion) for any interval. We detected 35 CNVs on chromosome 7, including 26 duplications and 9 deletions in 29 subjects, none of which disrupted *POT1*.

POT1 crystal structure. The homology-based three-dimensional model of human POT1 (UniProt, Q9NUX5, POTE1_HUMAN) was constructed using the Phyre2 algorithm[28,29]. The POT crystal structure of the N-terminal region (OB1 and OB2 domains) of the human POT1 protein (RCSB PDB, 1XJV) was visualized using Discovery Studio (v. 3.5, Accelrys). Multiple-sequence alignments were generated for homologous POT1 protein sequences using T-Coffee[45] to evaluate conservation. Alignments were generated with the following sequences: NP_056265.2 (*Homo sapiens*), NP_001127526.1 (*Pongo abelii*), NP_001131107.1 (*Sus scrofa*), NP_598692.1 (*Mus musculus*), NP_001019493.1 (*Rattus norvegicus*), EGW09411.1 (*Cricetulus griseus*), ELW65533.1 (*Tupaia chinensis*), ELK29766.1 (*Myotis davidii*), NP_996875.1 (*Gallus gallus*), EMP32613.1 (*Chelonia mydas*), NP_998876.1 (*Xenopus tropicalis*) and NP_001232889.1 (*Danio rerio*). Jalview[46] was used to visualize and format the alignment.

***POT1* gene sequencing.** A custom set of AmpliSeq primers from Life Technologies was used to amplify the coding region of the *POT1* gene at CGR, DCEG, NCI. Primers were created using the AmpliSeq primer design engine. Genomic DNA (40 ng) was used for each sample. The AmpliSeq process was performed according to the official process documentation found on the Ion Community. Samples were quality checked for proper amplicon length and quantity on an Agilent Bioanalyzer and were then sequenced according to the current Ion Torrent template and sequencing preparation documentation. Each PGM Ion 316 sequencing run consisted of 96 barcoded samples. The resulting sequencing data yielded coverage for each amplicon of >300× in depth.

Rare variant burden tests. To assess the overall genetic burden due to rare variants in *POT1*, we counted subjects carrying any rare variant in *POT1* and tested the difference between CMM cases and controls using Fisher's exact test. This test was based on *POT1* gene sequencing using Ion Torrent performed for 768 cases and 768 controls from Italy. We excluded variants and genotypes with low quality scores (<30), and we only included variants that were called by both the Torrent Variant Caller and GATK to ensure the accuracy of the burden test. This quality control step was performed with blinding to case or control status. We first included all variants with minor allele frequency less than 1% in ESP and the 1000 Genomes Project, regardless of their predicted functions. We then repeated the analysis with restriction to exonic variants. *P* values were two-sided. On the basis of simulation studies, we estimated that 768 cases and 768 controls had statistical power = 0.8 to identify OR = 2.3 for differences in the frequency of rare variant carriers between cases and controls, for overall rare variant allele frequency = 0.01 and OR = 5 for rare variant allele frequency = 0.002.

Genotyping of *POT1* variants. TaqMan assay (Life Technologies) validation and SNP genotyping were performed at CGR, DCEG, NCI. Details for the procedures are described in the **Supplementary Note**. Briefly, loci for which successful TaqMan assay designs were generated were validated using the 102 Variant GPS population panel and the 262 HapMap control panel (Coriell Cell Repositories). Assay validation was based on concordance (>99%) with reported HapMap genotypes (when available). SNP genotyping was performed using a 9700 Thermal Cycler (Life Technologies), and endpoint reads were evaluated using the 7900HT Sequence Detection System (Life Technologies). Cluster analysis was performed using SDS v2.2.2 software (Life Technologies).

Telomere restriction fragment analysis. Genomic DNA from PBMCs was digested with HinfI and RsaI and analyzed by standard Southern blotting using a [^{32}P-α]dCTP-labeled telomere probe under highly stringent conditions as previously described[47]. Experiments were repeated three times on each sample.

Quantitative telomere FISH and chromosome orientation FISH. PBMCs were briefly cultured with RPMI 1640 medium (Gibco BRL Life Technologies) supplemented with 2 mM L-glutamine, 25 mM HEPES and 20% FCS in the presence of 1 to 10% interleukin (IL)-2 (Advanced Biotechnologies) and phytohemagglutinin P (5 μg/ml; Sigma). To detect telomere sister chromatid exchange (T-SCE), PBMCs were subcultured in medium containing a 3:1 ratio of BrdU (5-bromo-2′-deoxyuridine) to BrdC (5-bromo2′-deoxycytidine) (Sigma) at a final concentration of 1×10^{-5} M and collected around 24 h. Colcemid (0.1 μg/ml) was added 4 h before collection. Cells were then used for Q-FISH and chromosome orientation FISH (CO-FISH) analyses, as described previously[48–50]. Briefly, for Q-FISH, metaphase spreads were hybridized with Cy3-labeled PNA $(CCCTAA)_3$ (Panagene). Images were captured using Cytovision software (Applied Imaging Corporation) on a fluorescence microscope (Axio2; Carl Zeiss); individual telomere fluorescence signals were quantified using TFL-Telo software (a kind gift from P. Lansdorp, University of Groningen). Metaphase spreads were scored for chromosome breaks and fusions, fragile telomeres (a chromatid with ≥2 telomere signals) and sister telomere losses (STLs) and signal-free ends (SFEs) (chromosomes or chromatids without detectable telomere signals). For CO-FISH, cells were hybridized with Alexa Fluor 488–labeled $(TTAGGG)_3$ and Cy3-labeled PNA $(CCCTAA)_3$ probes. A chromosome with more than two telomeric DNA signals using both probes was scored as T-SCE positive. For each individual, data were collected from at least 30 cells.

Telomere dysfunction–induced focus assays. Epstein-Barr virus (EBV)-immortalized cells derived from cryopreserved lymphocytes were spun onto Cytospin microscopic slides at 8*g* for 3 min. Cells were fixed in 2% paraformaldehyde, permeabilized with 0.5% Triton X-100 and blocked with 10% FBS for 1–2 h. Cells were stained with a rabbit antibody to γH2AX (1:200 dilution; sc-101696, Santa Cruz Biotechnology)[51] followed by Alexa Fluor 488–labeled secondary antibody (1:500 dilution; Molecular Probes). For combined telomere FISH, slides were fixed in 2% paraformaldehyde and then used to perform telomere FISH. z-stack images were captured on a fluorescence microscope (Axiovert 200M, Carl Zeiss). Experiments were repeated three times on each sample.

Telomeric G overhang measurements. Two methods were used to measure the length and abundance of G overhangs. The length of the telomeric G overhang was measured by telomere overhang protection assay as described[52]. The relative abundance of G overhangs was measured by non-denaturing

doi:10.1038/ng.2941

in-gel hybridization as described[52] with minor modifications. Briefly, 1 µg of genomic DNA was resuspended in 15 µl of gp32 protection buffer and treated with 1 µg of RNase A and with or without Exo1 at 37 °C for 1 h. Samples were size fractionated on a 0.8% agarose gel in 0.5× TBE (Tris-borate-EDTA) at 8 V/cm. The gel was then dried at 50 °C for 1 h. The dried gel was rinsed with 2× SSC (saline–sodium citrate) and then hybridized with the high-specificity C-rich probe at 42 °C overnight. After two washes with 0.1% SDS–0.1× SSC for 15 min per wash, the gel was exposed to a phosphorimager screen, and signals were acquired by Storm860 PhosphorImager and quantified with ImageQuant software. The gel was then denatured in 0.5 M NaOH–1.5 M NaCl for 1 h at room temperature, neutralized in 1.5 M NaCl–0.5 M Tris-HCl (pH 8.0) for 15 min and hybridized to C-rich probe again at 42 °C overnight. After two washes with 0.1× SSC–0.1% SDS for 15 min per wash, the gel was exposed to a phosphorimager screen. Relative amounts of overhang were calculated by normalizing signals from the native gel to signals from the denatured gel. Experiments were repeated three times on each sample.

Cell lysate preparation and telomerase assays. S-100 extracts from PBMCs were prepared, and 5 µg and 1 µg of cell extract were assayed for the presence of telomerase activity using the telomere repeat amplification protocol (TRAP) assay as previously described[53]. RNase A treatment was used for negative controls.

31. Tucker, M.A. *et al.* A natural history of melanomas and dysplastic nevi: an atlas of lesions in melanoma-prone families. *Cancer* **94**, 3192–3209 (2002).
32. Liang, X.S. *et al.* Common genetic variants in candidate genes and risk of familial lymphoid malignancies. *Br. J. Haematol.* **146**, 418–423 (2009).
33. Chaudru, V. *et al.* Influence of genes, nevi, and sun sensitivity on melanoma risk in a family sample unselected by family history and in melanoma-prone families. *J. Natl. Cancer Inst.* **96**, 785–795 (2004).
34. Bodelon, C. *et al.* On the interplay of telomeres, nevi and the risk of melanoma. *PLoS ONE* **7**, e52466 (2012).
35. Maccioni, L. *et al.* Variants at chromosome 20 (*ASIP* locus) and melanoma risk. *Int. J. Cancer* **132**, 42–54 (2013).
36. Landi, M.T. *et al.* A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am. J. Hum. Genet.* **85**, 679–691 (2009).
37. Landi, M.T. *et al.* Environment and Genetics in Lung cancer Etiology (EAGLE) study: an integrative population-based case-control study of lung cancer. *BMC Public Health* **8**, 203 (2008).
38. Ballew, B.J. *et al.* Germline mutations of regulator of telomere elongation helicase 1, *RTEL1*, in Dyskeratosis congenita. *Hum. Genet.* **132**, 473–480 (2013).
39. DePristo, M.A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
40. 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).
41. Browning, B.L. & Browning, S.R. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.* **84**, 210–223 (2009).
42. Kong, A. *et al.* Detection of sharing by descent, long-range phasing and haplotype imputation. *Nat. Genet.* **40**, 1068–1075 (2008).
43. Genin, E., Tullio-Pelet, A., Begeot, F., Lyonnet, S. & Abel, L. Estimating the age of rare disease mutations: the example of Triple-A syndrome. *J. Med. Genet.* **41**, 445–449 (2004).
44. Krumm, N. *et al.* Copy number variation detection and genotyping from exome sequence data. *Genome Res.* **22**, 1525–1532 (2012).
45. Notredame, C., Higgins, D.G. & Heringa, J. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217 (2000).
46. 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).
47. Hathcock, K.S., Hodes, R.J. & Weng, N.P. Analysis of telomere length and telomerase activity. *Curr. Protoc. Immunol.* Chapter 10, Unit 10.30 (2004).
48. Zijlmans, J.M. *et al.* Telomeres in the mouse have large inter-chromosomal variations in the number of T_2AG_3 repeats. *Proc. Natl. Acad. Sci. USA* **94**, 7423–7428 (1997).
49. Wang, Y. *et al.* An increase in telomere sister chromatid exchange in murine embryonic stem cells possessing critically shortened telomeres. *Proc. Natl. Acad. Sci. USA* **102**, 10256–10260 (2005).
50. Bailey, S.M., Goodwin, E.H. & Cornforth, M.N. Strand-specific fluorescence *in situ* hybridization: the CO-FISH family. *Cytogenet. Genome Res.* **107**, 14–17 (2004).
51. Vallabhaneni, H., O'Callaghan, N., Sidorova, J. & Liu, Y. Defective repair of oxidative base lesions by the DNA glycosylase Nth1 associates with multiple telomere defects. *PLoS Genet.* **9**, e1003639 (2013).
52. Chai, W., Shay, J.W. & Wright, W.E. Human telomeres maintain their overhang length at senescence. *Mol. Cell. Biol.* **25**, 2158–2168 (2005).
53. Liu, Y. *et al.* The telomerase reverse transcriptase is limiting and necessary for telomerase function *in vivo*. *Curr. Biol.* **10**, 1459–1462 (2000).