ARTICLE

Genome-wide linkage analysis in Spanish melanoma-prone families identifies a new familial melanoma susceptibility locus at 11q

Miriam Potrony [1,2] · Joan Anton Puig-Butille[2,3] · James M. Farnham[4] · Pol Giménez-Xavier[1,2] · Celia Badenas[2,3] · Gemma Tell-Martí[1,2] · Paula Aguilera[1,2] · Cristina Carrera[1,2] · Josep Malvehy[1,2] · Craig C. Teerlink[4] · Susana Puig[1,2]

Abstract
The main genetic factors for familial melanoma remain unknown in >75% of families. *CDKN2A* is mutated in around 20% of melanoma-prone families. Other high-risk melanoma susceptibility genes explain <3% of families studied to date. We performed the first genome-wide linkage analysis in *CDKN2A*-negative Spanish melanoma-prone families to identify novel melanoma susceptibility loci. We included 68 individuals from 2, 3, and 6 families with 2, 3, and at least 4 melanoma cases. We detected a locus with significant linkage evidence at 11q14.1-q14.3, with a maximum het-TLOD of 3.449 (rs12285365: A>G), using evidence from multiple pedigrees. The genes contained by the subregion with the strongest linkage evidence were: *DLG2, PRSS23, FZD4*, and *TMEM135*. We also detected several regions with suggestive linkage evidence (TLOD >1.9) (1q, 6p, 7p, 11q, 12p, 13q) including the region previously detected in melanoma-prone families from Sweden at 3q29. The family-specific analysis revealed three loci with suggestive linkage evidence for family #1: 1q31.1-q32.1 (max. TLOD 2.447), 6p24.3-p22.3 (max. TLOD 2.409), and 11q13.3-q21 (max. TLOD 2.654). Future next-generation sequencing studies of these regions may allow the identification of new melanoma susceptibility genetic factors.

These authors contributed equally: Miriam Potrony, Joan Anton Puig-Butille.

These authors jointly supervised this work: Craig C. Teerlink, Susana Puig.

Electronic supplementary material The online version of this article (https://doi.org/10.1038/s41431-018-0149-8) contains supplementary material, which is available to authorized users.

✉ Susana Puig
    susipuig@gmail.com

1 Department of Dermatology, Melanoma Unit, Hospital Clínic de Barcelona, IDIBAPS, Universitat de Barcelona, Barcelona, Spain

2 Centro de Investigación Biomédica en Red en Enfermedades Raras (CIBERER), Barcelona, Spain

3 Department of Biochemistry and Molecular Genetics, Melanoma Unit, Hospital Clínic de Barcelona, IDIBAPS, Universitat de Barcelona, Barcelona, Spain

4 Department of Medicine, Division of Genetic Epidemiology, University of Utah School of Medicine, Salt Lake City, UT, USA

Introduction

Melanoma etiology is complex and involves environmental, phenotypic, and genetic factors. Approximately 10% of melanoma cases occur in a familial context. To date, *CDKN2A* (NG_007485.1, NM_000077.4 (p16INK4A) and NM_058195.3 (p14ARF), LRG_11) is the main high-risk susceptibility gene and germline pathogenic variants are detected in around 20% of melanoma-prone families worldwide [1]. The prevalence of *CDKN2A* pathogenic variants varies across populations (5–72%) [2]. In the Mediterranean population, due to the low incidence of the disease, melanoma-prone families are considered as those with at least two melanoma patients in first- or second-degree relatives [3, 4]. Overall, 14% of Spanish melanoma-prone families carry *CDKN2A* pathogenic variants, with prevalence increasing with the number of cases in the family: 11% in families with 2 cases, 23% in families with 3 cases, and 36–43% in families with at least 4 melanoma cases [5, 6].

Beyond *CDKN2A*, other high-risk melanoma genes have been identified, but they account for <3% of the families studied worldwide [1]. Thus, the genetic factors underlying melanoma susceptibility remain unknown in a substantial number of high-risk melanoma families [1].

Previous genome-wide linkage analyses, either using microsatellite marker sets or high-density single-nucleotide polymorphism (SNP) arrays, have been conducted in *CDKN2A* wild-type melanoma-prone families, mostly from pedigrees of Northern European ancestry [7–9]. Altogether, these studies suggest 1p22, 9q21, and 17p12-p11 as melanoma susceptibility loci. Notably, the regions detected in these studies were restricted to each geographic population without overlap between studies, and results typically achieved suggestive evidence for linkage. To date, only one study has been conducted in Mediterranean melanoma pedigrees from Italy [10], which failed to detect results with suggestive or significant linkage evidence.

With the goal of identifying new familial melanoma susceptibility loci, we report the first genome-wide linkage analyses conducted in Spanish melanoma-prone families. This is the first study carried out in Mediterranean melanoma pedigrees that has been able to detect genomic regions reaching significant genome-wide linkage evidence.

Subjects and methods

Samples and pedigrees

The study included 29 melanoma cases and 39 non-affected individuals belonging to 11 Spanish melanoma-prone families (10 *CDKN2A*-negative families and one family with *CDKN2A*-positive and two *CDKN2A*-negative cases), with genome-wide genotyping data available from at least two melanoma cases (Figure S1). The family set was enriched with families with a high number of cases for our geographic location: six families with ≥4 melanoma cases, three families with 3 melanoma cases, and two families with 2 melanoma cases. All patients belonged to melanoma-prone families under dermatological follow-up at the Melanoma Unit of Hospital Clinic of Barcelona. For family and individual de-identification, the families included in the study were numbered consecutively from 1 to 11 and sex has been hidden on purpose.

The study was approved by the ethical committee of Hospital Clinic of Barcelona. All patients provided written, informed consent.

Linkage analysis

Subjects were genotyped on either the HumanOmni2.5 (Illumina) array versions v1.0 (81% of subjects) or v1.1 (19% of subjects). The GEO accession number for the genotyping data reported in this paper is GSE109208. Only SNPs common to both versions were included in the study (2,426,511 SNPs). We also excluded SNPs with missing genotypes in >95% of samples (2,332,767 SNPs remaining). Since linkage disequilibrium between markers can artificially inflate evidence for linkage [8], we reduced the set of markers to a non-linkage disequilibrium set by iteratively removing markers with heterozygosity <0.3, $r^2 > 0.16$ with a previously selected marker and a minimum distance of 0.1 cm between markers, which resulted in 24,225 SNPs for analysis [8].

Mcsim software was used to perform parametric linkage analysis. Mcsim uses Monte Carlo Markov Chain techniques to provide haplotype reconstructions to extract inheritance information in pedigrees [11]. In addition to standard multipoint logarithm of the odds (LOD) scores, the program calculates robust multipoint LOD scores (referred to as TLODs). TLOD score is preferable to standard multipoint because it incorporates the recombination frequency (theta) in the statistic's parameterization, preserving the robustness of the two-point LOD statistic to model misspecification while taking advantage of multipoint information [12]. The TLOD statistic follows the same theoretical distribution as other LOD score statistics (e.g., two-point, multipoint, and heterogeneity-LOD (het-LOD) scores) and can be interpreted with the same conventions. Lander and Kruglyak proposed using LOD >0.588 for nominal evidence, >1.9 for suggestive evidence, and >3.3 for significant evidence [13]. Evidence from multiple pedigrees was assessed with the heterogeneity-TLOD statistic (het-TLOD) [14]. Allele frequencies were estimated internally and general dominant and recessive models were used. We analyzed all pedigrees using an affected-only model that assumed a disease gene frequency of 0.005 for a dominant model and 0.05 for a recessive model. The penetrance estimates for carriers and non-carriers were 0.5 and 0.0005, respectively. The genome version GRCh37/hg19 was used to establish genomic positions. Reference sequence (RefSeq) database at NCBI and GeneCards Human Gene Database (http://www.genecards.org/) were used to obtain information on the genomic features in the regions of interest [15, 16].

Haplotype phasing

Regions of interest linked to multiple families were assessed for the presence of common haplotypes shared between linked families. The software SHAPEIT2 was used for haplotype phasing [17]. All genotyped SNPs in regions of interest were used for phasing.

Results

Genome-wide linkage analysis

The het-TLOD genome-wide analysis, using evidence summed across the pedigrees, revealed a region with

Table 1 Genome-wide suggestive het-TLODs (>1.9) and significant het-TLODs (>3.3)

Cytogenetic band	Genome location (GRCh37/hg19)	Model	Maximum score	SNP ID	Gene
1q31.1-q32.1	chr1: 186,946,386–205,336,875	Dom	2.456	rs7517688	None
3q29	chr3: 194,219,913–197,744,198	Rec	2.099	rs11185544	None
6p24.3-p23	chr6: 8,193,128–15,089,151	Dom	2.024	rs6925772	*HIVEP1*
7q21.11-q21.2	chr7: 85,216,272–91,167,397	Dom	2.546	rs10268943	None
11q14.1-q14.3	**chr11: 82,498,536–92,106,781**	**Dom**	**3.449**	**rs12285365**	***PRSS23***
11q22.1	chr11: 97,804,083–100,498,349	Dom	2.462	rs17577073	*CNTN5*
12p13.1	chr12: 13,066,220–13,617,099	Rec	1.950	rs12815655	None
13q12.3-q14.11	chr13: 30,484,862–43,478,867	Rec	2.365	rs2312972	None

The threshold used to define the boundary region was 1 het-TLOD difference from the maximum regional score. The significant het-TLODs (>3.3) are highlighted in bold

Dom dominant, *Rec* recessive

Fig. 1 Genome-wide het-TLOD scores. Genome-wide het-TLOD scores in dominant (continuous line) and recessive (dashed line) models are plotted. Significant linkage evidence threshold (het-TLOD >3.3) is denoted by the horizontal dashed line
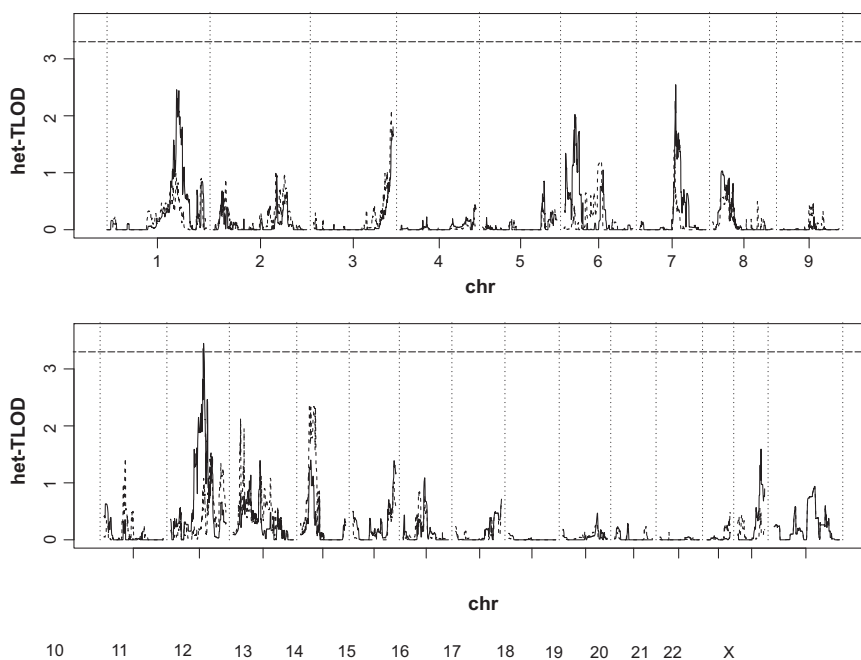

significant linkage (het-TLOD >3.3) on chromosome 11 (Table 1 and Fig. 1). This region had a maximum het-TLOD of 3.449 (rs12285365:A>G) and spanned the 11q14.1-q14.3 locus (when using one het-TLOD score support interval). The region contains 52 genomic features of which 38 are protein-coding genes. The strongest linkage evidence at this locus (all markers with het-TLOD >3.3) was detected in two regions: between rs1940085:G>A and rs7108021:T>G (chr11: 84.3–84.6 Mb) and between rs12285365:A>G and rs607530:T>C (chr11: 86.6–87.6 Mb). These regions contain four protein-coding genes: *DLG2* (NG_021375.1, NM_001142699.1), *PRSS23* (NM_007173.5), *FZD4* (NG_011752.1, NM_012193.3), and *TMEM135* (NM_022918.3). We phased haplotypes in the linked pedigrees using all available markers to determine whether the same haplotypes appeared in multiple linked families, but

failed to identify any such haplotype. Other regions showed suggestive linkage evidence (TLOD >1.9) at chromosome 1q, 6p, 7p, and 11q under a dominant model, and at chromosome 3q, 12p, and 13q under a recessive model (Table 1). SNPs description based on a genomic reference sequence is shown in Table S1.

Family-specific genome-wide linkage analysis

Data from previous studies suggest that certain high-risk melanoma factors may be restricted to a limited number of pedigrees such as germinal variants in *TERT* (NG_009265.1, NM_198253.2, LRG_343) [18, 19]. Thus, we conducted a separate genome-wide analysis for each family. We detected three regions with suggestive evidence for linkage (TLOD >1.9) in family #1 under a dominant

Fig. 2 Genome-wide TLOD scores for family #1. Genome-wide TLOD scores for dominant (continuous plot line) and recessive (dashed plot line) models are plotted. The suggestive linkage evidence threshold (TLOD >1.9) is denoted by the horizontal dashed line
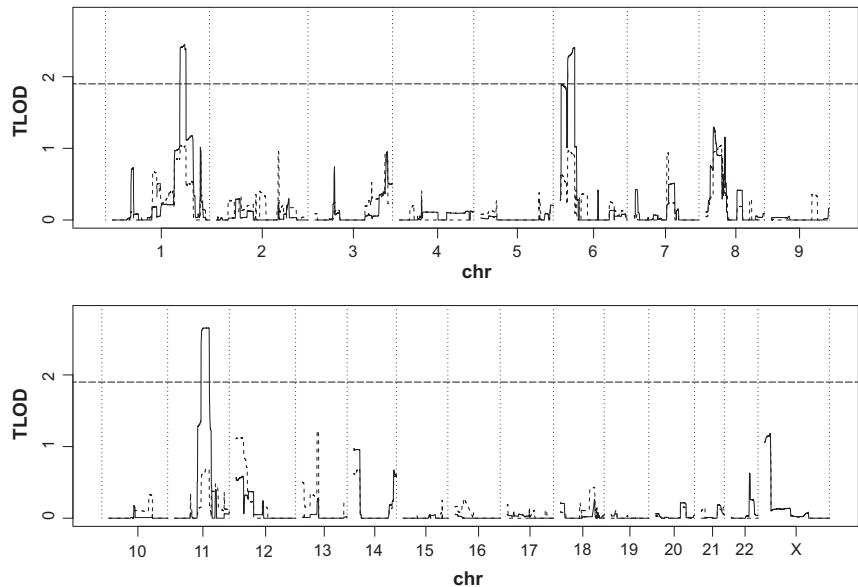
model (Fig. 2). This is a family with six *CDKN2A*-positive melanoma cases, two *CDKN2A*-negative melanoma cases, and other cancers in blood relatives (liver, lung, cervix, endometrial, and breast cancer cases). The *CDKN2A*-negative cases developed melanoma at a young age (32, 33 years old (y.o)) similar to the *CDKN2A*-positive cases (27, 34, 37, 37 y.o). Since, *CDKN2A*-negative cases did not carry medium melanoma risk variants such as *MC1R* (NG_012026.1, NM_002386.3) red-hair color variants or *MITF* (NG_011631.1, NM_000248.3, LRG_776) variant c.952G>A (p.(Glu318Lys)), we hypothesized that the melanoma risk observed in *CDKN2A*-negative cases may result from other melanoma susceptibility variants. We genotyped the *CDKN2A*-negative melanoma cases along with two *CDKN2A*-positive melanoma cases. The analyses identified three regions that segregate with all melanoma cases independently of *CDKN2A* status. The first region was in 1q31.1-q32.1 (chr1: 187.5–205.3 Mb) with a maximum TLOD of 2.447 at markers rs2246083:G>A and rs11590469:C>T (Figure S2). This region spans 17.8 Mb and contains 133 genomic features, of which 103 are protein-coding genes. The second region was in 6p24.3-p22.3 (chr6: 8.2–19.5 Mb) with a maximum TLOD of 2.409 at marker rs4712415:T>C (Figure S3). The region spans 11.3 Mb and contains 63 genomic features, of which 44 are coding protein genes. The third region was in 11q13.3-q21 (chr11: 68.7–95.5 Mb) with a maximum TLOD of 2.654 spanning >100 markers (Figure S4). This region spans 26.8 Mb and contains 239 genetic features, of which 171 are protein-coding genes.

Discussion

In Spain, the genetic background in melanoma-prone families remains unknown in >80% of families [5, 6].

Linkage analysis is likely to detect regions containing high-risk variants or genetic features segregating with the disease. Here, we report the results of a genome-wide linkage screen performed on 11 melanoma-prone families in which we detected significant linkage to the 11q14.1-q14.3 locus for melanoma susceptibility. Although the number of families included in the study is lower than previous studies, the subset of families was enriched by inclusion of highly informative families since 54.4% families had ≥4 melanoma cases.

A previous genome wide association study (GWAS) study performed in melanoma patients reported a melanoma locus at the 11q14.3 region. The study detected the strongest evidence of association near rs1393350:G>A encompassing *TYR* (NG_008748.1, NM_000372.4) gene, which plays a key role in human pigmentation and is a low-risk melanoma gene [20]. In the present study, the two subregions with strongest linkage evidence within 11q14.1-q14.3 do not include the *TYR* gene, suggesting that this genomic region is associated with melanoma susceptibility due to genetic factors other than pigment related alleles in the *TYR* gene. The *DLG2*, *PRSS23*, *FZD4*, and *TMEM135* genes are located in the regions with the strongest linkage evidence. The biological information about this set of genes is limited, but they are all plausible candidates for cancer susceptibility [21, 22]. However, further sequencing data and molecular studies are necessary to elucidate the possible role of these genes in melanoma susceptibility.

Moreover, we have detected seven additional loci (1q31.1-q32.1, 3q29, 6p24.3-p23, 7q21.11-q21.2, 11q22.1, 12p13.1, 13q12.3-q14.11) with suggestive linkage evidence in the studied families. Notably, the 3q28-q29 locus has been previously detected with suggestive evidence of melanoma linkage in *CDKN2A* wild-type Swedish families [9].

A subsequent analysis of those families reported a narrower region spanning 3.5 Mb (chr3: 192.1–195.6 Mb) [23], overlapping the linked region detected in Spanish families. The rest of the suggestive regions detected have not been previously reported. The finding of a common region in Spanish and Swedish melanoma-prone families, strongly suggests that this region may contain genetic factors associated with melanoma susceptibility. The overlap region from both populations contains 20 genetic features, of which 10 are protein-coding genes (Table S2) including plausible candidates involved in proliferation and apoptosis, lipid transport, serin/threonin phosphatase PP1 inhibition, or Notch activation [24–27].

Melanoma is one of the tumors with highest heritability [28]. In families with melanoma aggregation, melanoma susceptibility follows an autosomal dominant inheritance pattern with incomplete penetrance. Multiple genes can play a role in melanoma susceptibility in a family, by combination of high-risk gene/s and presence of medium-/low-risk variants modulating expressivity of the high-risk gene/s. High-risk variants or genetic features segregate with the disease in most affected cases in the family and may be detected by linkage analysis studies. We expect to identify one or very few high-risk variants in an individual. However, in >70% of families worldwide, these have still not been identified. Thus, studies such as the present one are needed to provide clues to new genomic regions to focus on in order to identify new high-risk variants that may explain part of the missing heritability of melanoma susceptibility. The combination of medium-/low-risk variants modulates the penetrance and expressivity of high-risk genes, but may vary within the family and may be inherited from different ancestors. Multiple medium-/low-risk variants have been described to date [1]. However, their specific role in the modulation of the expressivity of pathogenic variants in high-risk genes has only been well established for the highly polymorphic pigmentation gene *MC1R*. *CDKN2A* variant carriers with melanoma-associated variants in *MC1R* have an increased risk of developing melanoma than *CDKN2A* variant carriers with wild-type *MC1R* [29]. Although co-existence of a *CDKN2A* pathogenic variant with the rare *MITF* c.952G>A (p.(Glu318Lys)) variant has also been described [30], the implication of *MITF* in the modulation of melanoma penetrance in *CDKN2A* variant carriers is still unknown.

In our study, we included a *CDKN2A*-positive family in which two melanoma cases did not carry known high-risk nor medium-risk melanoma susceptibility variants. We detected three loci with suggestive linkage evidence indicating that, in addition to the *CDKN2A* pathogenic variant, other genetic factors underlie the increased melanoma risk observed in the members of this family. Knowing the gene, or combination of genes, involved in melanoma susceptibility is crucial for identification and better management of at-risk individuals. Furthermore, it allows the refinement of genetic counseling in melanoma, as specific measures can be included when genetic testing detects germline variants in known susceptibility genes [5, 30–32].

In conclusion, using linkage evidence from multiple pedigrees, we have identified a familial melanoma susceptibility locus at 11q14.1-q14.3, in Spanish melanoma-prone families. We have also detected suggestive evidence of linkage at 3q29, previously described in Swedish families. Future next-generation sequencing studies or candidate gene targeted sequencing from these regions may allow the identification of new genetic factors implicated in melanoma susceptibility.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Potrony M, Badenas C, Aguilera P, et al. Update in genetic susceptibility in melanoma. Ann Transl Med. 2015;3:210.
2. Goldstein AM, Chan M, Harland M, et al. Features associated with germline CDKN2A mutations: a GenoMEL study of melanoma-prone families from three continents. J Med Genet. 2007;44:99–106.
3. Badenas C, Aguilera P, Puig-Butille JA, Carrera C, Malvehy J, Puig S. Genetic counseling in melanoma. Dermatol Ther. 2012;25:397–402.

4. Leachman SA, Carucci J, Kohlmann W, et al. Selection criteria for genetic assessment of patients with familial melanoma. J Am Acad Dermatol. 2009;61:677.e1–14.

5. Potrony M, Puig-Butille JA, Aguilera P, et al. Increased prevalence of lung, breast, and pancreatic cancers in addition to melanoma risk in families bearing the cyclin-dependent kinase inhibitor 2A mutation: implications for genetic counseling. J Am Acad Dermatol. 2014;71:888–95.

6. Puig S, Potrony M, Cuellar F, et al. Characterization of individuals at high risk of developing melanoma in Latin America: bases for genetic counseling in melanoma. Genet Med. 2016;18:727–36.

7. Gillanders E, Juo SH, Holland EA, et al. Localization of a novel melanoma susceptibility locus to 1p22. Am J Hum Genet. 2003;73:301–13.

8. Cannon-Albright LA, Teerlink CC, Farnham JM, Thomas AW, Zone JJ, Leachman SA. Linkage analysis of extended high-risk pedigrees replicates a cutaneous malignant melanoma predisposition locus on chromosome 9q21. J Invest Dermatol. 2013;133:128–34.

9. Hoiom V, Tuominen R, Hansson J. Genome-wide linkage analysis of Swedish families to identify putative susceptibility loci for cutaneous malignant melanoma. Genes Chromosomes Cancer. 2011;50:1076–84.

10. Kerstann KF, Bradford PT, Steighner R, et al. No evidence for linkage with melanoma in Italian melanoma-prone families. Cancer Epidemiol Biomarkers Prev. 2008;17:1838–40.

11. Thomas A, Gutin A, Abkevich V, Bansal A. Multilocus linkage analysis by blocked Gibbs sampling. Stat Comput. 2000;10:259–69.

12. Goring HH, Terwilliger JD. Linkage analysis in the presence of errors II: marker-locus genotyping errors modeled with hypercomplex recombination fractions. Am J Hum Genet. 2000;66:1107–18.

13. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet. 1995;11:241–7.

14. Ott J. Linkage probability and its approximate confidence interval under possible heterogeneity. Genet Epidemiol Suppl. 1986;1:251–7.

15. Pruitt KD, Brown GR, Hiatt SM, et al. RefSeq: an update on mammalian reference sequences. Nucleic Acids Res. 2014;42: D756–63.

16. Safran M, Dalah I, Alexander J, et al. GeneCards Version 3: the human gene integrator. Database. 2010;2010:baq020.

17. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods. 2013;10:5–6.

18. Harland M, Petljak M, Robles-Espinoza CD, et al. Germline TERT promoter mutations are rare in familial melanoma. Fam Cancer. 2016;15:139–44.

19. Horn S, Figl A, Rachakonda PS, et al. TERT promoter mutations in familial and sporadic melanoma. Science. 2013;339:959–61.

20. Bishop DT, Demenais F, Iles MM, et al. Genome-wide association study identifies three loci associated with melanoma risk. Nat Genet. 2009;41:920–5.

21. Zhu J, Shang Y, Zhang M. Mechanistic basis of MAGUK-organized complexes in synaptic development and signalling. Nat Rev Neurosci. 2016;17:209–23.

22. Exil VJ, Silva Avila D, Benedetto A, et al. Stressed-induced TMEM135 protein is part of a conserved genetic network involved in fat storage and longevity regulation in *Caenorhabditis elegans*. PLoS ONE. 2010;5:e14228.

23. Tuominen R, Jonsson G, Enerback C, et al. Investigation of a putative melanoma susceptibility locus at chromosome 3q29. Cancer Genet. 2014;207:70–4.

24. Sullivan KD, Nakagawa A, Xue D, Espinosa JM. Human ACAP2 is a homolog of *C. elegans* CNT-1 that promotes apoptosis in cancer cells. Cell Cycle. 2015;14:1771–8.

25. Miranda E, Vizoso F, Martin A, et al. Apolipoprotein D expression in cutaneous malignant melanoma. J Surg Oncol. 2003;83:99–105.

26. Mahajan K, Mahajan NP. ACK1/TNK2 tyrosine kinase: molecular signaling and evolving role in cancers. Oncogene. 2015;34:4162–7.

27. Yu H, Takeuchi M, LeBarron J, et al. Notch-modifying xylosyl-transferase structures support a SNi-like retaining mechanism. Nat Chem Biol. 2015;11:847–54.

28. Mucci LA, Hjelmborg JB, Harris JR, et al. Familial risk and heritability of cancer among twins in Nordic countries. JAMA. 2016;315:68–76.

29. Demenais F, Mohamdi H, Chaudru V, et al. Association of MC1R variants and host phenotypes with melanoma risk in CDKN2A mutation carriers: a GenoMEL study. J Natl Cancer Inst. 2010;102:1568–83.

30. Potrony M, Puig-Butille JA, Aguilera P, et al. Prevalence of MITF p.E318K in patients with melanoma independent of the presence of CDKN2A causative mutations. JAMA Dermatol. 2016;152:405–12.

31. Helgadottir H, Hoiom V, Jonsson G, et al. High risk of tobacco-related cancers in CDKN2A mutation-positive melanoma families. J Med Genet. 2014;51:545–52.

32. Leachman SA, Lucero OM, Sampson JE, et al. Identification, genetic testing, and management of hereditary melanoma. Cancer Metastasis Rev. 2017;36:77–90.