SUPPORTING INFORMATION

SUPPLEMENTARY METHODS

Study participants

A total of 96 cancer-free unrelated Spanish individuals were recruited at the Department of Dermatology of two hospitals from Castellon Province, Spain (Castellon University General Hospital and La Plana University Hospital). All individuals provided a written informed consent. The study was approved by the Ethics Committee of the Jaume I University of Castellon (Spain).

Phenotypic data collection

Each participant completed a standardised questionnaire to collect information on sex, age, pigmentation traits (skin, hair and eye colour), Fitzpatrick's skin type classification, presence of pigmented spots, and personal and family cancer history. Detailed information related to oestrogen-related hyperpigmentation (dark spots, chloasma, linia nigra and/or melasma) during gestation or after the use of oestrogen-containing oral contraceptives was also recorded. Each participant completed the questionnaire under the supervision of a professional.

Tissue sampling

Volunteer participants donated a fresh-frozen normal skin sample obtained from the margin of skin excision biopsies undertaken to remove a cutaneous lesion. Thirty-nine samples were obtained from chronically sun-exposed skin areas (neck, face and hands), and 57 samples were collected from skin areas with intermittent sun exposure (back, chest, legs and upper arms).

Immediately after resection, tissue samples were submerged in RNA*later* Tissue Collection Solution (Thermo Fisher Scientific, Walham, MA, USA) and stored at 4ºC overnight. Then, the epidermis was separated from the dermis by incubating the tissue sample in 3.8% ammonium thiocyanate (Sigma-Aldrich, St Louis, MO, USA) in PBS (pH 7.4) at room temperature for 3 hours. Subsequently, the epidermis was immersed in RNA*later* solution and stored at -20ºC until sample processing.

Tissue sampling was performed under protocols approved by the Clinical Research Ethics Committee of the Castellon University General Hospital, and the Clinical Research Ethics Committee of the La Plana University Hospital.

Isolation of genomic DNA

Genomic DNA was isolated from epidermal samples with the QIAamp DNA Mini Kit, (Qiagen, Hilden, Germany). After DNA extraction, all samples were diluted to a concentration of 5 ng/µl in order to prepare them for genotyping.

SNP selection and genotyping

A total of 8 SNPs were selected to be genotyped, in order to entirely capture the *PDZK1* gene region, with a genomic sequence length of 36.5 kb. *PDZK1* germline variants, SNP codes, locations, derived and ancestral alleles and their frequencies were obtained from the Ensembl Variation database (www.ensembl.org/info/genome/variation). We included only SNPs with a reported minor allele frequency (MAF) greater than 5% in south-European populations from the International 1000 Genomes Project (http://www.1000genomes.org/), and with a minimum pairwise r^2 threshold of 0.8 (apart

from rs9726452 and rs9728619 that are in absolute linkage disequilibrium (r^2 =1.0) and were used for internal genotyping control).

SNP genotyping was conducted by the Spanish National Genotyping Centre (CEGEN-PRB2-ISCIII, Santiago de Compostela, Spain) as a contract service using the iPLEX Gold MassARRAY technology, according to the manufacturer's protocol (Sequenom, San Diego, CA, USA). Genotyping specificity was assessed by adding three DNA duplicates (two intra-assays and one inter-assay) per plate, yielding 100% consistent replication results.

Gene expression analysis

Total RNA was isolated from fresh-frozen epidermal samples using the RNeasy Mini Kit (Qiagen), according to manufacturer's instructions. Reverse transcription of each RNA sample was performed using the High Capacity cDNA Reverse Transcription Kit. The quantitative real-time PCR was performed on a StepOnePlus™ Real-Time PCR System using the TaqMan® Gene Expression Master Mix (Thermo Fisher Scientific). Beta-actin was used as internal control for data normalisation. TaqMan® Gene Expression Assays used were Hs00275727_m1 for *PDZK1* and Hs01060665_g1 for β-actin (Thermo Fisher Scientific). Gene expression analyses were performed in triplicate per sample, and included a negative control. All standard errors were less than 5% of the respective sample mean. The delta-delta Ct method was used to calculate the relative changes in *PDZK1* expression levels between different population subsets (S1).

Statistical analysis

Statistical analyses were performed using the R software (http://www.R-project.org), and two-sided *P*-values lower than 0.05 were considered as statistically significant. Unknown and missing values were excluded at each specific analysis.

Shapiro-Wilk test was used to test for normality of expression data distribution. One-way ANOVA test was used to account for differences in mRNA *PDZK1* levels between individuals included in different subsets, according to specific variables. Tukey's HSD was used for post-hoc pairwise comparisons. All genetic analyses were performed estimating the effect of the minor allele in the population.

Table S1. Distribution of phenotypes in the population according to the degree of sun exposure in each individual's epidermis

		Chronically sun-exposed (N=39)		Intermittently sun-exposed (N=56)	
		N	**%**	**N**	**%**
Sex	Females‡	11	28.21	34	60.71
	Males	28	71.79	22	39.29
	NA	0	0.00	0	0.00
Melasma in pregnancy	Absence	8	88.89	8	14.29
	Presence	1	11.11	8	14.29
	NA	0	0.00	0	0.00
Skin Color	Dark	17	43.59	24	42.86
	Fair/Pale	19	48.72	32	57.14
	NA	3	7.69	0	0.00
Skin phototype	III/IV	24	61.54	30	53.57
	I/II	12	30.77	26	46.43
	NA	3	7.69	0	0.00
Sun-induced spots	No	23	58.97	27	48.21
	Yes	13	33.33	27	48.21
	NA	3	7.69	2	3.57

N, number of individuals with each phenotype; %, percentage of individuals with each phenotype; NA, not available; p-value < 0.05 are presented in bold.

‡ Twenty-five out of 45 tissue-donating females had been pregnant at least once

Table S2. Differences in *PDZK1* expression levels according to *PDZK1* genotype

SNP rs#	Genomic Location	Genotype	MAF	ΔΔCt [†]	Fold-change [‡]	Non-adjusted *P*-value [§]	Adjusted *P*-value [¶]
rs12129861	1:145709377	TT	0.484	0.00 ± 0.50	1.00	0.082	0.050
		CT		0.51 ± 0.45	0.70		
		CC		-0.88 ± 0.75	1.85		
rs1298954	1:145704906	CC	0.307	0.00 ± 0.43	1.00	0.405	0.854
		CT		-0.48 ± 0.47	1.39		
		TT		1.99 ± 2.22	0.25		
rs9659930	1:145704365	CC	0.161	0.00 ± 0.39	1.00	0.980	0.827
		CG		-0.28 ± 0.57	1.22		
		GG		Not found	-		
rs9726452	1:145698757	TT	0.368	0.00 ± 0.45	1.00	0.857	0.321
		CT		-0.43 ± 0.46	1.35		
		CC		-0.78 ± 1.48	1.72		
rs9728619	1:145698627	GG	0.364	0.00 ± 0.44	1.00	0.904	0.365
		AG		1.20 ± 0.46	1.20		
		AA		1.55 ± 1.71	1.55		
rs72704277	1:145694551	AA	0.031	0.00 ± 0.11	1.00	0.094	0.059
		AG		1.55 ± 2.50	0.18		
		GG		Not found	-		
rs11576685	1:145691969	TT	0.078	0.00 ± 0.35	1.00	**0.016**	**0.030**
		CT		-1.71 ± 0.61	3.60		
		CC		Not found	-		
rs1284300	1:145687602	GG	0.068	0.00 ± 0.33	1.00	0.341	0.327
		AG		-0.74 ± 1.07	1.68		
		AA		Not found	-		

MAF, minor allele frequency; SEM, standard error of the mean.

[†] Homozygotes for major allele were set as reference group. Results are shown as mean ± SEM of the combined data from each population subset.

[‡] Ratio of gene expression change relative to the reference.

[§] *P*-value from the one-way ANOVA test. Bold indicates significant differences in *PDZK1* expression levels.

[¶] *P*-value from the one-way ANOVA test adjusted by sex and age. Bold indicates significant differences in *PDZK1* expression levels.

Tukey's HSD was used for post-hoc pairwise comparisons.

Two-sided P-values < 0.05 were considered as statistically significant.

SUPPLEMENTARY REFERENCES

S1. Livak K J, Schmittgen T D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods San Diego Calif 2001; 25:402–408.