## ONLINE METHODS

**Subjects.** Approval for the study was granted by the Icelandic National Bioethics Committee, the Icelandic Data Protection Authority and by local ethics committees for each of the non-Icelandic replication sample sets. The sample sets from Iceland, Eastern Europe, United States, Sweden and the Spanish CM cases and controls have been described previously (see **Supplementary Table 1**). All Icelandic BCC, SCC and CM cases had histologically confirmed diagnoses recorded by the Icelandic Cancer Registry, which has a nationwide catchment.

*Spain BCC.* BCC cases were recruited from the Oncology Department of Zaragoza Hospital between September 2007 and December 2008. Individuals with histologically proven invasive BCC were eligible to participate in the study. The median time interval from BCC diagnosis to collection of blood samples was 14 months (range 1–53 months). Median age at diagnosis was 69 years (range 21–91).

*Holland CM.* Individuals diagnosed with melanoma of the skin (ICD-O-3 code C44 and C80, morphology codes 8720-8790) in the period 2003–2007 were identified in the regional cancer registry held by the Comprehensive Cancer Centre East in Nijmegen, The Netherlands. This cancer center keeps a population-based cancer registry and covers the eastern part of the Netherlands, a region with 1.3 million inhabitants, one university clinic and seven community hospitals. All individuals diagnosed with melanoma at or before the age of 75 were invited to participate in the study. The invitation was done by the subjects' treating physicians (dermatologists and general or plastic surgeons) who all agreed to collaborate in this study. Informed consent was obtained for the collection of questionnaire data on lifestyle including sun exposure, medical history and family history, the collection of two 10 ml blood samples, possible linkage to population and disease registries (cancer registry, mortality registry, hospital information systems and the Dutch demographic register), collection of additional clinical data from their medical records, and the retention of identifying information for a duration of 25 years. The Comprehensive Cancer Centre East collects clinical and pathology data of all subjects in the cancer registry. In The Netherlands, lifestyle information, family history of cancer, reproductive and medical history as well as blood samples are available from a group of 6,700 population controls. These controls were collected in a survey in 2002–2003 by the Radboud University Nijmegen Medical Centre. This survey, The Nijmegen Biomedical Study, was based on an age-stratified random sample of the population of Nijmegen. From this group 1,832 male and female control individuals were selected and genotyped. Similar informed consent as described above was obtained from these controls.

*Austria CM.* After obtaining written informed consent, individuals attending the outpatient ward of the Department of Dermatology, Medical University of Vienna were invited to participate in the project. Subjects donated 3.4 ml of peripheral blood for DNA extraction. Information regarding skin phototype, sun exposure history and prior malignancies was assessed by interview questionnaire and from clinical records. In addition to the questionnaire, photographic documentation of the eyes, hair, back and arms (for assessment of nevi and sun exposure–related damage) was performed. Controls were recruited from individuals attending the Department of Dermatology for non-melanoma-related conditions. All subjects were of self-reported Austrian or central European descent. All samples and data were coded and archived anonymously. The study was approved by the ethics committee of the Medical University of Vienna (project number 59/2007).

*Italy CM.* Melanoma cases consisted of 389 individuals hospitalized for surgical treatment of melanoma at the Melanoma and Sarcoma Surgery Unit of the Fondazione IRCCS Istituto Nazionale Tumori, Milan between May 2006 and June 2007. A further 177 melanoma cases were individuals with advanced melanomas recruited for experimental immunotherapy trials in the same surgical unit between 1998 and 2008. Controls were healthy donors from the Immunohematology and Transfusion Medicine Department, Fondazione IRCCS Istituto Nazionale Tumori. All subjects were of Italian or European origin.

**Genotyping.** All Icelandic samples were typed using Illumina HumanHap300 or HumanCNV370-duo chips or by Nanogen Centaurus single-track genotyping assays as described previously[1]. For BCC, 930 Icelandic cases were typed on Illumina chips and the remaining 903 samples were typed by Centaurus assay. For non-Illumina SNPs, approximately 1,690 Icelandic BCC cases and 2,456 controls were genotyped using Centaurus assay. Primer sequences for Centaurus assays are available on request. Centaurus SNP assays were validated by genotyping the HapMap CEU samples and comparing genotypes with the published data. Assays were rejected if they showed $>1.5\%$ mismatches with the HapMap data. Approximately 10% of Icelandic case samples that were genotyped on the Illumina platform were also genotyped using the Centaurus assays and the observed mismatch rate was less than 0.5%. All samples from foreign cohorts were typed using the same Centaurus assays at the deCODE Genetics facility. Clustering algorithms were applied and manual editing was carried out in the same way as for the Icelandic samples. Two standard DNA samples and water blanks were included on every plate. Heterogeneity tests were used to monitor for deviant results originating from particular cohorts and the $P_{het}$ values are shown in the relevant results tables.

**Statistical analysis.** We calculated the OR for each SNP allele or haplotype assuming the multiplicative model, that is, that the relative risks of the two alleles that a person carries multiply. Allelic frequencies and OR are presented for the markers. The associated $P$ values were calculated with the standard likelihood ratio $\chi^2$ statistic. Confidence intervals were calculated assuming that the estimate of OR has a log-normal distribution. For SNPs that were in strong LD, whenever the genotype of one SNP was missing for an individual, the genotypes of the correlated SNPs were used to impute genotypes through a likelihood approach as previously described[1]. Some of the Icelandic cases and controls are related to each other, causing the $\chi^2$ statistic to have a mean $>1$. We estimated the inflation factor by simulating genotypes through the Icelandic genealogy and corrected the $\chi^2$ statistics for Icelandic OR's accordingly. The correction factors used were 1.20 for BCC, 1.07 for CM, 1.05 for SCC, 1.21 for CAD and 1.33 for T2D. Joint analyses of multiple case-control replication groups were carried out using a Mantel-Haenszel model in which the groups were allowed to have different population frequencies for alleles or genotypes but were assumed to have common relative risks. The tests of heterogeneity were performed by comparing the null hypothesis of the effect being the same in all populations to the alternative hypothesis of each population having a different effect using a likelihood ratio test. $I^2$ lies between 0% and 100% and describes the proportion of total variation in study estimates that is due to heterogeneity[29]. We calculated genotype-specific ORs by estimating the genotype frequencies in the population assuming Hardy-Weinberg equilibrium. No significant deviations from the multiplicative model were observed, all $P$ values being $>0.05$.

The population attributable risk (PAR) was calculated for each of nine variants as follows:

$$\mathrm{PAR} = 1 - (1 / ((1 - p)^2 + 2p\,(1 - p)\,\mathrm{OR} + p^2\,\mathrm{OR}^2))$$

where $P$ = frequency of the risk allele (estimated as the arithimetic average of the frequencies in each of the constituent population samples) and $OR$ = allelic odds ratio for the risk allele (combined from the constituent population samples using The Mantel-Haenszel method).

The joint population attributable risk (Joint PAR) for combinations of variants was calculated as follows:

$$\text{Joint PAR} = 1 - (\textstyle\prod_{1\to n}(1 - \mathrm{PAR_i}))$$

where $\mathrm{PAR}_i$ corresponds to the individual PAR for the $i$th SNP and $n$ is the number of variants considered. The full genotype-specific model did not provide a signficiantly better fit of the data than the multiplicative model for the variants reported here and for the variants previously reported[1,3]; therefore, we used the multiplicative model in determination of PAR. There was no significant heterogeneity in allele effect between constituent populations. To search for potential epistatic interactions, all pairwise combinations of the nine variants (a total of 36 tests) were tested using logistic regression, comparing models with and without an interaction term. The lowest $P$ value we obtained for any pairwise interaction was 0.03, which is not significant once the number of tests conducted is taken into account. Therefore, we assumed no epistatic interactions between loci in our calculation of joint PAR.

Combining the Icelandic genealogy and the method of long-range phasing[24] allowed us to determine the parental origins of the haplotypes in most of the Icelanders who were typed using an Illumina chip. In particular, long-range phasing was accomplished through identifying individuals who shared a long haplotype (identical by descent) with the proband, people referred to as surrogate parents. In general, there would be a group of surrogate parents for one haplotype and another group for the other haplotype, although in some

cases, only surrogate parents for one of the two haplotypes could be identified. For each haplotype of the proband, we determined, using the genealogy, the shortest meiotic distance to a surrogate parent through the father (minimum paternal distance) and the shortest distance through the mother (minimum maternal distance). For example, if the minimum paternal distance is substantially less than the minimum maternal distance, then the haplotype is likely to be inherited paternally. Moreover, the parental origins of the two haplotypes can be reliably determined if strong evidence exists for one of the two haplotypes. In general, a score is created by combining the results from both haplotypes. For the 1,118 cases and 34,869 controls in the BCC parent-of-origin analysis, only ten rs157935 heterozygotes in cases and 335 heterozygotes in controls had undetermined parental origins. To avoid bias, we included these individuals to estimate frequencies of the T allele in parental and maternal chromosomes for cases and controls, and likelihood approaches were used to properly take the incomplete information into account. Maximum likelihood and likelihood ratio tests were used to test/estimate the effect of the T allele when transmitted paternally and when transmitted maternally. To directly test whether the T allele confers an effect that depends on parental origin, we tested whether the number of ordered TG (first allele paternal, second allele maternal) heterozygotes (237) within the cases is greater than the number of ordered GT heterozygotes (182). If there is no parent-of-origin effect, then the two counts should be the same, in expectation. A binomial test gave a two-sided $P$ of 0.0083. Even though stratification is not an issue here, a variance adjustment is still needed to account for relatedness. Using genomic control, we converted the binomial $P$ to a 1-d.f. $\chi^2$ statistic and divided it by 1.04, the adjustment factor, resulting in a $P$ of 0.0096. By contrast, the number of TG heterozygotes and GT heterozygotes in controls showed no significance difference (7,452 versus 7,419, $P = 0.80$). All $P$ values are reported as two-sided.

**URLs.** Locations of epidermolysis bullosa variants and domain definitions for K5 were derived from the Human Intermediate Filament Database, http://www.interfil.org[30].

29. Higgins, J.P. & Thompson, S.G. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* **21**, 1539–1558 (2002).
30. Szeverenyi, I. *et al.* The Human Intermediate Filament Database: comprehensive information on a gene family involved in many human diseases. *Hum. Mutat.* **29**, 351–360 (2008).