ONLINE METHODS

Samples. Icelandic cases were ascertained through the Icelandic Cancer Registry or through national pathology registers as previously described[5]. Controls were drawn from non-cancer population-based projects conducted by deCODE genetics. Details of the non-Icelandic replication sample sets used are given in **Supplementary Table 2**.

Illumina SNP chip genotyping. The Icelandic chip-typed samples were assayed with the Illumina HumanHap300, HumanHap CNV370, HumanHap 610, 1M or Omni1-Quad BeadChips at deCODE genetics. Only the 317,503 SNPs from the HumanHap300 chip were used in the long-range phasing and subsequent SNP imputations. SNPs were excluded if they had (i) a yield lower than 95%, (ii) a minor allele frequency of less than 1% in the population or (iii) significant deviation from Hardy-Weinberg equilibrium in the controls ($P < 0.001$), (iv) if they produced an excessive inheritance error rate (>0.001) or (v) if there was substantial difference in allele frequency between chip types (from just a single chip if that resolved all differences but from all chips otherwise). All samples with a call rate <97% were excluded from the analysis. The final set of SNPs used for long-range phasing was composed of 297,835 autosomal SNPs.

Whole-genome sequencing and SNP calling. SNPs were imputed based on whole-genome sequence data from 457 Icelanders selected for various neoplastic, cardiovascular and psychiatric conditions. All of the individuals were sequenced at a depth of at least 10×. Approximately 16 million SNPs were imputed based on this set of individuals.

Sample preparation. Paired-end libraries for sequencing were prepared according to the manufacturer's instructions (Illumina). Briefly, approximately 5 μg of genomic DNA, isolated from frozen blood samples, was fragmented to a mean target size of 300 bp using a Covaris E210 instrument. The resulting fragmented DNA was end repaired using T4 and Klenow polymerases and T4 polynucleotide kinase with 10 mM dNTP followed by the addition of an A base at the ends using a Klenow exo fragment (3′→5′ exo–) and dATP (1 mM). Sequencing adaptors containing T overhangs were ligated to the DNA products followed by agarose (2%) gel electrophoresis. Fragments of about 400 bp were isolated from the gels (QIAGEN Gel Extraction Kit), and the adaptor-modified DNA fragments were PCR enriched for ten cycles using Phusion DNA polymerase (Finnzymes Oy) and PCR primers PE 1.0 and PE 2.0 (Illumina). Enriched libraries were further purified using agarose (2%) gel electrophoresis as described above. The quality and concentration of the libraries were assessed with the Agilent 2100 Bioanalyzer using the DNA 1000 LabChip (Agilent). Barcoded libraries were stored at −20 °C. All steps in the workflow were monitored using an in-house laboratory information management system with barcode tracking of all samples and reagents.

DNA sequencing. Template DNA fragments were hybridized to the surface of flow cells (Illumina PE flowcell, v4) and amplified to form clusters using the Illumina cBot. Briefly, DNA (8–10 pM) was denatured, followed by hybridization to grafted adaptors on the flowcell. Isothermal bridge amplification using Phusion polymerase was then followed by linearization of the bridged DNA, denaturation, blocking of the 3′ ends and hybridization of the sequencing primer. Sequencing-by-synthesis was performed on Illumina GAIIx instruments equipped with paired-end modules. Paired-end libraries were sequenced using 2 × 101 cycles of incorporation and imaging with Illumina sequencing kits, v4 or v5 (TruSeq). Each library or sample was initially run on a single lane for validation followed by further sequencing of ≥4 lanes with targeted raw cluster densities of 500–700 k/mm^2, depending on the version of the data imaging and analysis packages. Imaging and analysis of the data was performed using either the SCS2.6/RTA1.6 or SCS2.8/RTA1.8 software packages from Illumina, respectively. Real-time analysis involved the conversion of image data to base calling in real time.

Alignment. For each lane in the DNA sequencing output, the resulting qseq files were converted into fastq files using an in-house script. All output from the sequencing was converted, and the Illumina quality filtering flag was retained in the output. The fastq files were then aligned against Build 36 of the human reference sequence using BWA version 0.5.7 (ref. 24). All genomic locations quoted refer to HG18 Build 36.

BAM file generation. SAM file output from the alignment was converted into BAM format using SAMtools version 0.1.8 (ref. 25), and an in-house script was used to carry the Illumina quality filter flag over to the BAM file. The BAM files for each sample were then merged into a single BAM file using SAMtools. Finally, Picard version 1.17 (see URLs) was used to mark duplicates in the resulting sample BAM files.

SNP identification and genotype calling. A two-step approach was applied. The first step was to detect SNPs by identifying sequence positions where at least one individual could be determined to be different from the reference sequence with confidence (with a quality threshold of 20) based on the SNP calling feature of the pileup tool in SAMtools. SNPs that always differed heterozygous or homozygous from the reference were removed. The second step was to use the pileup tool to genotype the SNPs at the positions that were flagged as polymorphic. Because sequencing depth varies and, hence, the certainty of genotype calls also varies, genotype likelihoods rather than deterministic calls were calculated (**Supplementary Note**). Of the 2.5 million SNPs reported in the HapMap2 CEU samples, 96.3% were observed in the whole-genome sequencing data. Of the 6.9 million SNPs reported in the 1000 Genomes Project data, 89.4% were observed in the whole-genome sequencing data.

Methods for genotype imputation. Methods used for long-range phasing, genotype imputation, genealogy-based *in silico* genotyping and association testing are presented in the **Supplementary Note**.

Assessment of sun sensitivity. Sun sensitivity was self-assessed through questionnaires[14,15] using the Fitzpatrick score[26], where the lowest score (I) represents very fair skin that is very sensitive to ultraviolet radiation and the highest score (IV) represents dark skin that tans rather than burns in reaction to ultraviolet radiation exposure. Individuals scoring I and II were classified as being sensitive to sun and individuals scoring III and IV were classified as not being sensitive to sun.

Specification of the newly discovered SNP chr17:7,640,788. This SNP was identified by sequencing with the sequence context shown in **Supplementary Table 6**.

RNA analysis. RNA was isolated from blood using a QIAGEN RNA maxi kit according to the manufacturer's instructions. The concentration and quality of the RNA was determined with Agilent 2100 Bioanalyzers (Agilent Technologies). Complementary DNA (cDNA) was synthesized using the high-capacity cDNA reverse transcriptase kit (Applied Biosystems Inc.). Quantitative RT-PCR of *TP53* cDNA was performed with Applied Biosystems assay Hs99999147_m1 on an ABI 7900HT Real-time PCR system according to standard protocol. The RACE reaction was performed using a Smart-RACE cDNA amplification kit (Clontech) according to the manufacturer's protocol. Primer sequences are given in **Supplementary Table 6**. All sequencing was performed with BigDye R Terminator Chemistry on a 3730 system (Applied Biosystems Inc.).

24. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
25. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
26. Fitzpatrick, T.B. The validity and practicality of sun-reactive skin types I through VI. *Arch. Dermatol.* **124**, 869–871 (1988).

doi:10.1038/ng.926