Supplementary Material

Title

Association between a 46-SNP polygenic risk score and melanoma risk in Dutch familial melanoma patients

Supplementary Methods

Genotyping: Quality control procedures

Case series

Data analysis was performed using an in-house developed pipeline to generate the genotype data for all SNPs per sample. In brief, FastQ sequences were aligned to human reference genome GRCh37 using BWA. The obtained sequencing data had an average depth of >1,000 (>99% at least 30x) with horizontal coverage >99%. Variant calling was subsequently performed using the GATK HaplotypeCaller in ERC mode to produce a genomic VCF (gVCF) for each sample. Regions of high genome quality, with a minimum Genotype Quality (GQ) score of 90 (range 0–99) and >30 reads coverage, were selected. Using those regions, a single VCF file containing both variant and non-variant sites for all samples was generated through GATK's GenotypeGVCFs tool. Lastly, only the sites of interest (representing the SNPs included in this study, see supplementary Table S1) were selected using the tool 'SelectVariants', producing a table of genotypes for each sample. SNP data was complete for all individuals included in the case series.

Control series

Genome-wide genotype data was measured with the Illumina HumanOmniExpress-12 and -24 BeadChip available for 5363 samples; 5292 of these passed a call rate threshold of 95% and were

imputed using the 1000 Genomes phase1 v3[1] together with Genome of The Netherlands (GoNL) release 5[2] data as reference. Pre-imputation QC on the marker level consisted of a minor allele frequency (MAF) > 0.01, Hardy-Weinberg equilibrium (HWE) P-value > 10^{-4} and a single nucleotide polymorphism (SNP) yield > 95%, resulting in 609,046 SNPs to be used in the imputation process. Imputation was performed using the Impute2 pipeline developed by the GoNL team; see [http://www.bbmriwiki.nl/wiki/Impute2Pipeline].[3] This resulted in 20,011,335 SNPs. Post-imputation quality control consisted of exclusion of population outliers using principal component analysis, exclusion of sex discrepancies based on a comparison of genotype data and clinical data, and a relatedness check, which resulted finally in 4745 samples available for genome-wide analyses.

References

1. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. A map of human genome variation from population-scale sequencing. *Nature* 2010;**467**(7319):1061-73

2. Genome of the Netherlands Consortium. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat Genet* 2014;**46**(8):818-25

3. Kanterakis A, Deelen P, van Dijk F, Byelas H, Dijkstra M, Swertz MA. Molgenis-impute: imputation pipeline in a box. *BMC Res Notes* 2015;**8**:359

Supplementary Tables and Figures

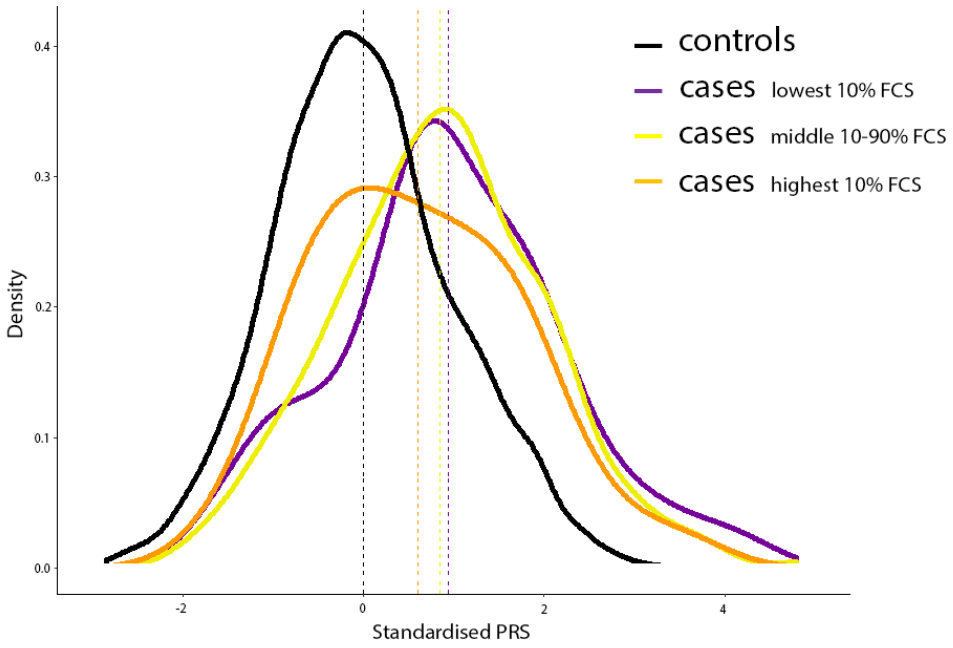
Supplementary Table S1: Genomic variants (SNPs) included in the 46-SNP PRS, and their associations with melanoma

SNP	Chromosome	Region/gene	Reference allele	Risk allele	Published OR (95% CI)[1]
rs7412746	1	*ARNT*	T	C	0.89 (0.85-0.92)
rs3219090	1	*PARP1*	C	T	0.88 (0.84-0.92)
rs1056837	2	*CYP1B1/RMDN2*	G	A	1.09 (1.03-1.14)
rs163092	2	*CYP1B1/RMDN2*	C	T	1.11 (1.06-1.16)
rs700635	2	*CASP8*	A	C	1.11 (1.07-1.15)
rs6554679	5	*TERT*	C	T	1.08 (1.04-1.12)
rs2736100	5	*TERT*	C	A	0.93 (0.87-0.98)
rs36115365	5	*TERT*	G	C	1.04 (0.99-1.08)
rs466502	5	*TERT*	A	G	1.16 (1.13-1.20)
rs2550948	5	*TERT*	T	C	1.05 (1.01-1.09)
rs16891982	5	*SLC45A2*	G	C	0.50 (0.24-0.76)
rs7776158	6	*CDKAL1*	G	A	1.11 (1.07-1.15)
rs12527588	6	*CDKAL1*	T	C	1.23 (1.16-1.31)
rs73069846	7	*AGR3*	C	T	1.11 (1.07-1.15)
rs34585474	7	*AGR3*	C	T	1.14 (1.09-1.20)
rs7781130	7	*AGR3*	T	C	1.22 (1.13-1.30)
rs6949072	7	*AGR3*	A	C	1.10 (1.04-1.16)
rs871024	9	*CDKN2A*	C	A	0.83 (0.77-0.88)
rs77560034	9	*CDKN2A*	G	C	1.18 (1.11-1.24)
rs3731217	9	*CDKN2A*	A	C	0.86 (0.81-0.92)
rs1011970	9	*CDKN2A*	G	T	1.14 (1.09-1.19)
rs4436178	9	*RAD23B*	G	A	1.18 (1.07-1.29)
rs113908778	9	*RAD23B*	C	T	1.28 (1.18-1.37)
rs1484375	9	*RAD23B*	G	A	1.12 (1.08-1.16)
rs2487999	10	*OBFC1*	C	T	1.14 (1.08-1.19)
rs76699054	11	*CCND1*	G	A	0.86 (0.78-0.94)
rs9651783	11	*CCND1*	T	G	1.12 (1.09-1.16)

rs1393350	11	*TYR*	G	A	1.22 (1.18-1.26)
rs1801516	11	*ATM*	G	A	0.84 (0.79-0.89)
rs4778138	15	*OCA2*	A	G	0.84 (0.78-0.90)
rs16953002	16	*FTO*	G	A	1.15 (1.10-1.20)
rs1805005	16	*MC1R*	G	T	1.00 (0.92-1.07)
rs1805006	16	*MC1R*	C	A	1.46 (1.19-1.79)
rs2228479	16	*MC1R*	G	A	1.12 (1.06-1.18)
rs11547464	16	*MC1R*	G	A	1.39 (1.20-1.61)
rs1805007	16	*MC1R*	C	T	1.85 (1.79-1.91)
rs1805008	16	*MC1R*	C	T	1.37 (1.31-1.43)
rs885479	16	*MC1R*	G	A	1.04 (0.95-1.12)
rs1110400	16	*MC1R*	T	C	1.30 (1.07-1.52)
rs1805009	16	*MC1R*	G	C	1.45 (1.34-1.57)
rs62211989	20	*ASIP*	G	C	1.43 (1.35-1.50)
rs6088372	20	*ASIP*	C	T	1.31 (1.24-1.37)
rs7274597	20	*ASIP*	C	T	0.87 (0.81-0.94)
rs6517661	21	*MX2*	A	C	0.91 (0.85-0.97)
rs45430	21	*MX2*	T	C	0.87 (0.83-0.90)
rs132985	22	*PLA2G6*	C	T	0.89 (0.85-0.94)

SNP = Single Nucleotide Polymorphism, OR = odds ratio

[1] Law MH, Bishop DT, Lee JE, et al: Genome-wide meta-analysis identifies five new susceptibility loci for cutaneous malignant melanoma. Nat Genet 47:987-995, 2015



Supplementary Figure S1. Distribution of the standardized PRS in cases divided into those with the lowest 10% familial clustering score (FCS) (purple), those with the middle 10-90% FCS (yellow) and those with the highest 10% FCS (orange). Dotted lines correspond to the means (see *Table 1*).