Genome-wide association study identifies a new melanoma susceptibility locus at 1q21.3

Supplementary table 1

Allele frequency information in Australian samples

SNP	Chr	Coordinate	OR	*P*	Coded allele	Other allele	Frequency of coded allele in cases	Frequency of coded allele in controls
rs7412746	1	149,127,095	0.82	2.5×10^{-7}	C	T	0.4012	0.4487
rs3219090	1	224,631,314	0.82	9.5×10^{-7}	T	C	0.289	0.3313
rs10170188	2	205,757,059	1.19	3.3×10^{-5}	A	G	0.2721	0.2388
rs17065828	3	62,017,865	0.83	3.9×10^{-5}	C	T	0.1873	0.2182
rs13177645	5	115,031,773	0.83	2.1×10^{-5}	G	A	0.4384	0.4836
rs7811987	7	136,176,803	1.19	1.1×10^{-5}	A	G	0.3807	0.3415
rs6478444	9	121,721,401	1.19	5.6×10^{-6}	G	A	0.3695	0.3295
rs10766295	11	16,061,966	1.17	2.1×10^{-5}	C	T	0.4619	0.4228
rs1584186	11	25,137,541	1.21	4.4×10^{-5}	A	G	0.2186	0.1884

Supplementary table 2

Allele frequency information in replication samples

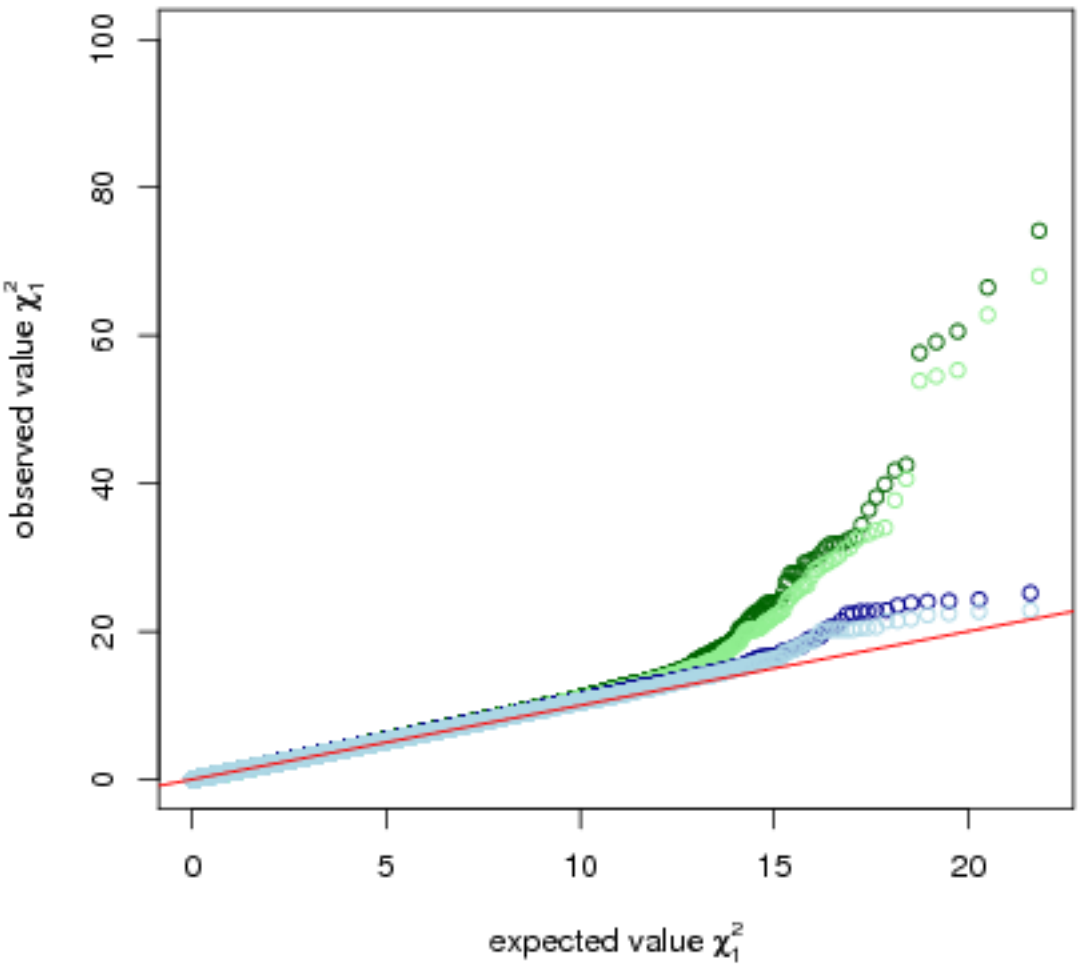
			Frequency of coded allele					
SNP	Coded allele	Other allele	Australian cases	Australian controls	USA 1 cases	USA 1 controls	USA 2 cases	USA 2 controls
rs7412746	C	T	0.4012	0.4487	0.4313	0.4744	0.4255	0.4625
rs3219090	T	C	0.289	0.3313	0.3127	0.3396	0.3075	0.3198

	Frequency of coded allele in Genomel site												
SNP	Israel cases	Israel controls	Italy cases	Italy controls	Paris cases	Paris controls	Poland cases	Poland controls	Scandinavia cases	Scandinavia controls	Spain cases	Spain controls	UK & Netherlands cases	UK & Netherlands controls
rs7412746	0.5357	0.6351	0.5491	0.5181	0.4596	0.4784	0.526	0.4674	0.4081	0.416	0.4787	0.4972	0.4048	0.4375
rs3219090	0.2589	0.3533	0.3161	0.4066	0.3049	0.3036	0.349	0.3315	0.3375	0.3976	0.2722	0.2391	0.3112	0.3218

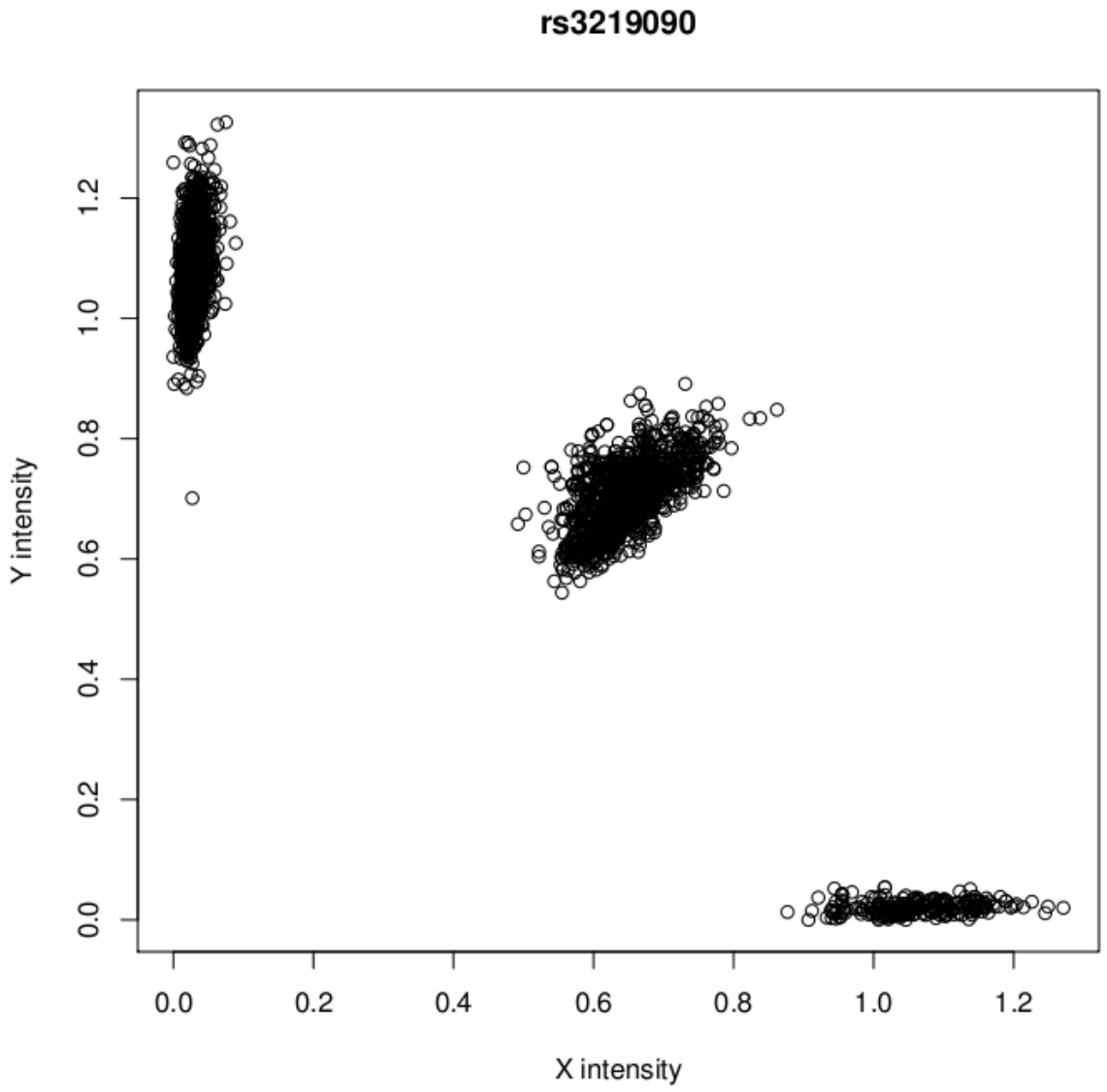
Supplementary Figures

Supplementary Figure 1 – Quantile-quantile (QQ) plot for SNPs directly genotyped on all Australian samples
Legend – Results for all SNPs uncorrected for the first 10 principle components are shown in dark green. Results for all SNPs corrected for the first 10 principle components are shown in light green.

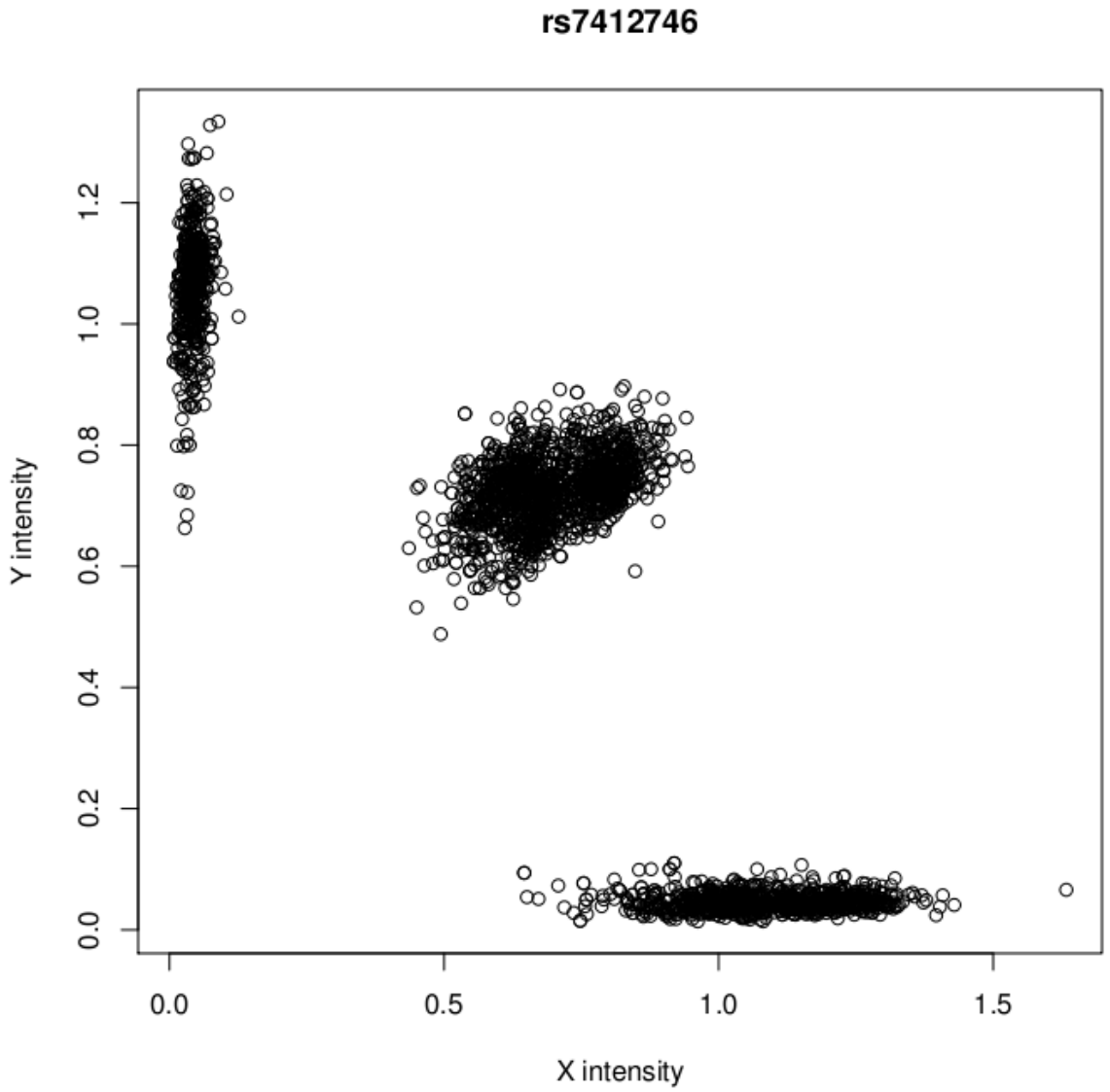
Results for all SNPs uncorrected for the first 10 principle components with previously identified regions removed (*MC1R* region: chr 16 87.5-89 megabases, *ASIP* region: chr 20 31-35 megabases, *OCA2*/*HERC2* region: chr 15 25-26.5 megabases, *CDKN2A* region: chr 9 21-22 megabases, *TYR* region: chr 11 88-89 megabases, *TYRP1* region: chr 9 12.2-13.2 megabases, *SLC45A2*: chr 5 33.5-34.5 megabases, all co-ordinates build 36) are shown in dark blue. Corrected results with previously identified regions removed are shown in light blue. The line y=x is shown in red.



Supplementary Figure 2
Supplementary Figure 2A
Cluster plot for lead SNPs in Australian discovery samples

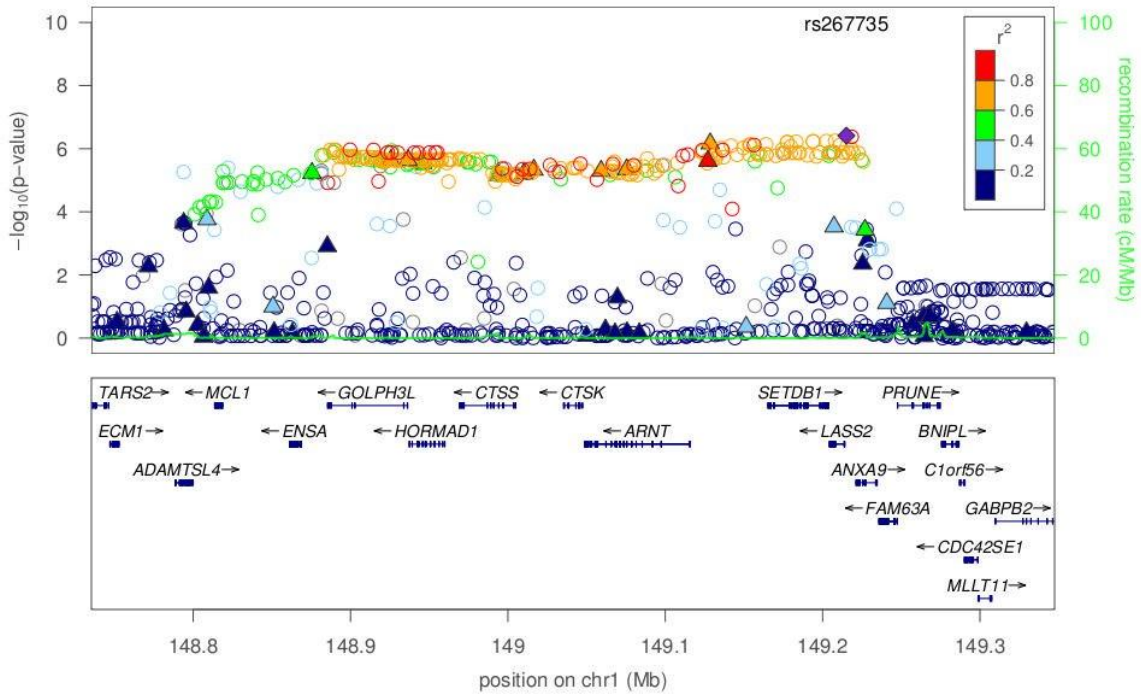


Supplementary Figure 2B

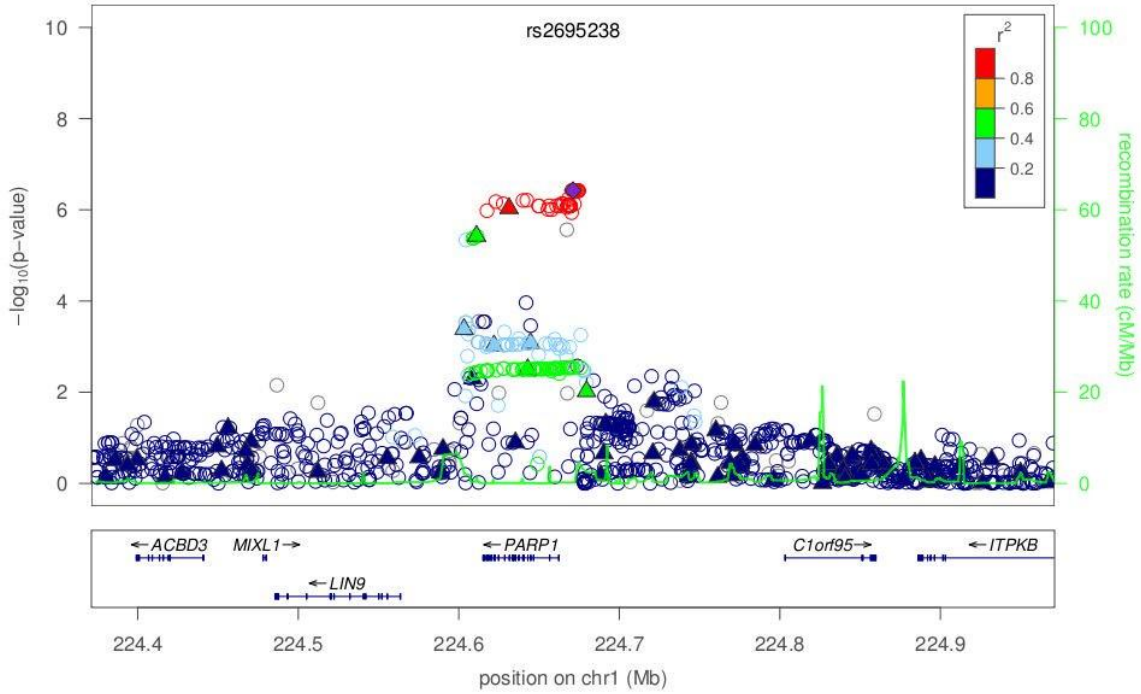


Supplementary Figure 3
Discovery sample association results at two novel melanoma susceptibility loci on chromosome 1 for both SNPs directly genotyped in all Australian samples and imputed SNPs. Results are corrected for the first 10 principal components. Genotyped SNPs are indicated by solid triangles and imputed SNPs by hollow circles. The top ranked SNP at each locus is shown as a solid purple diamond (this SNP is an imputed SNP at both loci). Imputation p-values for all SNPs are plotted. Note imputed and genotyped p-values for genotyped SNPs differ slightly because for the imputed result, analysis is based on dosage scores whereas with genotyped SNPs hard genotype calls are used. Association results shown are for (**A**) the chromosome 1 locus near 149 Mb, and (**B**) SNPs in the vicinity of the *PARP1* association signal. The color scheme indicates linkage disequilibrium between the most strongly-associated SNPs for the 149 Mb and *PARP1* region (shown in purple, rs267735 and rs2695238, respectively) and other genotyped SNPs in the two regions.

A



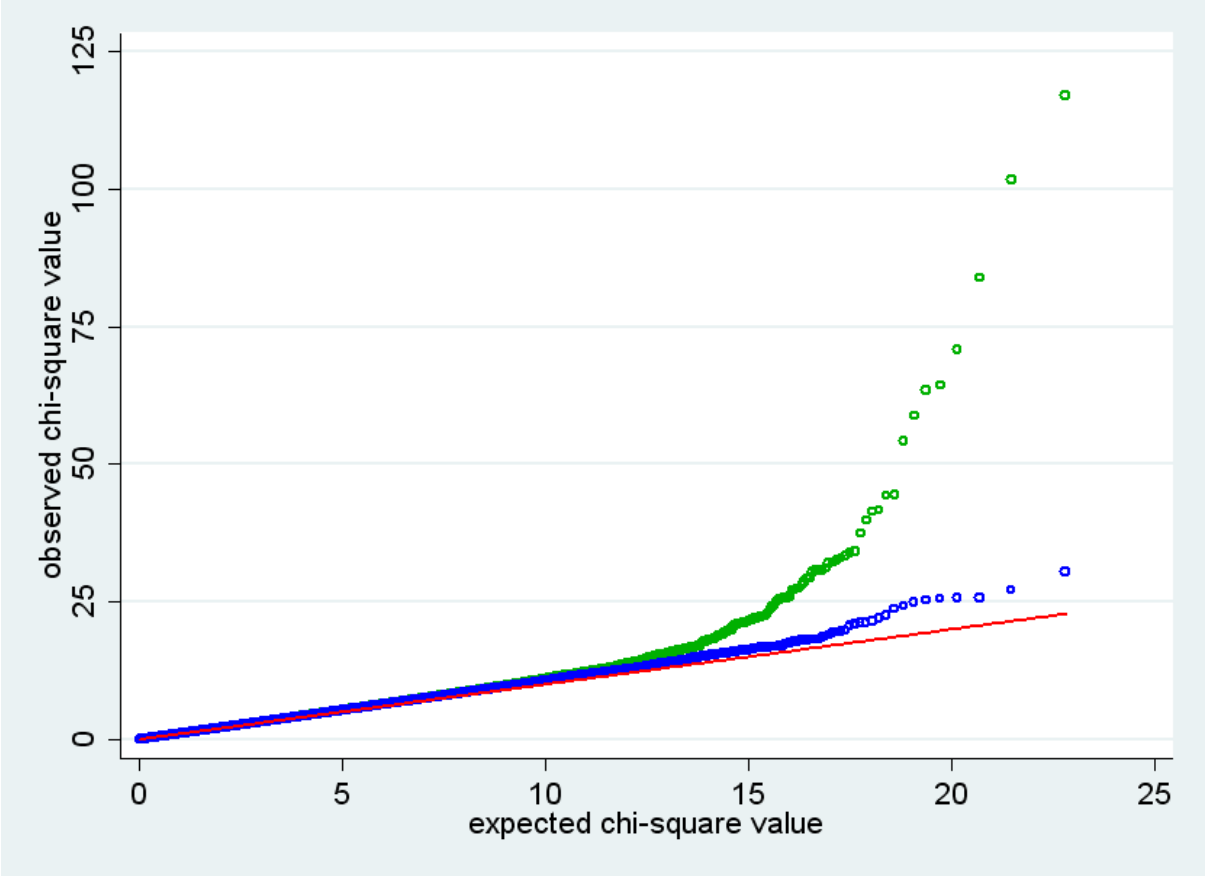
B


rs2695238
r²
0.8
0.6
0.4
0.2
−log10(p−value)
recombination rate (cM/Mb)
ACBD3
MIXL1
PARP1
C1orf95
ITPKB
LIN9
224.4
224.5
224.6
224.7
224.8
224.9
position on chr1 (Mb)

Supplementary Figure 4

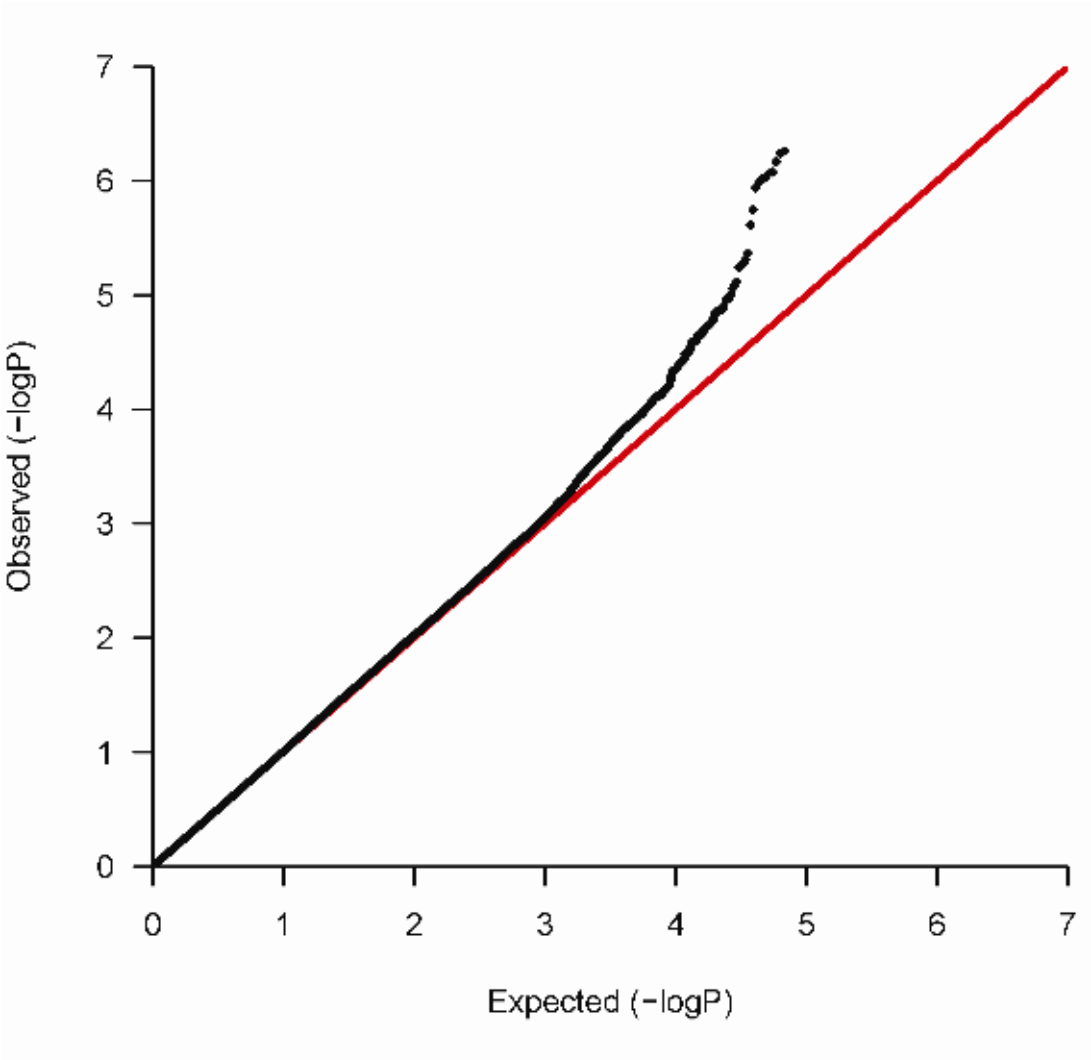
European replication cohort (Genomel) Q-Q plot

Genomic control lambda =1.05. The green line is using all SNPs with MAF>0.005 and the blue line is removing the 'known' regions, where known means chr 5 bp>0 & bp<2000000, chr 5 bp>33800000 & bp<34200000, chr 9 bp>21500000 & bp<22200000 chr 11 bp>87900000 & bp<89000000, chr 16 bp>87800000 & bp<89000000, chr 20 bp>31800000 & bp<33000000, chr 22bp>36500000 & bp<37000000. The line y=x is shown in red.



Supplementary Figure 5

United States 1 replication cohort Q-Q plot.
Genomic control lambda =1.01. The line y=x is shown in red.



ONLINE SUPPLEMENTARY NOTE

Study Subjects

Australian Discovery Sample.

2,265 melanoma cases were available for genotyping. These comprised a population based sample of 1,697 cases from Queensland, unselected for age-at-onset (Queensland study of Melanoma: Environment and Genetic Associations; Q-MEGA[1]) and 568 cases from a population-based case-control-family study of melanoma diagnosed before age 40 years, ascertained in Brisbane, Melbourne and Sydney (Australian Melanoma Family Study [2] ; AMFS). Approval for these studies was obtained from the Human Research Ethics Committees of QIMR, University of Sydney, University of Melbourne and cancer registries of NSW, Victoria and Queensland. Informed consent was obtained from all participants.

Age at onset (AAO) was ≤40 in 1,064 cases and >40 in 847 cases. 1,125 cases were histologically classified as invasive, while 307 were confirmed *in situ*. 1,080 cases had questionnaire-based nevus count based on a four point scale: "none", "a few", "moderate" and "very many". 1,146 cases had data on pigmentation variables, including: hair colour (fair, light brown, red, dark brown, black), eye colour (blue/grey, green hazel, brown/black), skin colour (light, medium, dark), and number of freckles (none, few, moderate, many).

Three sets of Australian controls sets were used. Firstly, a sample of 1,949 unrelated individuals was ascertained; they were mainly parents of adolescent twins (80% of the sample), together with a smaller number of twins and their siblings, recruited through schools to participate in the Brisbane Adolescent Twin Study[3,4]. Secondly, we also utilized as controls a sample of 2,318 genotyped endometriosis patients recruited by The Queensland Institute of Medical Research (QIMR) from 1995 to 2002[5]. Lastly, a set of 451 controls from AMFS were available for genotyping[2]. The AMFS and twin controls responded negatively when questioned about their personal history of melanoma, whilst the endometriosis samples were not screened. For all but very common diseases (prevalence >20%), using a large number of unscreened controls is more powerful than using a smaller number of screened controls[6]. In sum, 1,080 controls had questionnaire based nevus count based on a four point scale "none", "a few", "moderate" and "many", while 1,012 controls had data on pigmentation variables, as described above for cases. Approval for the twin and endometriosis studies was obtained from the QIMR Human Research Ethics Committee and the Australian Twin Registry. Approval for the AMFS controls was obtained from the Human Research Ethics Committees of University of Sydney, University of Melbourne and cancer registries of NSW, Victoria and Queensland. Informed consent was obtained from all participants.

Replication Sample 1 - GenoMEL

The GenoMEL data incorporated in the *in silico* replication come from a genome-wide association study of 2,804 cases and 1,835 controls for melanoma collected at 11 different centers across Europe plus Israel, supplemented by 3,878 controls from the WTCCC study (WTCCC [7]) and 1,905 French controls [8]. No Australian samples were included in the GenoMEL data used here.

Replication Sample 2 - MD Anderson Cancer Center

The study population was from two hospital-based case-control studies of cutaneous melanoma (CM) recruited from non-Hispanic white patients[9]. Controls were visitors at The University of Texas M. D. Anderson Cancer Center between March 1998 and August 2008. After performing data quality control of genome-wide SNP genotype data for these samples, we retained for analysis a total of 1,042 CM patients and 1,026 cancer-free controls frequency-matched on age and sex completed a self-administered comprehensive skin lifestyle questionnaire. This questionnaire was administered by an interviewer for 70% of cases and controls and was self-administered for 30% of cases and controls. An additional case series comprising 910 individuals presenting for treatment at M.D. Anderson Cancer Center were also studied. In total 1804 cases and 1026 controls were used. The study protocols were approved by the Institutional Review Board of the University of Texas M.D. Anderson Cancer Center. Informed consent was obtained from all participants.

Replication Sample 3 - Harvard

Melanoma nested case-control study within NHS and HPFS
We included the individuals in the melanoma nested case-control study within Nurses' Health Study (NHS) and Health Professionals Follow-up Study (HPFS). Recruitment of the study population was described previously[10]. Briefly, eligible cases in the NHS and the HPFS consisted of participants with pathologically confirmed melanoma, diagnosed any time after baseline up to the 2006 follow-up cycle (for both cohorts) who had no previously diagnosed cancer. One control per case was randomly selected from participants who were free of skin cancer up to and including the questionnaire cycle in which the case was diagnosed. Cases and their age-matched controls (+/- 1 year) were selected in the same cohort. We also included the controls in the case-control study of postmenopausal invasive breast cancer nested within the NHS cohort (NHS_BC)[11], as well as the participants from the type 2 diabetes case-control studies nested within the NHS and within the HPFS (NHS_T2D and HPFS_T2D)[12], and the coronary heart disease case-control studies nested within the NHS and within the HPFS (NHS_CHD and HPFS_CHD)[13]. Individuals diagnosed with skin cancer (melanoma, basal cell carcinoma or squamous cell carcinoma) were excluded and duplicated individuals were removed. All individuals were US non-Hispanics of European descent. Detailed descriptions for the five case-control studies were published previously[11-13]. Finally, we included 585 melanoma cases (mean age of diagnosis, 61.5 years) and 6,500 controls.

Description of Study Populations
Nurses' Health Study (NHS): The NHS comprises 121,700 female nurses aged 30-55. Nurses completed an initial self-administered questionnaire on their medical histories and baseline health-related exposures, forming the basis for the NHS cohort. From May 1989 through September 1990, we collected blood samples from 32,826 participants in the NHS cohort.

Health Professionals Follow-up Study (HPFS): In 1986, 51,529 men from all 50 U.S. states in health professions (dentists, pharmacists, optometrists, osteopath physicians, podiatrists, and veterinarians) aged 40-75 answered a detailed mailed questionnaire, forming the basis of the study. Between 1993 and 1994, 18,159 study participants provided blood samples by overnight courier.

Genotyping and data quality control

Australian Discovery Sample.

In main methods section.
Cluster plots confirming genotyping accuracy for rs7412746 and rs3219090 are given in supplementary figures 3A and 3B.

Replication Sample 1 - GenoMEL

The GenoMel data were genotyped in two phases, Phase 1 (Bishop *et al*., 2009[8]) on the Illumina HumanHap300 BeadChip version 2 duo array (317k SNPs) and Phase 2 on the Illumina Human610 quad array (610k SNPs). WTCCC study controls were genotyped on the Illumina HumanHap1.2 million array. French controls were genotyped by Centre National de Genotypage on the Illumina Humancnv370k array. Individuals were excluded for low call rate (<97% on the array on which the sample was genotyped), non-European ethnicity (determined by PCA in a similar manner to that described above), sex discrepancy with recorded phenotype information, first degree or closer relatedness with another sample or recommendation for exclusion by WTCCC. QC was applied to SNPs separately for each genotyping platform. SNPs were excluded for low call rate (<97%), Hardy-Weinberg equilibrium *P*-value < 10^{-20} or recommendation for exclusion by WTCCC (usually on the basis of poor clustering).
A Q-Q plot of the results across the genome is given in Supplementary Figure 4.

Replication Sample 2 - MD Anderson Cancer Center

Tissue samples were collected as whole blood, with various DNA extraction methods (including Gentra, Qiagen and phenol/chloroform). DNA samples for the first-stage genome-wide association study were genotyped using the Illumina Omnil-Quad array and were called using the BeadStudio algorithm, at the John Hopkins University Center for Inherited Disease Research (CIDR). Mean call rate for all samples was 99.86%. 41 failed genotyping with >10% missing rate across all SNPs, 11 samples had identity problems that could not be resolved. For this study, the IBD coefficients were estimated using 116,002 autosomal SNPs in PLINK[15]. In total, 126 duplicated, related (IBD), or outliers identified by PCA were excluded from the study. Following these exclusions there were 1,952 cases and 1,026 controls. Among 2,978 total case and control subjects passing quality control, 138 *in situ* cases were subsequently removed from the study as they had indeterminate phenotype. Ten atypical melanocytic proliferation (AMPs) patients were also excluded as not having invasive cancers. Finally, we analysed data from 1,804 cases and 1,026 controls available for the association study of melanoma susceptibility.
A Q-Q plot of the results across the genome is given in Supplementary Figure 5.

Replication Sample 3 - Harvard

We genotyped the two SNPs (rs7412746 and rs3219090) in the skin cancer case-control set using the OpenArray™ SNP Genotyping System (BioTrove, Woburn, MA). Laboratory personnel were blinded to the case-control status, and blinded quality control samples were inserted to validate genotyping procedures; concordance for the blinded samples was 100%. Primers, probes and conditions for genotyping assays are available upon request. For the

other five sets, we previously performed genotyping in BC_NHS using the Illumina HumanHap550 array [11], and performed genotyping using the Affymetrix 6.0 array in the NHS_T2D, HPFS_T2D, NHS_CHD and HPFS_CHD[12,13]. The SNP rs3219090 was directly genotyped in all of the five sets and rs7412746 was directly genotyped in NHS_BC. The rs7412746 was imputed in the other four sets with imputation R square larger than 0.95 in each set.

REFERENCES

1. Baxter, A.J. et al. The Queensland Study of Melanoma: environmental and genetic associations (Q-MEGA); study design, baseline characteristics, and repeatability of phenotype and sun exposure measures. *Twin Res Hum Genet* **11**, 183-96 (2008).
2. Cust, A.E. et al. Population-based, case-control-family design to investigate genetic and environmental influences on melanoma risk: Australian Melanoma Family Study. *Am J Epidemiol* **170**, 1541-54 (2009).
3. McGregor, B. et al. Genetic and environmental contributions to size, color, shape, and other characteristics of melanocytic naevi in a sample of adolescent twins. *Genet Epidemiol* **16**, 40-53 (1999).
4. Zhu, G. et al. A major quantitative-trait locus for mole density is linked to the familial melanoma gene CDKN2A: a maximum-likelihood combined linkage and association analysis in twins and their sibs. *Am J Hum Genet* **65**, 483-92 (1999).
5. Painter, J.N. et al. Genome-wide association study identifies a locus at 7p15.2 associated with endometriosis. *Nat Genet* (2010).
6. Moskvina, V., Holmans, P., Schmidt, K.M. & Craddock, N. Design of case-controls studies with unscreened controls. *Ann Hum Genet* **69**, 566-76 (2005).
7. WTCCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-78 (2007).
8. Bishop, D.T. et al. Genome-wide association study identifies three loci associated with melanoma risk. *Nat Genet* **41**, 920-5 (2009).
9. Li, C. et al. Genetic variants of the ADPRT, XRCC1 and APE1 genes and risk of cutaneous melanoma. *Carcinogenesis* **27**, 1894-901 (2006).
10. Zhang, M., Qureshi, A.A., Guo, Q. & Han, J. Genetic variation in DNA repair pathway genes and melanoma risk. *DNA Repair (Amst)*.
11. Hunter, D.J. et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* **39**, 870-4 (2007).
12. Qi, L. et al. Genetic variants at 2q24 are associated with susceptibility to type 2 diabetes. *Hum Mol Genet* **19**, 2706-15.
13. Rimm, E.B. et al. Prospective study of alcohol consumption and risk of coronary disease in men. *Lancet* **338**, 464-8 (1991).
14. Price, A.L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* **38**, 904-909 (2006).
15. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
16. Li, Y. & Abecasis, G. Mach 1.0: Rapid Haplotype Reconstruction and Missing Genotype Inference. *American Journal of Human Genetics* **S79, 2290**(2006).
17. Durbin, R.M. et al. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061-73 (2010).
18. Pruim, R.J. et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336-7 (2010).