nature genetics

Common sequence variants on 20q11.22 confer melanoma susceptibility

Kevin M Brown[1,23], Stuart MacGregor[2,23], Grant W Montgomery[2], David W Craig[3], Zhen Zhen Zhao[2], Kelly Iyadurai[1], Anjali K Henders[2], Nils Homer[4], Megan J Campbell[2], Mitchell Stark[2], Shane Thomas[2], Helen Schmid[5], Elizabeth A Holland[5], Elizabeth M Gillanders[6], David L Duffy[2], Judith A Maskiell[7], Jodie Jetann[8], Megan Ferguson[8], Dietrich A Stephan[3], Anne E Cust[7], David Whiteman[2], Adele Green[2], Håkan Olsson[9,21,22], Susana Puig[10,22], Paola Ghiorzo[11,22], Johan Hansson[12,22], Florence Demenais[13,22], Alisa M Goldstein[14], Nelleke A Gruis[15,22], David E Elder[16,22], Julia Newton Bishop[17,22], Richard F Kefford[5], Graham G Giles[18], Bruce K Armstrong[19], Joanne F Aitken[8], John L Hopper[7], Nicholas G Martin[2], Jeffrey M Trent[20], Graham J Mann[5] & Nicholas K Hayward[2]

We conducted a genome-wide association pooling study for cutaneous melanoma and performed validation in samples totaling 2,019 cases and 2,105 controls. Using pooling, we identified a new melanoma risk locus on chromosome 20 (rs910873 and rs1885120), with replication in two further samples (combined $P < 1 \times 10^{-15}$). The per allele odds ratio was 1.75 (1.53, 2.01), with evidence for stronger association in early-onset cases.

Cutaneous melanoma is an important health problem in fair-skinned populations worldwide, and its incidence is rising in most[1]. An understanding of the genetic factors influencing melanoma risk and the identification of susceptible individuals may aid in increasing sun protection and early detection of the disease in populations at risk. Four high-penetrance melanoma risk loci have been identified (*CDKN2A*, *ARF*, *CDK4* and a locus on 1p22)[2], and *MC1R* has been validated as a gene harboring low-penetrance risk alleles[3,4].

To identify additional low-penetrance risk alleles, we carried out a genome-wide association study (GWAS) involving the pooling of 864 cases drawn from a larger population-based sample of cases (individuals with melanoma) from Queensland, unselected for age at onset (Queensland study of Melanoma: Environment and Genetic Associations (Q-MEGA)[5]), and 864 controls (Q1). Each pool was hybridized to six Illumina HumanHap550 arrays, and SNPs were ranked after accounting for pooling error[6,7]. The proportion of SNPs with *P* values from pooling of <0.01 was consistent with what would be expected by chance if there were no true associations. Conversely, at smaller *P*-value thresholds, there were more SNPs than expected by chance. For example, at the 0.0001 threshold, we would expect to see ~55 SNPs under the null hypothesis of no association, but we in fact observed 90 SNPs, indicating that there were a number of true associations (**Supplementary Note** online).

Here we focus on only the most significant finding from pooling. The first-ranked (rs17305657, $P = 2.56 \times 10^{-7}$) and fourth-ranked (rs4911442, $P = 2.39 \times 10^{-6}$) SNPs are 1.5 Mb apart on chromosome 20. Multiple other SNPs in this region showed evidence for association (**Supplementary Fig. 1** online). When the pooling results were validated by individual genotyping, concordance was excellent (rs17305657, $P = 3.63 \times 10^{-6}$; rs4911442, $P = 1.03 \times 10^{-8}$). To fine map this region, we selected 31 additional SNPs that span ~2.78 Mb in chromosome bands 20q11.21–q11.22 (**Supplementary Methods** online). Selection was based on candidate genes, pooling results, ethnic allele frequency differences and linkage disequilibrium (LD) patterns (that is, SNPs whose presence correlated with that of both rs17305657 and rs4911442; **Supplementary Methods**). We then genotyped the set of 33 SNPs in three sample sets: 789 cases and 854 controls from the pooled Q-MEGA subjects (Q1); a second set of 725 cases and 797 controls sampled independently from Q-MEGA (Q2); and 505 cases and 454 controls (A1) from an independent population-based study of melanoma diagnosed before age 40 years, ascertained

[1]Integrated Cancer Genomics Division, The Translational Genomics Research Institute, Phoenix, Arizona 85028, USA. [2]The Population Studies and Genetics Division, Queensland Institute of Medical Research, Brisbane, Queensland 4029, Australia. [3]Neurogenomics Division, The Translational Genomics Research Institute, Phoenix, Arizona 85028, USA. [4]University of California Los Angeles, Los Angeles, California 90095, USA. [5]Westmead Institute of Cancer Research and Sydney Melanoma Unit, University of Sydney at Westmead Millennium Institute, Westmead, New South Wales 2145, Australia. [6]Inherited Disease Research Branch, National Human Genome Research Institute, National Institutes of Health, Baltimore, Maryland 21224, USA. [7]Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, School of Population Health, University of Melbourne, Melbourne, Victoria 3052, Australia. [8]Viertel Centre for Research in Cancer Control, The Cancer Council Queensland, Brisbane, Queensland 4004, Australia. [9]Lund Cancer Center Department of Oncology, University Hospital, SE-22185 Lund, Sweden. [10]Dermatology Department, Melanoma Unit, Hospital Clínic, Institut de Investigació Biomèdica August Pi Suñe, Universitat de Barcelona, 08036 Barcelona, Spain. [11]Department of Oncology, Biology, and Genetics, University of Genova, 16132 Genova, Italy. [12]Department of Oncology-Pathology, Karolinska Institute and Karolinska University Hospital, Solna, S-171 76 Stockholm, Sweden. [13]INSERM, U794, Fondation Jean-Dausset–CEPH, 75010 Paris, France. [14]Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland 20892, USA. [15]Department of Dermatology, Leiden University Medical Center, 2333 AL Leiden, The Netherlands. [16]Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [17]Section of Epidemiology and Biostatistics, Leeds Institute of Molecular Medicine, University of Leeds, Leeds LS9 7TF, UK. [18]Cancer Epidemiology Centre, The Cancer Council Victoria, Carlton, Victoria 3053, Australia. [19]School of Public Health, University of Sydney, Sydney, New South Wales 2006, Australia. [20]Genetic Basis of Human Disease Division, The Translational Genomics Research Institute, Phoenix, Arizona 85028, USA. [21]On behalf of the Lund Melanoma Group. [22]On behalf of the International Melanoma Genetics Consortium (GenoMEL). [23]These authors contributed equally to this work. Correspondence should be addressed to S.M. (stuart.macgregor@qimr.edu.au).

Received 6 March; accepted 30 April; published online 18 May 2008; doi:10.1038/ng.163

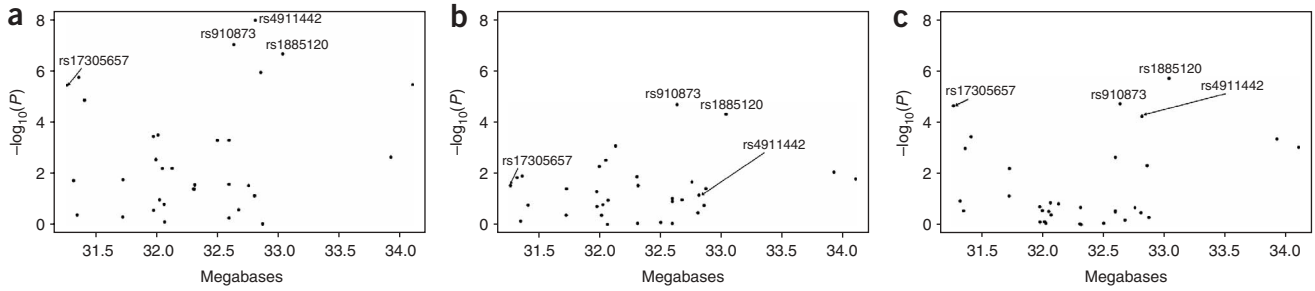

Figure 1 Association analysis of SNPs across a region of chromosome 20q11.22. *P* values for association testing for the three samples. (**a**) Q1 sample. (**b**) Q2 sample. (**c**) A1 sample.

in Brisbane, Melbourne and Sydney (Australian Melanoma Family Study (AMFS); A. Cust, unpublished data). The combined sample comprised 2,019 cases and 2,105 controls, all of European descent. All cases had incident primary melanomas (stage 1). **Figure 1** shows the association results for Q1, Q2 and A1 (see also **Supplementary Tables 1–5** online). Several SNPs showed stronger evidence for association than the two SNPs identified using the HumanHap550 array. In each of the samples, two SNPs were highly associated with melanoma, rs910873 and rs1885120 (**Fig. 1**), with combined $P < 1 \times 10^{-15}$ (**Fig. 2**).

To determine whether a single SNP could account for the association signal in this region, we analyzed multiple SNPs jointly by logistic regression. The effects of rs910873 and rs1885120 could not be separated (fitting one in the model rendered the other redundant, $r^2 > 0.9$ both in cases and controls). Given the high r^2 value between rs910873 and rs1885120, we cannot unambiguously identify the interval in which the causal variant(s) lie. When either SNP was fitted in the logistic regression, all other SNPs typed in the vicinity were redundant (**Supplementary Table 6** online). There was some evidence for residual association around rs17305657, indicating there may be two independent signals in this region of chromosome 20 (either in the same or in different genes). Alternatively, both SNPs may be in incomplete LD with a single causal variant.

We tested whether the observed association at rs910873 could be explained by any of the obvious candidate genes in the region. SNPs rs2071054 and rs3213182 lie in or near *E2F1*, which encodes a transcription factor regulated by the retinoblastoma protein, but neither showed evidence for association once the effect of rs910873 (or rs1885120) was accounted for (**Supplementary Table 6**). Similarly, rs819163, rs6059743 and rs819162, within or adjacent to *ASIP*, showed little evidence for residual association (smallest $P = 0.062$; **Supplementary Table 6**). *ASIP* encodes the human ortholog of the murine agouti gene product, a paracrine signaling molecule and antagonist of α-melanocyte–stimulating hormone (the ligand of the *MC1R* gene product), which regulates synthesis of melanin. In humans, a SNP in the 3′ untranslated region of *ASIP* has been associated with variation in pigmentation[8–10] but has not been independently associated with nevi or melanoma risk[4,8,10]. Although we have not genotyped all variants in *ASIP* and *E2F1*, our findings provide evidence that a risk variant for melanoma lies in a more telomeric region that includes several other candidate loci.

It is noteworthy that rs910873 (and, through LD, rs1885120) is polymorphic only in Caucasian (CEU) and not in Asian (JPT or CHB) or African (YRI) HapMap samples. Furthermore, in our dataset, the frequency of the risk alleles for rs910873 and rs1885120 was significantly higher in melanoma cases (frequency 0.15) than controls (frequency 0.09). Because population stratification may cause false-positive associations in non-homogeneous samples, we assessed the self-reported ancestry for all four grandparents in a subset of 1,779 controls and 597 cases from Q1 and Q2 for which data were available. The vast majority of controls ($n = 1{,}438$) and cases ($n = 585$) reported 100% Northern European ancestry (others had one or more grandparents with Southern European ancestry). In the Northern European subset, allele frequencies were very similar (within 0.01) to those found in the whole sample, implying that population stratification (even within Europe) could not explain the results.

To further describe the phenotypic associations of the peak SNPs, we examined age at onset using rs910873. In Q1 and Q2, we used two criteria, ≤40 years and ≤30 years. Because all A1 cases had age at onset ≤40 years, we stratified this sample only at ≤30. Results for Q2 indicated that the age-at-onset threshold of ≤40 was important

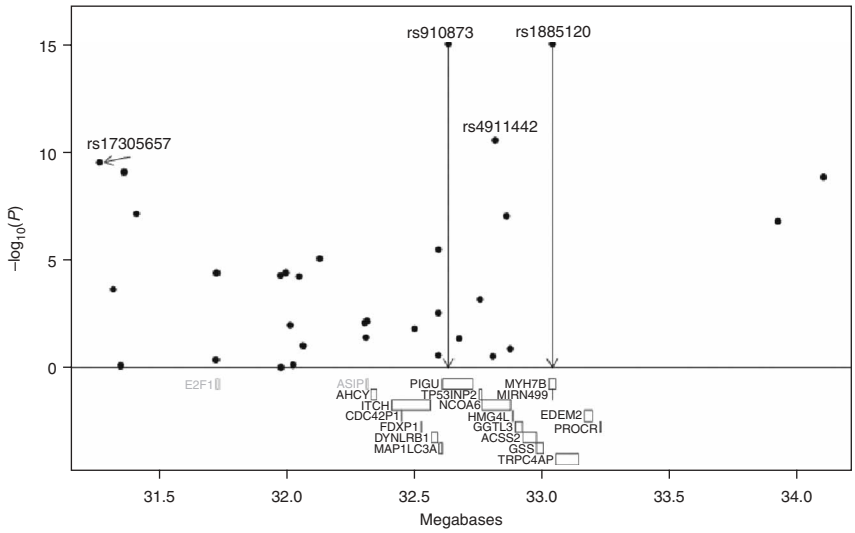

Figure 2 Association analysis of SNPs across a region of chromosome 20q11.22 for the combined sample. Genes from *AHCY* to *PROCP* are shown in black; two candidate genes, *E2F1* and *ASIP*, are shown in gray.

($\leq$40 subset per allele odds ratio (OR) = 1.83 (1.39, 2.41); $>$40 subset OR = 1.30 (1.00, 1.69)) but that the $\leq$30 threshold was not ($\leq$30 subset OR = 1.82 (1.31, 2.53)). Examining only the Q2 case sample, there was a significant effect of the rs910873 risk allele when age at onset was analyzed as a quantitative trait ($P = 0.03$, **Supplementary Note**). In A1, the association in the $\leq$30 subset (OR = 1.77 (1.22, 2.57)) was no different from that in the $>$30 subset (OR = 1.87 (1.40, 2.51)). Age at onset was not significant in Q1 (data not shown), the discovery sample that was ascertained without respect to age at onset. Overall, rs910873 (and, through LD, rs1885120) seems to be related to early onset.

To provide an unbiased estimate of the OR and population-attributable fraction (PAF), we used the two replication samples Q2 and A1. The ORs were 1.72 (1.48, 2.14) and 1.81 (1.52, 2.17) for all cases and $\leq$40 age at onset groups, respectively. The PAFs in these samples from the Australian population were 0.11 (0.08, 0.16) and 0.12 (0.08, 0.17) for all cases and $\leq$40 age at onset groups, respectively.

The most strongly associated region between rs910873 and rs1885120 is $\sim$400 kb in length. rs1885120 maps within an intron of *MYH7B* (myosin, heavy polypeptide 7B, cardiac muscle, beta), which is not expressed in a large panel of melanoma cell lines analyzed[11]. rs910873 maps within an intron of *PIGU* (also known as *CDC91L1*), which encodes phosphatidylinositol glycan anchor biosynthesis class U and is expressed in all melanoma cell lines assessed, as are *TP53INP2* (encoding tumor protein p53–inducible nuclear protein-2), *NCOA6* (encoding nuclear receptor coactivator-6), *GGTL3* (encoding gamma-glutamyltransferase-like-3), *ACSS2* (encoding acyl-CoA synthetase short-chain family member 2) and *GSS* (encoding glutathione synthetase), the other genes that map between these two SNPs.

In summary, we have identified a new melanoma risk locus and replicated this association in two independent samples, with a combined $P < 1 \times 10^{-15}$. The effect size for melanoma associated with this genomic region is of similar magnitude to that associated with *MC1R* (OR $\sim$2, PAF $\sim$0.1 for heterozygotes)[3,4], the only low-penetrance melanoma susceptibility gene identified to date whose association has been robustly replicated. Identification of the causal variants associated with melanoma predisposition at 20q11.22 will help refine the estimates of risk for this increasingly common cancer.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

This work was supported by the National Cancer Institute (NCI) of the US National Institutes of Health (NIH) (CA88363, CA83115), the National Health and Medical Research Council of Australia (NHMRC) (380385, 389892, 496675, 402761), the Cancer Councils New South Wales, Victoria and Queensland, the Cancer Institute New South Wales, the Melanoma Research Foundation (MRF) and a charitable contribution from F. Najafi. N.K.H. and G.W.M. are supported by the NHMRC Fellowships scheme and J.L.H. is an Australia Fellow of the NHMRC. S.M. and K.M.B. are recipients of Career Development Awards from the NHMRC (496674) and MRF, respectively. K.M.B. and J.M.T. are supported by the NCI/NIH (respectively, CA109544 and CA083115; and CA109544). D.A.S. is supported by the US National Heart, Lung, and Blood Institute of the NIH (HL086528). B.K.A. is supported by the University of Sydney Medical Foundation. A.E.C. is supported by an NHMRC postdoctoral fellowship. The authors are grateful to M. Huentelman and S. Szelinger for technical assistance. The AMFS and Q-MEGA gratefully acknowledge all of their participants, the hard work of all its research interviewers and examiners, and C. Agha-Hamilton for managing the AMFS biospecimens. Q-MEGA thanks A. Baxter, M. de Nooyer, I. Gardner, D. Statham, B. Haddon, J. Palmer, B. Castellano, L. Bardsley, D. Smyth and H. Beeby for their input into project management, sample processing and database development.

Published online at http://www.nature.com/naturegenetics/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Lens, M.B. & Dawes, M. *Br. J. Dermatol.* **150**, 179–185 (2004).
2. de Snoo, F.A. & Hayward, N.K. *Cancer Lett.* **230**, 153–186 (2005).
3. Palmer, J.S. *et al. Am. J. Hum. Genet.* **66**, 176–186 (2000).
4. Landi, M.T. *et al. J. Natl. Cancer Inst.* **97**, 998–1007 (2005).
5. Baxter, A. *et al. Twin Res. Hum. Genet.* **11**, 183–196 (2008).
6. Macgregor, S. *et al. Nucleic Acids Res.* **36**, e35 (2008).
7. Macgregor, S. *et al. Nucleic Acids Res.* **34**, e55 (2006).
8. Kanetsky, P.A. *et al. Am. J. Hum. Genet.* **70**, 770–775 (2002).
9. Bonilla, C. *et al. Hum. Genet.* **116**, 402–406 (2005).
10. Meziani, R. *et al. J. Dermatol. Sci.* **40**, 133–136 (2005).
11. Johansson, P., Pavey, S. & Hayward, N. *Pigment Cell Res.* **20**, 216–221 (2007).