A germline variant in the *TP53* polyadenylation signal confers cancer susceptibility

Simon N Stacey[1,54], **Patrick Sulem**[1,54], **Aslaug Jonasdottir**[1], **Gisli Masson**[1], **Julius Gudmundsson**[1], **Daniel F Gudbjartsson**[1], **Olafur T Magnusson**[1], **Sigurjon A Gudjonsson**[1], **Bardur Sigurgeirsson**[2,3], **Kristin Thorisdottir**[2,3], **Rafn Ragnarsson**[2,3], **Kristrun R Benediktsdottir**[2,3], **Bjørn A Nexø**[4], **Anne Tjønneland**[5], **Kim Overvad**[6], **Peter Rudnai**[7], **Eugene Gurzau**[8], **Kvetoslava Koppova**[9], **Kari Hemminki**[10], **Cristina Corredera**[11], **Victoria Fuentelsaz**[12], **Pilar Grasa**[11], **Sebastian Navarrete**[13], **Fernando Fuertes**[13], **Maria D García-Prats**[14], **Enrique Sanambrosio**[15], **Angeles Panadero**[16], **Ana De Juan**[17], **Almudena Garcia**[17], **Fernando Rivera**[17], **Dolores Planelles**[18], **Virtudes Soriano**[19], **Celia Requena**[20], **Katja K Aben**[21,22], **Michelle M van Rossum**[23], **Ruben G H M Cremers**[21,24], **Inge M van Oort**[24], **Dick-Johan van Spronsen**[25], **Jack A Schalken**[24], **Wilbert H M Peters**[26], **Brian T Helfand**[27], **Jenny L Donovan**[28], **Freddie C Hamdy**[29], **Daniel Badescu**[30], **Ovidiu Codreanu**[30], **Mariana Jinga**[30], **Irma E Csiki**[31], **Vali Constantinescu**[31], **Paula Badea**[31], **Ioan N Mates**[32], **Daniela E Dinu**[32], **Adrian Constantin**[32], **Dana Mates**[31], **Sjofn Kristjansdottir**[33], **Bjarni A Agnarsson**[2,3], **Eirikur Jonsson**[2,3], **Rosa B Barkardottir**[2,3], **Gudmundur V Einarsson**[2], **Fridbjorn Sigurdsson**[2], **Pall H Moller**[2], **Tryggvi Stefansson**[2], **Trausti Valdimarsson**[34], **Oskar T Johannsson**[2], **Helgi Sigurdsson**[2], **Thorvaldur Jonsson**[2], **Jon G Jonasson**[2,3,35], **Laufey Tryggvadottir**[35], **Terri Rice**[36], **Helen M Hansen**[36], **Yuanyuan Xiao**[37], **Daniel H Lachance**[38], **Brian Patrick O'Neill**[38], **Matthew L Kosel**[39], **Paul A Decker**[39], **Gudmar Thorleifsson**[1], **Hrefna Johannsdottir**[1], **Hafdis T Helgadottir**[1], **Asgeir Sigurdsson**[1], **Valgerdur Steinthorsdottir**[1], **Annika Lindblom**[40], **Swedish Low-risk Colorectal Cancer Study Group**[41], **Robert S Sandler**[42], **Temitope O Keku**[42], **Karina Banasik**[43], **Torben Jørgensen**[44,45], **Daniel R Witte**[46], **Torben Hansen**[43,47,48], **Oluf Pedersen**[43,45,48,49], **Viorel Jinga**[30], **David E Neal**[50], **William J Catalona**[27], **Margaret Wrensch**[36,37], **John Wiencke**[36,37], **Robert B Jenkins**[51], **Eduardo Nagore**[20], **Ulla Vogel**[52], **Lambertus A Kiemeney**[21,24], **Rajiv Kumar**[10], **José I Mayordomo**[53], **Jon H Olafsson**[2,3], **Augustine Kong**[1], **Unnur Thorsteinsdottir**[1,3], **Thorunn Rafnar**[1], and **Kari Stefansson**[1,3]

[1]deCODE genetics, Reykjavik, Iceland. [2]Landspitali-University Hospital, Reykjavik, Iceland. [3]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [4]Institute of Human Genetics, University of Aarhus, Aarhus, Denmark. [5]Institute of Cancer Epidemiology, Danish Cancer Society, Copenhagen, Denmark. [6]Department of Epidemiology, School of Public Health, University of Aarhus, Aarhus, Denmark. [7]National Institute of Environmental Health, Budapest, Hungary. [8]Environmental Health Centre, Cluj, Cluj-Napoca, Romania. [9]State Health Institute, Banska Bystrica, Slovakia. [10]Division of Molecular Genetic Epidemiology, German Cancer Research Centre, Heidelberg, Germany. [11]Division of Dermatology, University of Zaragoza, Zaragoza, Spain. [12]Division of Dermatology, Gregorio Marañon University Hospital, Madrid, Spain. [13]Division of Radiation Oncology, University of Zaragoza, Zaragoza, Spain. [14]Division of Pathology, San Jorge General Hospital, Huesca, Spain. [15]Division of Surgery, General Hospital, Soria, Spain. [16]Division of Medical Oncology, Ciudad de Coria Hospital, Coria, Spain. [17]Division of Medical Oncology, Marques de Valdecilla University Hospital, Santander, Spain. [18]Laboratory of Histocompatibility-Molecular Biology, Centro de Transfusión de la Comunidad Valenciana, Valencia, Spain. [19]Department of Oncology, Instituto Valenciano de Oncologia, Valencia, Spain. [20]Department of Dermatology, Instituto Valenciano de Oncologia, Valencia, Spain. [21]Department of Epidemiology, Biostatistics and Health Technology Assessment, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands. [22]Comprehensive Cancer Center The Netherlands, Location Nijmegen, Nijmegen, The Netherlands. [23]Department of Dermatology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands. [24]Department of Urology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands. [25]Department of Internal Medicine, Canisius Wilhelmina Ziekenhuis, Nijmegen, The Netherlands. [26]Department of Gastroenterology, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands. [27]Department of Urology, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA. [28]School of Social and Community Medicine, University of Bristol, Bristol, UK. [29]Nuffield Department of Surgical Science, University of Oxford, John Radcliffe Hospital, Oxford, UK. [30]University of Medicine and Pharmacy Carol Davila, Bucharest, Romania. [31]National Institute of Public Health, Bucharest, Romania. [32]University of Medicine and Pharmacy Carol Davila, St. Mary Surgical Clinic, Bucharest, Romania. [33]Speglun ehf, Reykjavik, Iceland. [34]Domus Medica, Reykjavik, Iceland. [35]Icelandic Cancer Registry, Reykjavik, Iceland. [36]Department of Neurological Surgery, University of California, San Francisco, San Francisco, California, USA. [37]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, California, USA. [38]Department of Neurology, Mayo Clinic, Rochester, Minnesota, USA. [39]Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA. [40]Department of Molecular Medicine and Surgery, Karolinska Institute, Stockholm, Sweden. [42]University Center for Gastrointestinal Biology and Disease, University of North Carolina, Chapel Hill, North Carolina, USA. [43]The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, University of Copenhagen, Copenhagen, Denmark. [44]Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark. [45]Research Centre for Prevention and Health, Glostrup University Hospital, Glostrup, Denmark. [46]Steno Diabetes Center, Gentofte, Denmark. [47]Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark. [48]Hagedorn Research Institute, Gentofte, Denmark. [49]Faculty of Health Sciences, University of Aarhus, Aarhus, Denmark. [50]Oncology Centre, Addenbrooke's Hospital, University of Cambridge, Cambridge, UK. [51]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota, USA. [52]National Research Centre for the Working Environment, Copenhagen, Denmark. [53]Division of Medical Oncology University Hospital, Zaragoza, Spain.

Abstract

To identify new risk variants for cutaneous basal cell carcinoma, we performed a genome-wide association study of 16 million SNPs identified through whole-genome sequencing of 457 Icelanders. We imputed genotypes for 41,675 Illumina SNP chip-typed Icelanders and their

relatives. In the discovery phase, the strongest signal came from rs78378222[C] (odds ratio (OR) = 2.36, $P = 5.2 \times 10^{-17}$), which has a frequency of 0.0192 in the Icelandic population. We then confirmed this association in non-Icelandic samples (OR = 1.75, $P = 0.0060$; overall OR = 2.16, $P = 2.2 \times 10^{-20}$). rs78378222 is in the 3′ untranslated region of *TP53* and changes the AATAAA polyadenylation signal to AATACA, resulting in impaired 3′-end processing of *TP53* mRNA. Investigation of other tumor types identified associations of this SNP with prostate cancer (OR = 1.44, $P = 2.4 \times 10^{-6}$), glioma (OR = 2.35, $P = 1.0 \times 10^{-5}$) and colorectal adenoma (OR = 1.39, $P = 1.6 \times 10^{-4}$). However, we observed no effect for breast cancer, a common Li-Fraumeni syndrome tumor (OR = 1.06, $P = 0.57$, 95% confidence interval 0.88–1.27).

Basal cell carcinoma (BCC) is the most common cancer in people of European ancestry. Sun exposure is the primary risk factor for BCC, but genetic predisposition also plays a substantial role[1,2]. High penetrance mutations in Hedgehog pathway genes (*PTCH1*, *PTCH2* and *SUFU*) cause Gorlin syndrome, also known as basal cell nevus syndrome[1]. Candidate gene and genome-wide association studies have shown that common, low-penetrance susceptibility alleles exist at *MC1R* and several other loci[3–7].

Previously, we described a large genome-wide association study of the Icelandic population using common SNPs and showed how genotypes can be phased over long distances[8,9]. For BCC, we initially generated Illumina SNP chip data for 1,366 affected individuals (cases) and 40,309 controls. Haplotype association analysis based on long-range phasing showed that several 0.3-cM haplotypes at 17p13 were strongly associated with BCC. The most significant signals were produced by haplotype A6 (OR = 2.04, $P = 2.0 \times 10^{-10}$), spanning the region chr17: 7,186,095–7,425,536 and by a highly correlated haplotype, A8 (OR = 2.00, $P = 3.0 \times 10^{-10}$), spanning an adjacent region, chr17:7,431,901–7,680,389. The region covered by these haplotypes is illustrated in Figure 1.

To search for variants that might not be covered well by the chips, we used high-capacity DNA sequencing techniques to sequence the entire genomes of 457 Icelanders to an average depth of over 10× (Online Methods), which identified approximately 16 million SNPs. To ensure that all the rare risk alleles that might be carried on the A6 or A8 backgrounds would be sequenced, we included ten individuals who carried these haplotypes among the 457 individuals selected for sequencing. Using imputation assisted by long-range haplotype phasing, we used the sequence data to determine the genotypes of the 16 million SNPs in the 41,675 Icelanders who had been genotyped on the SNP chips. Moreover, knowledge of Icelandic genealogy allowed us to propagate genotypic information into individuals for whom we have neither SNP chip nor sequence data, a process we refer to as 'genealogy-based *in silico* genotyping'. We refer to the combined method of imputing sequence-derived data into phased chromosomes from chip-typed individuals and using genealogy-based *in silico* genotyping to infer the sequence of ungenotyped individuals as 'two-way imputation' (Supplementary Note).

We conducted a two-way-imputation–based genome-wide BCC association analysis of the 16 million SNPs, which we designated the 'discovery phase'. This analysis identified a number of SNPs with strong associations in the region covered by the two haplotypes. The strongest signal (OR = 2.36, $P = 5.2 \times 10^{-17}$) came from rs78378222, located in the 3′ untranslated region of *TP53* (Fig. 1 and Table 1). This signal was not only the strongest in the region covered by the two candidate haplotypes, but it was also the strongest signal of the 16 million SNPs observed genome wide. The minor (C, at risk) allele is present in the Icelandic population at a frequency of 0.0192. There was no deviation from Hardy-Weinberg equilibrium, and we observed C/C homozygotes (although they are rare), so the variant is not recessive lethal. The SNP was not on the Illumina chips and is not in HapMap2 or HapMap3. The best on-chip tag SNP was rs4796305, which produced only a

modest signal (OR = 1.18, P = 0.0024, minor allele frequency = 0.119, r^2 = 0.15). A quantile-quantile plot is shown in Supplementary Figure 1. Aside from the 17p13 locus, all other SNPs that surpassed the $P = 3.0 \times 10^{-9}$ Bonferroni-adjusted threshold for genome-wide significance were in previously published loci[5,6].

To further investigate the rs78378222 association, we conducted a follow-up phase. The design of the discovery and follow-up phases is detailed in Supplementary Figure 2. We devised a Centaurus[10] single-track assay for rs78378222 and used it to genotype all available samples from Icelandic cases with BCC (n = 2,322) and 7,200 controls. Among these, 1,044 out of the 2,322 single-track genotyped cases had not previously been genotyped by the Illumina chip. An association analysis with these 1,044 cases and 7,200 controls yielded OR = 2.19 and $P = 3.2 \times 10^{-7}$ (Table 1, see the Iceland follow-up single-track genotyped category). Note that this is not a fully independent replication of the discovery phase result, as some of the individuals who had been assigned genealogy-based *in silico* genotypes in the discovery phase were single-track genotyped in the follow-up phase (Supplementary Fig. 2). We then repeated the two-way-imputation–based analysis using a non-overlapping sample that excluded the 1,044 single-tracked cases (Table 1, see the Iceland follow-up two-way–imputation category). Combining the results from the two follow-up groups provided strong evidence for the association of rs78378222 with BCC (OR = 2.25, $P = 5.4 \times 10^{-19}$; Table 1, see the Iceland follow-up phase combined category). Association based on all 2,322 cases who had been single-track genotyped (irrespective of whether or not they had been chip typed) produced comparable results (Supplementary Table 1).

We then typed rs78378222 in replication samples from Denmark, eastern Europe and Spain (the replication sets are described in Supplementary Table 2). We found the risk allele in all the populations tested, with frequencies that seemed to decline with each population's distance from Iceland (Table 1). Combined, the evidence for replication of the BCC association in non-Icelandic samples was significant and showed no evidence of heterogeneity (OR = 1.75, P = 0.0060, P of heterogeneity (P_{het}) = 0.27; Table 1). Combined with the Icelandic follow-up phase data, the overall association was highly significant (OR = 2.16, $P = 2.2 \times 10^{-20}$).

The second strongest signal in the genome originated from a previously unreported SNP (designated chr17:7,640,788; OR = 2.41, $P = 1.1 \times 10^{-13}$) which was also in the region covered by the two candidate haplotypes (Fig. 1). Based on single-track genotyping of 2,281 cases and 6,858 controls for both SNPs, chr17:7,640,788 was correlated (r^2 = 0.61) with rs78378222 and so may capture the same signal. When adjusted for the effect of rs78378222, there was no residual signal from chr17:7,640,788 (OR$_{adj}$ = 1.07, P_{adj} = 0.72), whereas rs78378222 remained significant after adjustment for chr17:7,640,788 (OR$_{adj}$ = 2.00, $P_{adj} = 3.0 \times 10^{-5}$). Therefore, we did not investigate chr17:7,640,788 further. A common germline variant in *TP53* (rs1042522, p.Pro72Arg) has been studied extensively for cancer susceptibility, generally with equivocal results[11]. We saw no evidence that rs1042522 was associated with BCC (OR = 1.00, P = 0.98). We confirmed this by single-track genotyping. Thus, it appears that p.Pro72Arg does not confer risk of BCC and is unrelated to the rs78378222 signal.

p53 is induced by ultraviolet irradiation of skin and is a primary mediator of the tanning response[12,13]. We looked for associations between rs78378222 and sensitivity of skin to sun in 11,131 samples from Iceland and The Netherlands[14,15]. There was no significant association between rs78378222[C] and self-reported sun sensitivity (Fitzpatrick score I and II compared to III and IV; OR = 1.14, P = 0.23, 95% confidence interval (CI) 0.92–1.41). We also noted that rs78378222[C] was somewhat more frequent in individuals with tumors

at sun-exposed sites (0.0412) than in those with tumors at non–sun-exposed sites (0.0302); however, this difference was not significant ($P = 0.18$).

Inspection of the sequence surrounding rs78378222 indicated that it occurs in the sole polyadenylation signal of *TP53*, with the risk-associated variant changing the sequence AATAAA to AATACA, thus disrupting the signal sequence. This class of mutations was first observed in the polyadenylation signal of *HBA2* (encoding alpha 2 globin), leading to alpha-thalassemia[16]. We obtained RNA from blood and adipose tissue from rs78378222[A/C] heterozygotes and rs78378222[A/A] homozygotes. Using RT-PCR with primers internal to the *TP53* mRNA, we observed that rs78378222[A/C] heterozygotes expressed somewhat less *TP53* transcript than wild-type homozygotes ($P = 0.041$; Supplementary Fig. 3a). To investigate polyadenylation site usage of wild-type and variant *TP53*, we selected total RNA samples from rs78378222[A/C] heterozygotes and carried out 3′ rapid amplification of complementary DNA ends (RACE; Supplementary Fig. 3b). Amplification using a *TP53* gene-specific forward primer produced a band of 1.3 kb, which is the expected length of correctly terminated mRNA (Supplementary Fig. 3c). However, sequencing of this band indicated that correctly terminated polyA(+) mRNAs were produced predominantly from the wild-type allele, with 73% of mRNAs containing the wild-type A allele and 27% containing the variant C allele ($P = 1.6 \times 10^{-6}$; Supplementary Fig. 3c). We then carried out RT-PCR using a 'run-on' reverse primer, located in the genomic sequence approximately 320 bp beyond the normal 3′ end of *TP53* (Supplementary Fig. 3b). Sequencing of RT-PCR products from heterozygotes showed that this RNA species was comprised almost entirely of variant C allele transcripts (Supplementary Fig. 3d). Taken together, these data suggest that the rs78378222[C] variant impairs proper termination and polyadenylation of the *TP53* transcript.

Next, we searched for associations between rs78378222 and 20 major tumor types by cross-referencing genotypes to the Icelandic Cancer Registry and national pathology records. After correcting for multiple phenotype testing, we observed significant associations for prostate cancer, brain cancers and colorectal adenoma (but not colorectal cancer) (Supplementary Table 3). We conducted a follow-up phase in an attempt to get further evidence of these associations. The follow-up phase was analogous to that described in Supplementary Figure 2, and the sample numbers are detailed in Supplementary Table 4. We directly genotyped rs78378222 in all available Icelandic cases with prostate cancer, colorectal adenoma or brain cancer and determined follow-up phase two-way–imputation-based and single-track genotype-based association values (Table 2). For prostate cancer, we further genotyped replication samples from five countries. The association with prostate cancer outside Iceland was significant (OR = 1.63, $P = 1.1 \times 10^{-4}$), as was the combined Iceland follow-up and replication sample result (OR = 1.44, $P = 2.4 \times 10^{-6}$; Table 2). For the colorectal adenoma follow-up phase, single-track genotyping gave a comparable result to two-way imputation, and the combined analysis yielded a significant association result (OR = 1.39, $P = 1.6 \times 10^{-4}$; Table 2 and Supplementary Table 1). However, we still observed no significant association with colorectal cancer after genotyping replication samples from four countries (combined OR = 1.06, $P = 0.51$, 95% CI 0.89–1.27; Supplementary Table 5). This raises the possibility that rs78378222[C] might predispose to colorectal adenomas with a low propensity for progression to invasive cancer.

The discovery two-way–imputation category of 'all brain cancers' contained meningiomas (International Classification of Diseases (ICD) 10 code C70) and gliomas (C71–C72). When glioma and meningioma were considered separately, the association appeared stronger for glioma (OR = 2.50, $P = 0.0055$) than for meningioma (OR = 1.48, $P = 0.36$). Although this difference was not significant ($P = 0.88$), we focused on glioma in the follow-up phase (Table 2). In Iceland, the follow-up phase yielded a suggestive association with glioma (OR

= 2.36, P = 0.0036). This was confirmed in two case-control samples of adult glioma from the United States (OR = 2.34, P = 9.2×10^{-4}). Combined with the Icelandic data, we obtained firm evidence that rs78378222[C] was associated with glioma (OR = 2.35, P = 1.0 $\times 10^{-5}$; Table 2).

We also investigated potential associations with melanoma (because of the link between rs78378222 and skin cancer risk) and breast cancer (because of the involvement of *TP53* mutations in Li-Fraumeni and Li-Fraumeni–like syndromes). We did not see any convincing evidence that rs78378222 confers risk of either of these tumors (OR = 1.07, P = 0.64, 95% CI 0.81–1.42 for melanoma and OR = 1.06, P = 0.57, 95% CI 0.88–1.27 for breast cancer; Supplementary Table 5). There was no evidence of specific associations with estrogen-receptor–negative breast cancer (OR = 1.15, P = 0.61) or high-risk breast cancer (defined as an age at diagnosis under 50 or a history of multiple primary breast cancers; OR = 0.87, P = 0.36).

Given its central role in tumor biology and the high frequency of somatic mutation[17], intensive efforts have been devoted to searching for germline cancer susceptibility variants in *TP53*. Aside from gain-of-function mutations typically found in Li-Fraumeni and Li-Fraumeni–like syndromes, no germline variants of *TP53* have so far been implicated reproducibly in cancer predisposition[11,18]. Cross-referencing the Icelandic genealogical database with cancer registry records did not reveal any rs78378222[C] carriers with family histories fitting the criteria of Li-Fraumeni or Li-Fraumeni–like syndrome[19,20]. *TP53* mutations leading to Li-Fraumeni and Li-Fraumeni–like syndromes are rare, occurring in 1 in 5,000 to 1 in 20,000 births[21,22]. Although the penetrance of rs78378222[C] for the cancers it affects is much less, carriers are expected in frequencies of up to 4% in some populations. The effect of rs78378222[C] on 3′-end processing suggests a new mechanism by which *TP53* may promote oncogenesis, with a distinctive spectrum of tumors being affected.

URLs. Picard version 1.17, http://picard.sourceforge.net/.

Methods

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.
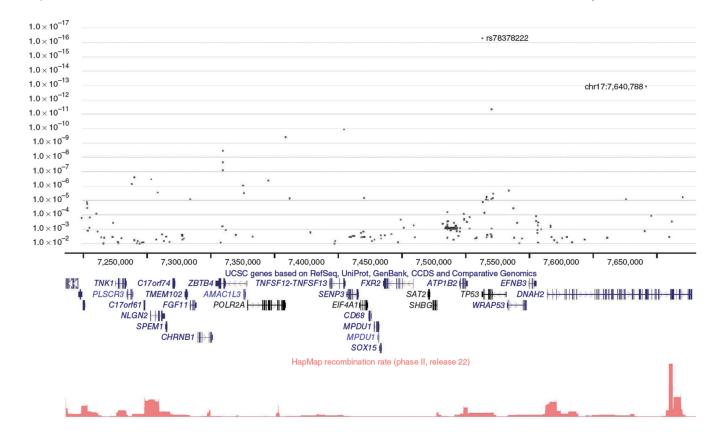
Acknowledgments

References

1. de Zwaan SE, Haass NK. Genetics of basal cell carcinoma. Australas. J. Dermatol. 2010; 51:81–92. [PubMed: 20546211]

2. Epstein EH. Basal cell carcinomas: attack of the hedgehog. Nat. Rev. Cancer. 2008; 8:743–754. [PubMed: 18813320]

3. Box NF, et al. Melanocortin-1 receptor genotype is a risk factor for basal and squamous cell carcinoma. J. Invest. Dermatol. 2001; 116:224–229. [PubMed: 11179997]

4. Gudbjartsson DF, et al. *ASIP* and *TYR* pigmentation variants associate with cutaneous melanoma and basal cell carcinoma. Nat. Genet. 2008; 40:886–891. [PubMed: 18488027]

5. Rafnar T, et al. Sequence variants at the *TERT-CLPTM1L* locus associate with many cancer types. Nat. Genet. 2009; 41:221–227. [PubMed: 19151717]

6. Stacey SN, et al. Common variants on 1p36 and 1q42 are associated with cutaneous basal cell carcinoma but not with melanoma or pigmentation traits. Nat. Genet. 2008; 40:1313–1318. [PubMed: 18849993]

7. Stacey SN, et al. New common variants affecting susceptibility to basal cell carcinoma. Nat. Genet. 2009; 41:909–914. [PubMed: 19578363]

8. Kong A, et al. Detection of sharing by descent, long-range phasing and haplotype imputation. Nat. Genet. 2008; 40:1068–1075. [PubMed: 19165921]

9. Kong A, et al. Parental origin of sequence variants associated with complex diseases. Nature. 2009; 462:868–874. [PubMed: 20016592]

10. Kutyavin IV, et al. A novel endonuclease IV post-PCR genotyping system. Nucleic Acids Res. 2006; 34:e128. [PubMed: 17012270]

11. Whibley C, Pharoah PD, Hollstein M. p53 polymorphisms: cancer implications. Nat. Rev. Cancer. 2009; 9:95–107. [PubMed: 19165225]

12. Cui R, et al. Central role of p53 in the suntan response and pathologic hyperpigmentation. Cell. 2007; 128:853–864. [PubMed: 17350573]

13. Miller AJ, Tsao H. New insights into pigmentary pathways and skin cancer. Br. J. Dermatol. 2010; 162:22–28. [PubMed: 19863502]

14. Sulem P, et al. Two newly identified genetic determinants of pigmentation in Europeans. Nat. Genet. 2008; 40:835–837. [PubMed: 18488028]

15. Sulem P, et al. Genetic determinants of hair, eye and skin pigmentation in Europeans. Nat. Genet. 2007; 39:1443–1452. [PubMed: 17952075]

16. Higgs DR, et al. Alpha-thalassaemia caused by a polyadenylation signal mutation. Nature. 1983; 306:398–400. [PubMed: 6646217]

17. Junttila MR, Evan GI. p53—a Jack of all trades but master of none. Nat. Rev. Cancer. 2009; 9:821–829. [PubMed: 19776747]

18. Palmero EI, Achatz MI, Ashton-Prolla P, Olivier M, Hainaut P. Tumor protein-53 mutations and inherited cancer: beyond Li-Fraumeni syndrome. Curr. Opin. Oncol. 2010; 22:64–69. [PubMed: 19952748]

19. Li FP, et al. A cancer family syndrome in twenty-four kindreds. Cancer Res. 1988; 48:5358–5362. [PubMed: 3409256]

20. Birch JM, et al. Prevalence and diversity of constitutional mutations in the p53 gene among 21 Li-Fraumeni families. Cancer Res. 1994; 54:1298–1304. [PubMed: 8118819]

21. Gonzalez KD, et al. Beyond Li Fraumeni syndrome: clinical characteristics of families with p53 germline mutations. J. Clin. Oncol. 2009; 27:1250–1256. [PubMed: 19204208]

22. Lalloo F, et al. Prediction of pathogenic mutations in patients with early-onset breast cancer by family history. Lancet. 2003; 361:1101–1102. [PubMed: 12672316]

23. McVean GA, et al. The fine-scale structure of recombination rate variation in the human genome. Science. 2004; 304:581–584. [PubMed: 15105499]

Figure 1.
Overview of single-point SNP association data obtained from genomic sequencing in the 17p13 region covered by haplotypes A6 and A8. The region shown is chr17:7,186,095–7,680,389 (HG18 Build 36). The upper panel shows BCC association *P* values for SNPs in the region identified by whole-genome sequencing of 457 individuals. We determined association by two-way imputation (see the text for details); only SNPs with *P* < 0.01 are plotted. The positions of the *TP53* SNP rs78378222 and the newly discovered SNP giving the second-highest signal in the region (chr17:7,640,788) are indicated. The locations of UCSC genes in the region are shown in the middle panel. The lower panel shows recombination rates calculated as described previously[23] from HapMap2 release 22 data.

Table 1

Association between BCC and rs78378222[C]

Sample set	P	OR	95% CI	N_{cases}	Frequency in cases	$N_{controls}$	Frequency in controls	P^a_{het}
Iceland discovery phase two-way imputation	5.2×10^{-17}	2.36	1.93–2.89	$2,121^b$	0.0442	$>39,614^b$	0.0192^c	
Iceland follow-up phase two-way imputation[d]	3.1×10^{-13}	2.28	1.83–2.85	$1,813^b$	0.0428	$>36,709^b$	0.0192^c	
Iceland follow-up phase single-track genotyped[d]	3.2×10^{-7}	2.19	1.62–2.96	1,044	0.0398	7,200	0.0185	
Iceland follow-up phase combined	$\mathbf{5.4 \times 10^{-19}}$	**2.25**	**1.88–2.69**	$\mathbf{2,857}^b$	**NA**	$\mathbf{>43,909}^b$	**NA**	**0.83**
Denmark	0.0057	2.06	1.23–3.44	308	0.0341	3,441	0.0168	
Eastern Europe[e]	0.12	2.04	0.84–4.93	526	0.0133	532	0.0066	
Spain	0.77	0.86	0.34–2.21	628	0.0040	3,928	0.0046	0.79
Combined non-Icelandic replication	**0.0060**	**1.75**	**1.17–2.62**	**1,462**	**NA**	**7,901**	**NA**	**0.27**
Combined Iceland follow up and replication	$\mathbf{2.2 \times 10^{-20}}$	**2.16**	**1.83–2.54**	$\mathbf{4,319}^b$	**NA**	$\mathbf{>51,810}^b$	**NA**	**0.27**

NA, not applicable; OR, odds ratio; CI, confidence interval.

[a] The P value for heterogeneity based on combining groups using the Mantel-Haenszel model.

[b] Effective sample size estimate taking into account the efficiency of the Icelandic genealogy-based *in silico* genotyping.

[c] The control frequency given is derived from 40,309 directly chip-typed control individuals without BCC.

[d] The Iceland follow-up phase two-way imputation set consists of individuals who were chip-typed and individuals whose genotypes were determined by genealogy-based *in silico* genotyping. This set differs from the Iceland discovery phase two-way imputation set in that it lacks those individuals whose genotypes were determined by genealogy-based *in silico* genotyping in the discovery phase but who were subsequently directly single-track genotyped. Such individuals form part of the Iceland follow-up phase single-track genotyped set, which consists of individuals whose genotypes were determined solely by a Centaurus single-track assay. The follow-up phase two-way imputation and single-track genotyped sets are non-overlapping (Supplementary Fig. 2).

[e] Samples from Hungary, Slovakia and Romania.

Table 2

Association between rs78378222[C] and prostate cancer, glioma and colorectal adenoma

Sample set	Tumor type	P	OR	95% CI	N_{cases}	Frequency in cases	$N_{controls}$	Frequency in controls	P^a_{het}
Iceland follow-up phase two-way imputation[b]	Prostate cancer	0.0058	1.35	1.09–1.67	2,671c	0.0258	>36,331c	0.0192d	
Iceland follow-up phase single-track genotyped[b]	Prostate cancer	0.27	1.28	0.82–1.99	635	0.0236	7,200	0.0185	
Iceland follow-up phase combined	**Prostate cancer**	**0.0030**	**1.34**	**1.10–1.62**	**3,306c**	**NA**	**>43,531c**	**NA**	**0.83**
The Netherlands	Prostate cancer	0.093	1.41	0.94–2.10	1,085	0.0207	1,796	0.0147	
UK	Prostate cancer	0.24	1.39	0.80–2.42	521	0.0192	1,407	0.0139	
Romania	Prostate cancer	0.41	1.38	0.64–2.95	639	0.0110	815	0.0080	
United States	Prostate cancer	0.0026	2.24	1.33–3.79	1,454	0.0155	1,293	0.0070	
Spain	Prostate cancer	0.037	2.52	1.06–5.99	785	0.0070	1,787	0.0028	
Combined non-Icelandic replication	**Prostate cancer**	**1.1×10^{-4}**	**1.63**	**1.27–2.09**	**4,484**	**NA**	**7,098**	**NA**	**0.5**
Combined Iceland follow up and replication	**Prostate cancer**	**2.4×10^{-6}**	**1.44**	**1.24–1.68**	**7,790c**	**NA**	**>50,629c**	**NA**	**0.43**
Iceland follow-up phase two-way imputation[b]	Glioma	0.0074	2.61	1.29–5.27	135c	0.0486	>37,881c	0.0192d	
Iceland follow-up phase single-track genotyped[b]	Glioma	0.22	1.90	0.69–5.27	72	0.0347	7,200	0.0185	
Iceland follow-up phase combined	**Glioma**	**0.0036**	**2.36**	**1.32–4.20**	**207c**	**NA**	**>45,081c**	**NA**	**0.62**
USA UCSF	Glioma	0.0063	2.65	1.32–5.32	658	0.0228	573	0.0087	
USA Mayo Clinic	Glioma	0.053	2.05	0.99–4.25	530	0.0321	283	0.0159	
Combined non-Icelandic replication	**Glioma**	**9.2×10^{-4}**	**2.34**	**1.42–3.88**	**1,188**	**NA**	**856**	**NA**	**0.62**
Combined Iceland follow up and replication	**Glioma**	**1.0×10^{-5}**	**2.35**	**1.61–3.44**	**1,395c**	**NA**	**>45,937c**	**NA**	**0.88**
Iceland follow-up phase two-way imputation[b]	Colorectal adenoma	0.0064	1.32	1.08–1.61	3,057c	0.0252	>36,022c	0.0192d	
Iceland follow-up phase single-track genotyped[b]	Colorectal adenoma	0.0052	1.60	1.15–2.23	1,038	0.0294	7,200	0.0185	
Iceland follow-up phase combined	**Colorectal adenoma**	**1.6×10^{-4}**	**1.39**	**1.17–1.65**	**4,095c**	**NA**	**>43,022c**	**NA**	**0.33**

NA, not applicable; OR, odds ratio; CI, confidence interval.

[a]The P value for heterogeneity based on combining groups using the Mantel-Haenszel model.

[b]The Iceland follow-up phase two-way imputation set consists of individuals who were chip-typed and individuals whose genotypes were determined by genealogy-based *in silico* genotyping. This set differs from the Iceland discovery phase two-way imputation set shown in Supplementary Table 2 in that it lacks those individuals whose genotypes were determined by genealogy-based *in silico* genotyping in the discovery phase but who were subsequently directly single-track genotyped. Such individuals form part of the Iceland follow-up phase single-track genotyped set, which consists of

individuals whose genotypes were determined solely by Centaurus single-track assay. The follow-up phase two-way imputation and single-track genotyped sets are non-overlapping. See Supplementary Figure 2 and Supplementary Table 3 for further details.

[c]The effective sample size estimate taking into account efficiency of Icelandic genealogy-based *in silico* genotyping.

[d]The control frequency given is derived from 40,309 directly chip-typed control individuals without BCC.