# Common variants affecting susceptibility to develop multiple basal cell carcinomas

**J.A.C. Verkouteren***
L.M. Pardo*
A.G. Uitterlinden
A. Hofman
T. Nijsten
* shared first authorship

*J Invest Dermatol. 2015 Aug;135(8):2135-2138.*

**To the editor**

Among the millions of people who develop a first basal cell carcinoma (BCC) annually, ~30% will get subsequent BCCs (Flohil et al., 2013). The majority of BCCs occur on the head and neck, where tumor growth and surgery can lead to functional and cosmetic morbidity. Because of the high incidence, risk of multiple tumors, and morbidity, the disease burden and healthcare costs are considerable (Housman et al., 2003; Hollestein et al., 2014).

Several candidate gene approaches (CGAs) suggest that polymorphisms in the genes encoding cytochrome P450, glutathione S-transferase, and HLA are implicated in the development of multiple BCCs (mBCC; Cerimele et al., 1988; Rompel et al., 1995; Lear et al., 1996; Ramachandran et al., 2000). Most of these studies have a small sample size and include only a few variants per gene. To date, there are no studies investigating whether more recently identified BCC loci also confer susceptibility to mBCC. We investigated whether single-nucleotide polymorphisms (SNPs) previously associated with BCC increase the risk of mBCC using a CGA. In addition, we conducted a genome-wide association study (GWAS) to identify previously unreported loci associated with the risk of mBCC.

A detailed description of all the methods is presented in the Supplementary Material online. We used participants from the Rotterdam Study (RS), which is a population-based follow-up study that consists of three cohorts (RS-I, II, and III; Hofman et al., 2013). The Medical Ethics Committee of the Erasmus Medical Center and the review board of the Dutch Ministry of Health, Welfare and Sport have ratified the RS. All participants who gave informed consent were linked with a nationwide registry of histopathology in The Netherlands (PALGA, up to 31 December 2013) to identify histopathologically confirmed BCCs, squamous cell carcinomas (SCCs), and melanomas (Casparie et al., 2007).

DNA from whole blood was extracted and genotyped following standard protocols (Hofman et al., 2013). Quality control procedures were applied to the genotyped SNP data. The GWA data sets were imputed to the 1,000 Genomes data set using MACH-minimac v1.0.18 (Howie et al., 2012). In total, 30,072,738 markers were genotyped and/or imputed. We excluded markers with a minor allele frequency <3% and an imputation quality <0.3. After quality control, 7,260,691 markers were available for analysis.

From the 9,810 RS participants with genotype and phenotype information, 1,219 individuals with BCC were identified, of whom 472 had mBCC (38.7%). Participants with mBCC had a significantly higher proportion of SCCs and/or melanomas compared with those with single BCC (sBCC; Supplementary Table S1 online).

First, 19 candidate SNPs and 17 loci from well-powered GWASs/CGA of BCC were selected (Supplementary Table S2 and Supplementary Material online). We then

conducted SNP- and gene-based logistic regression analyses in two data sets, comparing BCC to no BCC (i.e., validation set) and mBCC to sBCC to investigate whether these BCC loci also increase the risk of mBCC. The analyses were adjusted for age at study entry or age at first BCC, sex, and four principal components (PCs). As these SNPs and loci were previously significantly associated with BCC, we only adjusted for multiple testing in the mBCC versus sBCC data set using the Bonferroni correction. All analyses were performed in PLINK v1.07 (Purcell et al., 2007).

In the CGA on BCC against no BCC, 12/19 (63%) candidate SNPs and 5/17 (29%) loci were replicated, demonstrating a good external validity of the study (Table 1 and Supplementary Table S3A online). Interestingly, the CGA comparing sBCC to mBCC did not yield any significant associations between these BCC-related SNPs/loci and risk of mBCC (Table 1 and Supplementary Table S3B online).

**Table 1.** SNP-based and gene-based association analyses

| | | BCC vs. no BCC | | | mBCC vs. sBCC | | |
|---|---|---|---|---|---|---|---|
| **SNP** | **Locus** | **SNP-based p-value** | **Gene-based p-value** | **SNP-based OR** | **SNP-based p-value** | **Gene-based p-value** | **SNP-based OR** |
| rs1126809 | TYR | **0.0027** | 0.14 | 1.16 | 0.99 | 0.16 | 1.0010 |
| rs4911414 | 20q11.22 | 0.067 | 0.96 | 1.086 | 0.49 | 0.77 | 0.94 |
| rs1015362 | 20q11.22 | 0.83 | 0.96 | 1.010 | 0.15 | 0.77 | 0.87 |
| rs7538876 | PADI6 | **0.00062** | 0.075 | 1.16 | 0.063 | 0.11 | 1.17 |
| rs801114 | 1q42.13 | **0.0070** | **0.049** | 1.13 | 0.63 | 0.75 | 0.96 |
| rs11170164 | KRT5 | 0.17 | **0.0020** | 1.12 | 0.13 | 0.24 | 1.27 |
| rs2151280 | CDKN2B-AS1 | **0.0043** | 0.13 | 0.88 | 0.97 | 0.64 | 1.00 |
| rs157935 | LINC-PINT | **0.00077** | 0.10 | 0.85 | 0.18 | 0.86 | 0.88 |
| rs16891982 | SLC45A2 | **0.0081** | **0.038** | 0.65 | 0.82 | 0.47 | 0.93 |
| rs401681 | CLPTM1L | **0.00053** | **0.022** | 0.86 | 0.36 | 0.83 | 0.92 |
| rs12210050 | EXOC2 | **0.030** | 0.35 | 1.15 | 0.69 | 0.13 | 0.95 |
| rs7335046 | UBAC2 | 0.19 | 0.35 | 1.087 | 0.10 | 0.016 | 1.22 |
| rs1805007 | MC1R | 0.069 | 0.60 | 1.16 | 0.43 | 0.25 | 0.88 |
| rs78378222 | TP53 | **0.037** | 0.17 | 1.34 | 0.13 | 0.39 | 1.49 |
| rs12203592 | IRF4 | **7.7E-05** | 0.063 | 1.31 | 0.35 | 0.30 | 0.89 |
| rs12202284 | EXOC2 | 0.12 | 0.35 | 1.096 | 0.88 | 0.13 | 1.017 |
| rs8015138 | 14q22.1 | 0.79 | 0.088 | 0.99 | 0.91 | 0.72 | 0.99 |
| rs214782 | TGM3 | **1.2E-05** | **0.0070** | 1.27 | 0.056 | 0.19 | 1.22 |
| rs7006527 | RGS22 | **0.0018** | 0.17 | 0.82 | 0.49 | 0.85 | 1.088 |

Abbreviations: BCC, basal cell carcinoma; mBCC, multiple basal cell carcinomas; OR, odds ratio; sBCC, single basal cell carcinoma; SNP, single-nucleotide polymorphism. Numbers in bold display significant differences (P-value <0.05). The gene-based P-value for mBCC versus sBCC should be corrected for multiple testing.

Finally, we conducted a pilot GWAS using logistic regression (additive model) to test for associations between markers and mBCC, adjusting for age at diagnosis, sex, and four PCs (Supplementary Material online). A meta-analysis of the GWAS results per cohort was performed. Despite the low overall power to detect genome-wide significant hits (Supplementary Figure S1 online), we identified genome-wide suggestive associations in chromosomes 2, 3, 18, and 22 (P-values $<5 \times 10^{-6}$, Supplementary Figure S2 online and Table 2). The most significant SNP was rs78857623 (P-value $=1.2 \times 10^{-7}$, Table 2), which mapped to an intron in FHIT, and it was in linkage disequilibrium ($r^2=0.65$) with another significantly associated intronic FHIT SNP (rs78316259; P-value$=4.6 \times 10^{-6}$). FHIT is a tumor suppressor gene that encodes a diadenosine polyphosphate hydrolase involved in purine metabolism (Barnes et al., 1996). Aberrant FHIT transcripts as well as germline mutations in this gene have been found in different cancers including BCC (Ohta et al., 1996; Goldberg et al., 2006; Ding et al., 2008).

Given the high proportion of SCCs and melanomas in mBCC cases, a sensitivity analysis investigating the influence of other cutaneous cancers was performed (Supplementary Table S1 online). We observed changes in the P-values of the associations due to a 13% decrease in sample size, but all top SNPs remained significant with a P-value $<5 \times 10^{-5}$ (data not shown), showing that our findings were driven by BCC cases.

In contrast to other BCC GWASs, we performed a GWAS on histopathologically confirmed mBCC. By combining national pathology data with genomewide SNP data from a population-based study, we accurately distinguished between sBCC and mBCC. It is a pilot GWAS because the sample size is small and replication data are not easily available. All existing cohorts, which have performed genetic epidemiology on skin cancer, do not have data on mBCC. Like in other GWASs, the significant associations are only statistical and therefore any inference about the functional impact of the variants to the risk of mBCC needs to be investigated with other approaches. Despite these limitations, our data set contains the largest collection of cases with mBCC to date and may serve as a valuable reference for future studies.

Because of the high risk of subsequent BCCs in individuals with a first BCC, we expected that loci predisposing to sBCCs would also contribute to the risk of mBCC. However, the CGA analysis did not support this hypothesis, suggesting that other biological factors, including genes, may confer an increased risk to mBCC. The earlier onset of first BCCs in patients with mBCC may indeed suggest that these patients have a stronger genetic burden compared with those with sBCC (68.7 vs. 72.4 years; P-value $<0.05$, Supplementary Table S1 online). A heritability analysis comparing mBCC against sBCC in well-powered samples will help validate this hypothesis. It could also be argued that other yet-to-be-identified loci conferring susceptibility to BCC may also increase risk of mBCC, which will require larger consortia on the genetics of BCC.

**Table 2.** SNPs with highly significant associations from GWAS on mBCC

| Chr. | Marker | Frequency[1] | P-value | Direction[2] | HetISq[3] | HetPVal[4] | Gene | Functional consequence marker |
|---|---|---|---|---|---|---|---|---|
| 3 | rs78857623 | 0.048 | 1.2E-07 | --- | 0.4 | 0.37 | FHIT | intronic |
| 2 | 2:168946822:D | 0.084 | 3.9E-07 | --- | 0 | 0.53 | STK39 | intronic |
| 18 | rs4371253 | 0.18 | 2.1E-06 | +++ | 0 | 0.76 | SLC14A2 | intronic |
| 22 | rs4824031 | 0.34 | 2.2E-06 | --- | 0 | 0.88 | | intergenic |
| 9 | rs17717641 | 0.94 | 3.4E-06 | --- | 0 | 0.57 | | intergenic |
| 16 | rs41305755 | 0.20 | 3.6E-06 | --- | 56.1 | 0.10 | ABAT | intronic |
| 16 | rs45545237 | 0.80 | 3.6E-06 | +++ | 56.1 | 0.10 | ABAT | intronic |
| 6 | rs79899616 | 0.96 | 3.6E-06 | +++ | 0 | 0.57 | | intergenic |
| 22 | rs8138971 | 0.38 | 3.6E-06 | --- | 0 | 0.92 | | intergenic |
| 18 | rs4890560 | 0.19 | 3.7E-06 | +++ | 0 | 0.87 | SLC14A2 | intronic |
| 18 | rs4890291 | 0.19 | 3.8E-06 | +++ | 0 | 0.87 | SLC14A2 | intronic |
| 11 | rs80160790 | 0.031 | 4.0E-06 | --? | 0 | 0.58 | GRIA4 | intronic, upstream variant 2KB & 5'UTR variant |
| 22 | rs9616609 | 0.36 | 4.1E-06 | --- | 0 | 0.80 | | intergenic |
| 18 | rs8093237 | 0.18 | 4.2E-06 | +++ | 0 | 0.75 | SLC14A2 | intronic |
| 18 | rs11875624 | 0.18 | 4.5E-06 | +++ | 0 | 0.74 | SLC14A2 | intronic |
| 18 | 18:43175190:D | 0.18 | 4.5E-06 | +++ | 0 | 0.74 | | deletion |
| 3 | rs78316259 | 0.057 | 4.6E-06 | --- | 0 | 0.89 | FHIT | intronic |
| 9 | rs1412279 | 0.46 | 5.2E-06 | --- | 0 | 0.47 | | intergenic |
| 9 | 9:78186385:D | 0.065 | 5.2E-06 | +++ | 0 | 0.69 | | deletion |
| 9 | rs75858454 | 0.065 | 5.2E-06 | +++ | 0 | 0.63 | | intergenic |
| 9 | rs74403342 | 0.94 | 5.4E-06 | --- | 0 | 0.68 | | intergenic |

Abbreviations: Chr., chromosome; GWAS, genome-wide association study; mBCC, multiple basal cell carcinomas; SNP, single-nucleotide polymorphism; UTR, untranslated region.

[1] Frequency: weighted average of frequency for allele 1 across all cohorts.

[2] Direction: summary of effect direction for each study, with one '+' or '–' per cohort ('?' is unknown).

[3] HetISq: $I^2$ statistic which measures heterogeneity on scale of 0–100%.

[4] HetPVal: P-value for heterogeneity statistic.

In conclusion, genetic loci previously associated with BCC do not increase the risk of mBCC. A pilot GWAS on mBCC identified to our knowledge previously unreported susceptibility variants, but these findings need to be replicated in other mBCC cohorts.

## ACKNOWLEDGEMENTS

This research was funded by a Vidi Grant of ZonMw (91711315). The Rotterdam Study is supported by the Netherlands Organisation of Scientific Research NWO Investments (175.010.2005.011, 911-03-012) and further funded by the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NOW, 050-060-810), Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. We thank the department of internal medicine and Karol Estrada and Maksim V Struchalin for their work on the genetic data. The authors are grateful to the study participants, the staff from the Rotterdam Study, and the participating general practitioners and pharmacists. We further thank Esther van den Broek and Lucy Overbeek from the PALGA foundation and dermatopathologist Senada Koljenovic from the Erasmus Medical Center.

## REFERENCES

Barnes LD, Garrison PN, Siprashvili Z et al. (1996) Fhit, a putative tumor suppressor in humans, is a dinucleoside 5′,5‴-P1,P3-triphosphate hydrolase. Biochemistry 35:11529–35

Casparie M, Tiebosch AT, Burger G et al. (2007) Pathology databanking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. Cell Oncol 29: 19–24

Cerimele D, Contu L, Carcassi C et al. (1988) HLA and multiple skin carcinomas. Dermatologica 176:176–81

Ding Y, Larson G, Rivas G et al. (2008) Strong signature of natural selection within an FHIT intron implicated in prostate cancer risk. PLoS One 3:e3533

Flohil SC, van der Leest RJ, Arends LR et al. (2013) Risk of subsequent cutaneous malignancy in patients with prior keratinocyte carcinoma: a systematic review and meta-analysis. Eur J Cancer 49:2365–75

Goldberg M, Rummelt C, Laerm A et al. (2006) Epigenetic silencing contributes to frequent loss of the fragile histidine triad tumour suppressor in basal cell carcinomas. Br J Dermatol 155:1154–8

Hofman A, Darwish Murad S, van Duijn CM et al. (2013) The Rotterdam Study: 2014 objectives and design update. Eur J Epidemiol 28: 889–926

Hollestein LM, de Vries E, Aarts MJ et al. (2014) Burden of disease caused by keratinocyte cancer has increased in The Netherlands since 1989. J Am Acad Dermatol 71: 896–903

Housman TS, Feldman SR, Williford PM et al. (2003) Skin cancer is among the most costly of all cancers to treat for the Medicare population. J Am Acad Dermatol 48:425–9

Howie B, Fuchsberger C, Stephens M et al. (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet 44:955–9

Lear JT, Heagerty AH, Smith A et al. (1996) Multiple cutaneous basal cell carcinomas: glutathione S-transferase (GSTM1, GSTT1) and cytochrome P450 (CYP2D6, CYP1A1) polymorphisms influence tumour numbers and accrual. Carcinogenesis 17:1891–6

Ohta M, Inoue H, Cotticelli MG et al. (1996) The FHIT gene, spanning the chromosome 3p14.2 fragile site and renal carcinoma-associated t(3;8) breakpoint, is abnormal in digestive tract cancers. Cell 84:587–97

Purcell S, Neale B, Todd-Brown K et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–75

Ramachandran S, Hoban PR, Ichii-Jones F et al. (2000) Glutathione S-transferase GSTP1 and cyclin D1 genotypes: association with numbers of basal cell carcinomas in a patient subgroup at high-risk of multiple tumours. Pharmacogenetics 10: 545–56

Rompel R, Petres J, Kaupert K et al. (1995) Human leukocyte antigens and multiple basal cell carcinomas. Recent Results Cancer Res 139: 297–302

## SUPPLEMENTARY MATERIAL

## Materials and methods

*Study population*

The Rotterdam Study (RS) is a prospective population-based follow-up study of the determinants and prognosis of chronic diseases, including skin diseases and cancer, in the elderly. Only participants living in Rotterdam, The Netherlands, are included. The RS consists of a major cohort (RS-I) and two extensions (RS-II and RS-III). RS-I started in 1990 and initially included 7,983 participants living in the Ommoord district in Rotterdam. RS-II started in 2000 and now includes 3,011 participants. RS-III is a further extension of the cohort, started in 2006, and now includes 3,932 participants. By the end of 2008, the RS comprised 14,926 subjects aged 45 years or over. The overall response rate for all three cycles at baseline was 72.0%. The cohort consists predominantly (90%) of participants of North-European ancestry. A detailed description of the design of the RS is presented in Hofman et al. (2013). The Medical Ethics Committee of the Erasmus Medical Center and the review board of the Dutch Ministry of Health, Welfare and Sport have ratified the RS. Our study was conducted according to Declaration of Helsinki Principles. From each participant, written informed consent was obtained.

*Ascertainment of (multiple) basal cell carcinoma cases*

All the participants of the RS who gave informed consent (n = 14,628) were linked with the nationwide network and registry of histopathology and cytopathology in The Netherlands (PALGA; up to 31st December 2013) to identify histopathologically confirmed BCCs, squamous cell carcinomas and melanomas. PALGA was founded in 1971 and achieved complete national coverage in 1991. Every obtained pathology excerpt contains encrypted patient data, a report identifier, the conclusion of the pathologist and a PALGA diagnosis line derived from the Systematized Nomenclature of Medicine (Casparie et al., 2007). The majority of excerpts extracted from PALGA include an anatomical location, which makes it possible to distinguish between participants with single or subsequent tumors. If location was not available, we assumed that a biopsy followed by an excision for the same type of skin cancer, concerned the same tumor. The next BCC following a radical excision was always counted as a new neoplasm. If an excision was incomplete, the next reported tumor on the same/adjacent location was regarded as recurrent and not a new BCC. If the diagnosis or the number of unique BCCs remained unclear, the medical files were searched by hand and a consensus decision was made based on these data.

*DNA collection, genotyping, imputation and quality control*

DNA from whole blood was extracted following standard protocols (Hofman et al., 2013). The Illumina Infinium II HumanHap550 BeadChips were used to genotype the RS-I (n = 6,291) and RS-II (n = 2,157) cohorts while Illumina Human610-Quad BeadChips were used to genotype the RS-III cohort (n = 3,048). Quality control criteria included the removal of SNPs with Hardy-Weinberg equilibrium deviations ($p<10-6$), genotyping call rate < 97%, gender mismatch and a high mean autosomal heterozygosity. We also excluded duplicates or first-degree relatives using identity-by-descent (IBD) estimates and outliers (three standard deviations away from the population mean) using multi-dimensional scaling (MDS) analysis with four principal components. After excluding related participants, 9,810 participants remained in our dataset.

To increase the coverage of the genome, we imputed the RS-I, RS-II and RS-III cohorts separately, using 1000Genomes (GIANT Phase I version 3) as the reference panel and using a two-step procedure imputation algorithm implemented in the program MACH-Minimac v1.0.18 using default parameters (Howie et al., 2012). In total 30,072,738 markers were genotyped and/or imputed. We filtered out markers with a MAF < 3% and an imputation quality score ($r^2$_pihat) < 0.3. This resulted in 7,260,691 markers that passed quality control and were used in the meta-analysis.

*Selection of SNPs and candidate loci associated with BCC*

We selected English publications indexed in PubMed until December 2013 that reported associations between common SNP variants (MAF > 1%) and BCC or keratinocyte carcinoma. To reduce the burden of multiple testing for our CGA we limited our CGA to variants identified in high-powered GWASs and CGAs of BCC in humans. For a GWAS we considered SNPs with a p-value < $10^{-6}$ to choose candidate genes.

To select the candidate gene/locus of an associated SNP located within a gene/locus, we retrieved the coordinates of the RefSeq longest transcript from the UCSC Genome Browser (http://www.http://genome.ucsc.edu/; GRCh37/hg19 assembly) and added 15Kb downstream and upstream of the locus region (Kent et al., 2002). The 30Kb region was selected to only include regulatory regions nearby the locus to which the variants have been previously annotated. For intergenic SNPs located more than 15Kb from a gene, we added 30Kb (15Kb downstream and 15Kb upstream) to the position of the candidate SNP. Further, we used the genomic coordinates of the selected loci and extracted all SNPs available from the RS genome-wide SNP data for the gene/locus association analyses. The list of SNPs and loci associated with the SNPs from the studies we included for our analysis, are shown in Supplementary Table S2 online.

*Statistics and analyses*

*Candidate SNP and gene association analyses on (m)BCC*

To test for associations between (m)BCC and common SNPs and candidate genes, best-guessed genotypes of the three RS cohorts were estimated from the imputed data using the Genome wide Complex Trait Analysis (GCTA) software with default parameters (Yang et al., 2011).

We first carried-out a SNP-based association analysis on SNPs that were previously associated with BCC using two case-control designs, namely: all prevalent cases of BCC against no BCC (BCC/controls), and all prevalent cases with mBCC against all prevalent cases with sBCC. We used the first design as a validation dataset and the second design to address the research questions. For the SNP association we carried-out a logistic regression analysis, adjusting the model for age at study entry/age at first BCC, sex and four principal components. Since the SNPs were already replicated in previous GWAS we only adjusted for multiple testing in the mBCC/sBCC design using the Bonferroni correction. These analyses were carried-out in PLINK v1.07 (Purcell et al., 2007).

To screen for additional variants in the genomic region (15Kb up- and downstream) to which the candidate SNPs are mapped to, a gene-based logistic regression analysis using a set-based test implemented in PLINK v1.07 was performed with the following parameters: $r^2 = 0.5$, p-value = 0.05, maximum number of SNPs = 15 and permutations = 1000. This test calculates whether the mean statistics of SNP associations within a gene are larger than those calculated under the null hypothesis of no association, taking into account the linkage disequilibrium (LD) between SNPs within a set/gene and adjusting for multiple testing per gene using permutations. For these analyses we adjusted for the same covariates as in the SNP-based associations. P-values were further adjusted for the number of evaluated genes in the mBCC/sBCC dataset.

*Genome-wide association analysis*

To discover to our knowledge previously unreported loci associated with an increased risk for mBCC, a pilot GWAS per RS cohort with participants with mBCC as cases and subjects with sBCC as controls was performed. We used logistic regression with an additive model to test for associations between SNPs and the phenotype, adjusting for age at diagnosis, sex and four principal components. The significance of the association was tested using the likelihood ratio test with one degree of freedom. The GWAS analyses were implemented in the ProbABEL package, which incorporates dose imputed data within a logistic regression framework (Aulchenko et al., 2010). Subsequently, because the standard error of the effects may not be accurate due to the small sample size, we used the p-values from the likelihood-ratio tests as summary statistics to meta-analyze.

The meta-analysis of the three RS cohorts was carried out using the METAL software, allowing for genomic control correction and heterogeneity (Willer et al., 2010).

**Supplementary table S1.** Study population characteristics

| Characteristics | sBCC[1] | mBCC[2] |
|---|---|---|
| Number of participants | 747 | 472 |
| Mean age (range) at study entry, years | 66.4 (46-99) | 65.8 (48-92) |
| Mean age (range) at first BCC diagnosis, years | **72.4 (36-99)** | **68.7 (30-96)** |
| Sex (%) | | |
| Female | 429 (57.4) | 233 (49.4) |
| Male | **318 (42.6)** | **239 (50.6)** |
| Other cutaneous cancers (%) | | |
| SCC[3] | 62 (8.3) | 76 (16.1) |
| Melanoma | 6 (0.8) | 13 (2.8) |
| SCC & Melanoma | 2 (0.3) | 4 (0.8) |
| All combined | **70 (9.4)** | **93 (19.7)** |

[1] Single basal cell carcinoma
[2] Multiple basal cell carcinoma
[3] Squamous cell carcinoma
Numbers in bold display significant differences (p-value < 0.05) according to the independent-samples t-test or Chi-squared test

**Supplementary table S2.** Selected candidate SNPs and loci

| Study type | SNP[1] | Position[2] | Chr[3] | Mutation | Locus | Genomic coordinates[4] | Other genes within region |
|---|---|---|---|---|---|---|---|
| CGA[5] | rs1126809[7] | 89017961 | 11 | missense | TYR | 88896039- 89043927 | |
| CGA | rs4911414[7] | 32729444 | 20 | intergenic | 20q11.22 | 32714444-32879444 | ASIP, AHCY (partial) |
| CGA | rs1015362[7] | 32738612 | 20 | intergenic | 20q11.22 | 32714444-32879444 | ASIP, AHCY (partial) |
| GWAS[6] | rs7538876[8] | 17722363 | 1 | intronic | PADI6 | 17683690-17743195 | RCC2 (partial) |
| GWAS | rs801114[8] | 228997835 | 1 | intergenic | 1q42.13 | 228982835-229147835 | |
| GWAS | rs11170164[9] | 52913668 | 12 | missense | KRT5 | 52893358-52929243 | |
| GWAS | rs2151280[9] | 22034719 | 9 | intronic | CDKN2B-AS1 | 21979789-22136093 | CDKN2B, CDKN2A (partial) |
| GWAS | rs157935[9] | 130585553 | 7 | intronic | LINC-PINT | 130550750-130809675 | |
| GWAS | rs16891982[9] | 33951693 | 5 | 3'UTR | SLC45A2 | 33930971-33999780 | AMACR (partial), RXFP3 (partial) |
| GWAS | rs401681[9] | 1322087 | 5 | intronic | CLPTM1L | 1302999-1360002 | |
| GWAS | rs12210050[9] | 475489 | 6 | intergenic | EXOC2 | 470137-708141 | HUS1B |
| GWAS | rs7335046[10] | 100041738 | 13 | intergenic | UBAC2 | 99837678-100053753 | GPR18, GPR183, UBAC2-AS1, FKSG29 |
| GWAS | rs1805007[10] | 89986117 | 16 | missense | MC1R | 89969286-90002385 | TUBB3 (partial) |
| GWAS | rs78378222[11] | 7571752 | 17 | 3'UTR | TP53 | 7556719- 7605868 | WRAP53 (partial) |
| GWAS | rs12203592[12] | 396321 | 6 | intronic | IRF4 | 376738- 426443 | |
| GWAS | rs12202284[12] | 471136 | 6 | intergenic | EXOC2 | 470137-708141 | HUS1B |
| GWAS | rs8015138[12] | 52310104 | 14 | intergenic | 14q22.1 | 52295104-52325104 | |
| GWAS | rs214782[13] | 2281970 | 20 | intronic | TGM3 | 2261612-2336725 | |
| GWAS | rs7006527[13] | 101024505 | 8 | intronic | RGS22 | 100958165-101133344 | |

[1] Single nucleotide polymorphism
[2] Base pair positions were based on GRCh37/hg19 assembly
[3] Chromosome
[4] Genomic coordinates were based on GRCh37/hg19 assembly
[5] Candidate gene approach
[6] Genome-wide association study
[7] Gudbjartsson *et al.* (2008); [8] Stacey *et al.* (2008); [9] Stacey *et al.* (2009); [10] Nan *et al.* (2011); [11] Stacey *et al.* (2011); [12] Zhang *et al.* (2013); [13] Stacey *et al.* (2014)

**Supplementary table S3a.** SNP-based and gene-based associations for BCC[1]

| Set[2] | Total no. SNPs/set | No. significant SNPs[3] | No. significant independent SNPs[4] | P-value[5] | SNPs/insertions/deletions[6] |
|---|---|---|---|---|---|
| TYR | 552 | 110 | 6 | 0.14 | rs28521275\|rs12363772\|rs147546939\|rs140959774\|11:88897675:D\|rs150457098 |
| 20q11.22 | 378 | 5 | 4 | 0.96 | rs819135\|rs11907546\|rs6087563\|20:32783615:I |
| PADI6 | 339 | 89 | 7 | 0.075 | rs2800696\|rs144412112\|rs12127366\|rs72637458\|rs3094884\|rs10788668\|rs76455718 |
| 1q42.13 | 816 | 60 | 15 | **0.049** | rs10916373\|rs116307216\|rs61824935\|rs142159847\|rs114257119\|1:229010041:D\|rs710805\|rs6680713\|rs710813\|rs6685751\|rs11582420\|rs78351254\|1:229101395:D\|rs190427483\|rs188101343 |
| KRT5 | 230 | 1 | 1 | **0.0020** | rs142879390 |
| CDKN2B-AS1 | 574 | 94 | 15 | 0.13 | rs72652409\|rs10757265\|9:21980969:D\|9:22062751:I\|rs597816\|rs72655437\|rs10757281\|rs74992648\|rs141348408\|9:22036252:D\|rs3731212\|rs72655422\|rs72654280\|rs10965228\|rs2811711 |
| LINC-PINT | 927 | 46 | 15 | 0.10 | rs157934\|rs17737947\|7:130567125:I\|7:130670612:D\|rs2895195\|rs73159889\|rs205755\|rs28564708\|7:130675843:D\|rs56219258\|rs150200398\|rs113171594\|rs7781295\|rs55968167\|rs34211697 |
| SLC45A2 | 191 | 24 | 4 | **0.038** | 5:33956141:D\|rs35406\|5:33944592:D\|rs173662 |
| CLPTM1L | 236 | 94 | 15 | **0.022** | rs55901723\|rs37002\|rs37004\|rs428499\|rs190785038\|rs182017427\|rs76879431\|rs27066\|rs182898174\|rs186023279\|rs144439878\|rs140648021\|rs62329688\|rs186156459\|rs71575565 |
| EXOC2 | 1310 | 74 | 15 | 0.35 | rs111739182\|rs150161396\|rs186862400\|rs151237719\|rs115663764\|rs9504290\|rs1213054\|rs1747586\|rs139546405\|rs181005467\|rs115686195\|rs181391484\|6:505139:I\|rs12210050\|rs2985278 |
| UBAC2 | 967 | 19 | 13 | 0.35 | rs192022243\|rs138435721\|rs188034987\|rs116740777\|rs142525959\|rs61970285\|13:99959963:I\|rs145902748\|rs148859462\|rs1160294\|rs138258516\|rs17472050\|rs139945409 |
| MC1R | 180 | 6 | 5 | 0.60 | rs76302987\|rs3212379\|16:89974729:D\|rs3212349\|rs140585935 |
| TP53 | 232 | 27 | 12 | 0.17 | rs1641548\|rs2908807\|rs9894946\|rs76032029\|17:7600126:D\|rs17883670\|17:7579643:D\|rs12938947\|rs78378222\|rs74351250\|rs1642785\|rs72829452 |
| IRF4 | 227 | 8 | 4 | 0.063 | rs12203592\|rs9405192\|rs190846031\|6:384956:I |
| 14q22.1 | 256 | 4 | 3 | 0.088 | rs2748146\|rs2884137\|rs2993998 |
| TGM3 | 284 | 131 | 13 | **0.0070** | rs214748\|rs188612684\|rs151233784\|rs192816818\|rs192147751\|rs6048207\|rs2014017\|20:2336433:I\|rs6082627\|rs214816\|rs214827\|rs6082867\|rs6515233 |

**Supplementary table S3a.** SNP-based and gene-based associations for BCC[1] (*continued*)

| Set[2] | Total no. SNPs/set | No. significant SNPs[3] | No. significant independent SNPs[4] | P-value[5] | SNPs/insertions/deletions[6] |
|---|---|---|---|---|---|
| RGS22 | 618 | 52 | 15 | 0.17 | 8:101111131:D\|rs17391663\|rs13340637\|rs76209450\|rs6994747\|rs187405154\|rs2885 5477\|8:101081191:I\|rs140377173\|rs28457202\|rs146325009\|rs141450089\|rs1153461-50\|rs181290644\|rs150640895 |

[1] BCC: 1,219 cases (≥1 BCC) were compared to 8,591 controls (no BCC)
[2] Locus region, see Supplementary Table S2 for exact genomic coordinates
[3] Total number of SNPs with p-value < 0.05
[4] Total number of significant SNPs (p-value < 0.05) also passing LD-criterion ($r^2 < 0.5$)
[5] Numbers in bold display significant associations
[6] List of significant SNPs/deletions/insertions also passing the LD criterion

**Supplementary table S3b.** SNP-based and gene-based associations for mBCC[1]

| Set[2] | Total no. SNPs/set | No. significant SNPs[3] | No. significant independent SNPs[4] | P-value[5] | SNPs/insertions/deletions[6] |
|---|---|---|---|---|---|
| TYR | 552 | 9 | 1 | 0.16 | rs35963892 |
| 20q11.22 | 378 | 4 | 4 | 0.77 | rs146491521\|rs139914983\|rs13042965\|rs7262908 |
| PADI6 | 339 | 62 | 5 | 0.11 | rs183603588\|rs3094884\|rs7191\|rs76550244\|rs7556072 |
| 1q42.13 | 816 | 4 | 4 | 0.75 | rs146666914\|rs115475209\|rs77541106\|rs187651720 |
| KRT5 | 230 | 52 | 3 | 0.24 | rs11170161\|12:52898781:D\|rs11549951 |
| CDKN2B-AS1 | 574 | 1 | 1 | 0.64 | rs61743293 |
| LINC-PINT | 927 | 31 | 14 | 0.86 | rs117798053\|rs205730\|rs11767836\|rs6959008\|rs11760691\|rs181060962\|rs11980581\|rs116519273\|7:130755763:D\|7:130657365:I\|rs62473520\|7:130652737:D\|rs6953084\|rs17165262 |
| SLC45A2 | 191 | 3 | 2 | 0.47 | rs77175283\|rs141549575 |
| CLPTM1L | 236 | 7 | 3 | 0.83 | rs12332579\|rs189368366\|rs144439878 |
| EXOC2 | 1310 | 36 | 13 | 0.13 | rs145280257\|rs58022076\|rs17756753\|rs6933777\|rs113608863\|rs72835985\|rs3823124\|rs186862400\|6:675542:D\|rs3799307\|rs143388344\|rs150891820\|rs17236435 |
| UBAC2 | 967 | 148 | 9 | 0.016 | rs2390237\|rs2296046\|13:100022283:D\|13:99934484:I\|13:99933087:D\|rs79145601\|rs9513608\|rs61968345\|rs188179841 |
| MC1R | 180 | 32 | 8 | 0.25 | rs72813442\|rs76302987\|rs12920483\|rs4785736\|rs111398992\|rs188455950\|rs182927424\|rs186542844 |
| TP53 | 232 | 2 | 2 | 0.39 | rs17883670\|MERGED_DEL_2_87391 |
| IRF4 | 227 | 10 | 8 | 0.30 | rs79324228\|rs11756234\|6:425200:D\|rs9378774\|rs3823305\|rs6935510\|6:393837:D\|rs72833949 |
| 14q22.1 | 256 | 1 | 1 | 0.72 | rs140107698 |
| TGM3 | 284 | 26 | 8 | 0.19 | rs6515233\|rs6082627\|rs112763701\|rs2422689\|rs2076408\|rs214755\|rs45440896\|rs214802 |
| RGS22 | 618 | 5 | 5 | 0.85 | rs191382884\|rs62532694\|8:101065714:D\|rs140899826\|8:100989341:I |

[1] mBCC: 472 cases (>1 BCC) were compared to 747 controls (1 BCC)

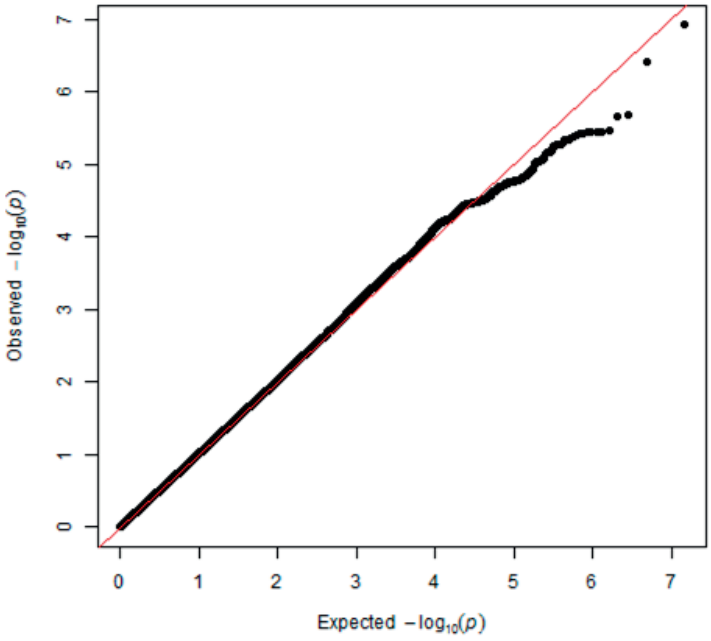
[2] Locus region, see Supplementary Table S2 for exact genomic coordinates

[3] Total number of SNPs with p-value < 0.05

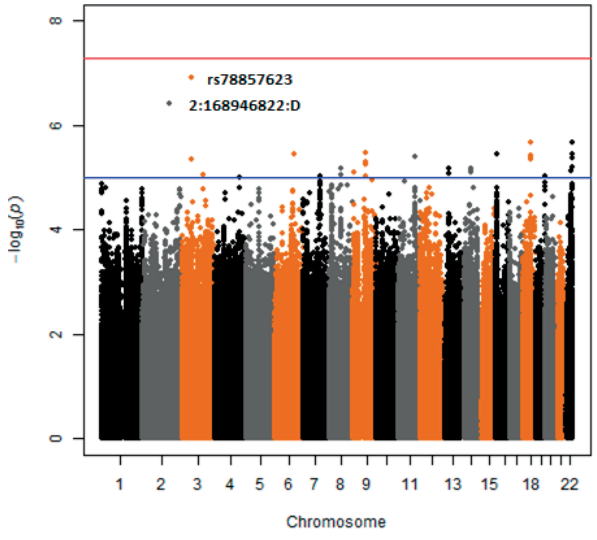
[4] Total number of significant SNPs (p-value < 0.05) also passing LD-criterion ($r^2 < 0.5$)

[5] There are no significant associations, because all p-values need to be corrected for multiple testing

[6] List of significant SNPs/deletions/insertions also passing the LD criterion




**Supplementary Figure S1.** Q-Q-plot of the meta-analysis of the Rotterdam Study on mBCC
The $-\log_{10}$ of observed p-values of the associations between mBCC and SNPs (Y-axis) are plotted against the expected p-values under the assumption of no association (X-axis).




**Supplementary Figure S2.** Manhattan plot of the meta-analysis of the Rotterdam Study on mBCC
The $-\log_{10}$ of observed p-values of the associations between mBCC and SNPs (Y-axis) for all SNPs (dots) are represented per chromosome (X-axis). The horizontal lines indicate the significant threshold of p-value = 5×10-6 (blue) and p-value = $5\times10^{-8}$ (red).

## References

Aulchenko YS, Struchalin MV, van Duijn CM (2010) ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics* 11:134.

Casparie M, Tiebosch AT, Burger G, *et al.* (2007) Pathology databanking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. *Cell Oncol* 29:19-24.

Gudbjartsson DF, Sulem P, Stacey SN, *et al.* (2008) ASIP and TYR pigmentation variants associate with cutaneous melanoma and basal cell carcinoma. *Nat Genet* 40:886-91.

Hofman A, Darwish Murad S, van Duijn CM, *et al.* (2013) The Rotterdam Study: 2014 objectives and design update. *Eur J Epidemiol* 28:889-926.

Howie B, Fuchsberger C, Stephens M, *et al.* (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* 44:955-9.

Kent WJ, Sugnet CW, Furey TS, *et al.* (2002) The human genome browser at UCSC. *Genome Res* 12:996-1006.

Nan H, Xu M, Kraft P, *et al.* (2011) Genome-wide association study identifies novel alleles associated with risk of cutaneous basal cell carcinoma and squamous cell carcinoma. *Hum Mol Genet* 20:3718-24.

Purcell S, Neale B, Todd-Brown K, *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559-75.

Stacey SN, Gudbjartsson DF, Sulem P, *et al.* (2008) Common variants on 1p36 and 1q42 areassociated with cutaneous basal cell carcinoma but not with melanoma or pigmentation traits. *Nat Genet* 40:1313-8.

Stacey SN, Sulem P, Gudbjartsson DF, *et al.* (2014) Germline sequence variants in TGM3 and RGS22 confer risk of basal cell carcinoma. *Hum Mol Genet* 23:3045-53.

Stacey SN, Sulem P, Jonasdottir A, *et al.* (2011) A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. *Nat Genet* 43:1098-103.

Stacey SN, Sulem P, Masson G, *et al.* (2009) New common variants affecting susceptibility to basal cell carcinoma. *Nat Genet* 41:909-14.

Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26:2190-1.

Yang J, Lee SH, Goddard ME, *et al.* (2011) GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 88:76-82.

Zhang M, Song F, Liang L, *et al.* (2013) Genome-wide association studies identify several new loci associated with pigmentation traits and skin cancer risk in European Americans. *Hum Mol Genet* 22:2948-59.