



**HAL**  
open science

## Genome-wide pooling approach identifies *SPATA5* as a new susceptibility locus for alopecia areata

Regina C. Betz, Lina M. Forstbauer, Felix F. Brockschmidt, Valentina Moskvina, Christine Herold, Silke Redler, Alexandra Herzog, Axel M. Hillmer, Christian Meesters, Stefanie Heilmann, et al.

### ► To cite this version:

Regina C. Betz, Lina M. Forstbauer, Felix F. Brockschmidt, Valentina Moskvina, Christine Herold, et al.. Genome-wide pooling approach identifies *SPATA5* as a new susceptibility locus for alopecia areata. *European Journal of Human Genetics*, 2011, 10.1038/ejhg.2011.185 . hal-00691329

**HAL Id: hal-00691329**

**<https://hal.science/hal-00691329>**

Submitted on 26 Apr 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Genome-wide pooling approach identifies *SPATA5* as a new susceptibility locus for**  
2 **alopecia areata**

3  
4 Lina M. Forstbauer<sup>1\*</sup>, Felix F. Brockschmidt<sup>1,2\*</sup>, Valentina Moskvina<sup>3</sup>, Christine  
5 Herold<sup>4</sup>, Silke Redler<sup>1</sup>, Alexandra Herzog<sup>1</sup>, Axel M. Hillmer<sup>5</sup>, Christian Meesters<sup>4,6</sup>,  
6 Stefanie Heilmann<sup>1,2</sup>, Florian Albert<sup>1</sup>, Margrieta Alblas<sup>1,2</sup>, Sandra Hanneken<sup>7</sup>, Sibylle  
7 Eigelshoven<sup>7</sup>, Kathrin A. Giehl<sup>8</sup>, Dagny Jagielska<sup>1,9</sup>, Ulrike Blume-Peytavi<sup>9</sup>, Natalie  
8 Garcia Bartels<sup>9</sup>, Jennifer Kuhn<sup>10,11,12</sup>, Hans Christian Hennies<sup>10,11,12</sup>, Matthias  
9 Goebeler<sup>13</sup>, Andreas Jung<sup>13</sup>, Wiebke K. Peitsch<sup>14</sup>, Anne-Katrin Kortüm<sup>15</sup>, Ingrid Moll<sup>15</sup>,  
10 Roland Kruse<sup>16</sup>, Gerhard Lutz<sup>17</sup>, Hans Wolf<sup>7</sup>, Bettina Blaumeiser<sup>18</sup>, Markus Böhm<sup>19</sup>,  
11 George Kirov<sup>3</sup>, Tim Becker<sup>4,6</sup>, Markus M. Nöthen<sup>1,2</sup>, Regina C. Betz<sup>1</sup>

12  
13  
14 <sup>1</sup> Institute of Human Genetics, University of Bonn, Bonn, Germany

15 <sup>2</sup> Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany

16 <sup>3</sup> MRC Centre for Neuropsychiatric Genetics & Genomics, Henry Wellcome Building,  
17 Cardiff University, Heath Park, Cardiff, United Kingdom

18 <sup>4</sup> German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany

19 <sup>5</sup> Genome Technology and Biology Group, Genome Institute of Singapore, Singapore

20 <sup>6</sup> Institute for Medical Biometry, Informatics, and Epidemiology, University of Bonn,  
21 Bonn, Germany

22 <sup>7</sup> Department of Dermatology, University of Düsseldorf, Düsseldorf, Germany

23 <sup>8</sup> Department of Dermatology, University of Munich, Munich, Germany

24 <sup>9</sup> Clinical Research Center for Hair and Skin Science, Department for Dermatology  
25 and Allergy, Charité-Universitätsmedizin Berlin, Berlin, Germany

26 <sup>10</sup> Division of Dermatogenetics, Cologne Center for Genomics, University of Cologne,  
27 Cologne, Germany

28 <sup>11</sup> Center for Molecular Medicine Cologne, University of Cologne, Cologne, Germany

29 <sup>12</sup> Cologne Cluster of Excellence on Cellular Stress Responses in Aging-associated  
30 Diseases, University of Cologne, Cologne, Germany

31 <sup>13</sup> Department of Dermatology, University of Giessen, Giessen, Germany

32 <sup>14</sup> Department of Dermatology, University Medical Centre Mannheim, University of  
33 Heidelberg, Mannheim, Germany

34 <sup>15</sup> Department of Dermatology and Venerology, University Hospital Hamburg-  
35 Eppendorf, Hamburg, Germany

36 <sup>16</sup> Dermatological Practice, Paderborn, Germany

37 <sup>17</sup> Hair and Nail, Wesseling, Germany

38 <sup>18</sup> Department of Medical Genetics, University & University Hospital of Antwerp,  
39 Antwerp, Belgium

40 <sup>19</sup> Department of Dermatology, University of Münster, Münster, Germany

41

42

43 Correspondence to:

44 Prof. Regina C. Betz, MD

45 Institute of Human Genetics

46 University of Bonn

47 Sigmund-Freud-Str. 25

48 D-53127 Bonn

49 Germany

50 Phone: +49 228 287 51023

51 Fax: +49 228 287 51011

52 E-mail: [regina.betz@uni-bonn.de](mailto:regina.betz@uni-bonn.de)

53

54

55 \* These authors have contributed equally to this work.

56

57

58 **Running title:** SPATA5 identified in GWAS for alopecia areata

59

60 **Abstract**

61 Alopecia areata (AA) is a common hair loss disorder which is thought to be a tissue-  
62 specific autoimmune disease. Previous research has identified a few AA susceptibility  
63 genes, most of which are implicated in autoimmunity. To identify new genetic variants  
64 and further elucidate the genetic basis of AA, we performed a genome-wide association  
65 study using the strategy of pooled DNA genotyping (729 cases; 656 controls). The  
66 strongest association was for variants in the HLA region, which confirms the validity of  
67 the pooling strategy. The selected top 61 single nucleotide polymorphisms (SNPs) were  
68 analyzed in an independent replication sample (454 cases; 1,364 controls). Only one  
69 SNP outside of the HLA region (rs304650) showed significant association. This SNP  
70 was then analyzed in a second independent replication sample (537 cases; 657 controls).  
71 The finding was not replicated on a significant level, but showed the same tendency. A  
72 combined analysis of the two replication samples was then performed, and the SNP  
73 rs304650 showed significant association with  $P = 3.43 \times 10^{-4}$  (OR = 1.24 [1.10 - 1.39]).  
74 This SNP maps to an intronic region of the *SPATA5* (spermatogenesis-associated  
75 protein 5) gene on chromosome 4. The results therefore suggest the *SPATA5* locus is a  
76 new susceptibility locus for AA.

77

78

79 **Keywords**

80 Alopecia areata, genome-wide association study, autoimmunity, hair loss

81

82 **Introduction**

83 Alopecia areata (AA) is a common hair loss disorder which affects approximately 1-2 %  
84 of the general population. It affects both sexes and all age groups.<sup>1</sup> Affected individuals  
85 present with a non-scarring, circumscribed hair loss which has a sudden onset and a  
86 recurrent course. The scalp is the most commonly affected site, although all hair-bearing  
87 areas of the skin may be involved. Episodes of hair loss typically start with isolated  
88 hairless patches. These extend centrifugally and may coalesce. AA is divided into three  
89 main clinical types on the basis of the degree of hair loss and the sites affected: (i)  
90 patchy AA; (ii) AA totalis (AT), affecting the whole scalp; and (iii) AA universalis  
91 (AU), affecting the whole body. Individuals with AA may experience complete  
92 remission, a chronic course, or progression towards AT or AU. Familial AA also  
93 occurs, and recurrence risks of 5–6% have been reported for the children of affected  
94 individuals.<sup>2,3</sup> The pattern of familiarity and a limited twin study which reports a  
95 concordance rate in monozygotic twins of 55%<sup>3</sup> suggest that multiple genetic factors  
96 and environmental factors are involved in the pathology of AA. Nevertheless, the  
97 etiopathogenesis of AA remains poorly understood. One hypothesis is that AA is a  
98 tissue-specific autoimmune disease of the hair follicle.<sup>4</sup> This is supported by reports of  
99 association between AA and specific HLA alleles.<sup>5-8</sup> Association has also been reported  
100 between AA and the *W620* variant of the *PTPN22* gene (protein tyrosine phosphatase,  
101 nonreceptor type 22), which has been implicated in several autoimmune disorders.<sup>9,10</sup>  
102 To date, only one systematic genome-wide linkage study of AA has been performed in  
103 humans. This identified four susceptibility loci on chromosomes 6, 10, 16, and 18,  
104 respectively.<sup>11</sup> The pattern of familiarity and a limited twin study which reports a concordance  
105 rate in monozygotic twins of 55%

106 To identify new genetic variants and thus further elucidate the genetic basis of AA, we  
107 performed a genome-wide association (GWA) study in 729 AA cases and 656 controls  
108 using the strategy of pooled DNA genotyping. The top 61 single nucleotide  
109 polymorphisms (SNPs) were selected for individual genotyping in the initial sample of  
110 the pooling approach for the purpose of confirmation. We further analyzed these SNPs  
111 in an independent replication sample. The best association finding was then analyzed in  
112 a second independent replication sample. The strongest association was found for  
113 variants in the *HLA* region. These results thus confirm the validity of the pooling-based  
114 approach to DNA genotyping. Borderline significance was found for rs304650 in the  
115 *SPATA5* (spermatogenesis-associated protein 5) gene, which may therefore represent a  
116 new susceptibility gene locus for AA.

117 During the final stages of the present study, a GWA study of an AA sample from the  
118 US was published by Petukhova et al.<sup>12</sup> This identified eight gene loci for AA, most of  
119 which are implicated in autoimmunity. Although *SPATA5* was not among these loci, six  
120 SNPs in *SPATA5*, of all which show evidence for linkage disequilibrium with rs304650,  
121 were reported with a *P*-value < 10<sup>-4</sup>.

122

### 123 **Materials and Methods**

124 *DNA samples* The analyses involved 1 720 individuals with AA (cases) and 2 677  
125 controls. For cases, the inclusion criterion was a dermatologist-assigned diagnosis of  
126 AA. Patients with Down syndrome or Turner syndrome were excluded. The cases were  
127 recruited from: (i) outpatient clinics in Belgium (University Hospitals of Antwerp and  
128 Gent) and Germany (University Hospitals of Munich, Münster, Düsseldorf, Berlin,  
129 Bonn, Gießen, Hamburg and Mannheim); (ii) a private dermatology practice

130 (Wesseling, Germany); and (iii) AA self-help support groups (Germany, Netherlands).  
131 All cases were asked whether or not they had a positive family history of AA. This was  
132 defined as a history of at least one first- or second-degree relative with any form of the  
133 disorder. The controls were drawn from the general population, and were therefore not  
134 specifically screened for the absence of AA. They were healthy unrelated blood donors  
135 from the University Hospital of Bonn (n = 1 313) or participants in one of three  
136 population-based epidemiological studies: (1) PopGen<sup>13</sup> (n = 490); (2) KORA<sup>14</sup> (n =  
137 490); and (3) HNR<sup>15</sup> (n = 384). All cases and controls were of central European origin.  
138 Ethical approval was obtained from the respective ethics committees, and all  
139 participants provided written informed consent prior to blood sampling. The DNA of  
140 patients and blood donors was extracted from peripheral blood leukocytes. This was  
141 achieved by salting out with saturated NaCl solution according to standard methods, or  
142 through the use of a Chemagic Magnetic Separation Module I (Chemagen, Baesweiler,  
143 Germany) in accordance with the manufacturer's instructions. The DNA was stored in  
144 liquid nitrogen until use. For the controls from the three population-based  
145 epidemiological studies, the genotypes were taken from previously generated genome-  
146 wide association data sets. The PopGen and KORA datasets were generated within the  
147 German National Genome Research Network to serve as a national research resource.  
148 The HNR data set was generated as part of a collaboration to generate a set of universal  
149 controls for genetic studies.

150 *DNA Pooling* The DNA samples of 729 AA cases and 656 blood-donor controls were  
151 selected at random. These samples were then used to generate a control pool and two  
152 AA case pools. The first AA case pool consisted of the DNA of all 729 AA cases  
153 (including the 224 AA cases with a positive family history). The second AA case pool

154 consisted of the DNA of ‘positive family history’ AA patients only (n = 224). To ensure  
155 equal amounts of DNA in each pool, each individual DNA sample was subjected to a  
156 series of dilution steps (above 100ng/μl to 20ng/μl, and so on to a final concentration of  
157 10ng/μl). Each DNA sample was double-quantified using a NanoDrop device  
158 (PEQLAB Biotechnologie GmbH, Erlangen, Germany) and adjusted to +/-10% of the  
159 required concentration. For the 10ng/μl dilution, a broader range of -60% to +80% was  
160 used to reflect the measurement error of the NanoDrop device at low DNA  
161 concentrations. Equimolar amounts of each DNA sample (100ng) were pipetted  
162 manually into a large tube to form the pool. Each pool was then concentrated to 60ng/μl  
163 using the Microcon YM-100 Centrifugal Filter Device (Millipore GmbH, Schwalbach,  
164 Germany), and then requantified with the NanoDrop device.

165 *Genotyping of the pooled DNA and data analysis* The Illumina Sentrix  
166 HumanHap550v3 genotyping BeadChips were used. These contain more than 550 000  
167 tag SNPs (Illumina Inc., San Diego, CA). To avoid inter-experimental variation, each  
168 pool was genotyped on five chips in accordance with the manufacturer’s  
169 recommendations for individual samples. SNP allele frequencies were estimated using  
170 data from the BeadArray™ Reader imaging (Illumina Inc., San Diego, CA) and  
171 Illumina’s genotyping software Bead Studio 2.0 (Illumina Inc., San Diego, CA). Three  
172 analyses were performed on the basis of the allele frequency estimates obtained from  
173 the genotyping experiment: (i) 729 cases vs. 656 controls; (ii) 224 cases with a positive  
174 family history vs. 656 controls; and (iii) a sliding window analysis using the dataset of  
175 729 cases vs. 656 controls, in accordance with recommendations described elsewhere.<sup>16</sup>  
176 For each replicate, the following approximation of allele frequencies in cases and  
177 controls was performed on the basis of the raw data:  $f_i = X_{raw}/(X_{raw}+Y_{raw})$ , where



178 Xraw and Yraw are the intensities of the two dyes (Cy5 and Cy3) used to genotype  
179 SNPs on the Illumina platform, and  $i$  is the replicate number. The obtained allele  
180 frequency was then averaged over the number of replicates of each pool  $f=(f_1+\dots+f_M)/M$   
181 (where  $M$  is the number of replicates) and corrected for unequal amplification using the  
182 formula  $f_{corrected}=f/(f+k\cdot(1-f))$ .<sup>17</sup> For each SNP, the correction factor  $k$  was estimated  
183 using HapMap CEPH data (<http://hapmap.ncbi.nlm.nih.gov/>) and the formula  $k=(f_{con}$ -  
184  $f_{con}:f_{CEPH})/(f_{CEPH}-f_{con}:f_{CEPH})$ , where  $f_{cont}$  is the allele frequency in controls obtained from  
185 the pooling experiment and  $f_{CEPH}$  is the allele frequency calculated from HapMap CEPH  
186 data.

187 The correction factor  $k$  was used as the quality control measure as it provides an  
188 indication of how close the estimated allele frequencies in controls are to those of the  
189 CEPH data. SNPs were excluded from the further analysis when  $k$  was  $>4$  or  $<0.25$ .

190 The coefficient of variation was also used as a quality control measure for each SNP in  
191 each pool. This was calculated using the formula  $\text{sqrt}(\text{var}_e)/f$ , where  $\text{var}_e$  is the  
192 experimental variance estimated from the replicates data. This indicates how close the  
193 SNP is to the replicates, while taking into account allele frequencies and the number of  
194 replicates per pool. SNPs were excluded from further analysis when the coefficient of  
195 variation was  $>1$  (upper boundary of the 95% confidence interval) in at least in one of  
196 the three pools (i.e. cases, controls, familial cases).

197 For the remaining 487 932 SNPs, association analyses were performed to compare 729  
198 cases and 224 familial cases with 656 controls. Modified Z-statistics were used, as  
199 described by Abraham et al.<sup>18</sup> These take into account both the experimental variance  
200 (see above) and the sampling variance.

201 A signal processing analysis technique was used to interpret the results of the case-  
202 control association analysis. This enables the detection of genetic association with  
203 disease, while taking into account the significance of several subsequent genetic  
204 markers in a sliding window. Moskvina has described a theoretical approach to  
205 calculating the probability of at least one false alarm being flagged by the detection  
206 statistic under the Null hypothesis of no signal (association).<sup>14</sup> This is equivalent to the  
207 probability of type I error when taking the number of comparisons in the window into  
208 account, and thus provides a “genome-wide window-based” significance level.<sup>19</sup>

209 *SNP selection, confirmation of pooling results, and independent replication* A number  
210 of SNPs were selected to confirm the pooling results on an individual genotyping level.  
211 A replication step was performed in two independent case-control samples (for an  
212 overview see Fig. 1). The following SNPs were selected from the three pooling-based  
213 analyses: (i) the top 50 SNPs from the analysis ‘729 cases vs. 656 controls’; (ii) the top  
214 20 SNPs from the analysis ‘224 cases with a positive family history vs. 656 controls’;  
215 and (iii) the top 35 SNPs from the ‘sliding window analysis’. All selected SNPs had  
216 shown  $P$ -values of  $< 1 \times 10^{-4}$  in the respective pooling analysis. As SNPs from the  
217 previously known HLA region were highly overrepresented, all but three of these SNPs  
218 were excluded. The three remaining HLA SNPs, rs3115553 (chr. 6: 32 353 805bp),  
219 rs9275141 (chr. 6: 32 759 095bp), and rs9275572 (chr. 6: 32 786 977bp), were selected  
220 since they were the best findings in the HLA region in the ‘sliding window analysis’.  
221 These were used as positive controls in subsequent analyses. Regrettably, rs9275141  
222 failed during the assay design of the Sequenom iPLEX reaction (Sequenom GmbH,  
223 Hamburg, Germany). It was therefore replaced by the next best SNP in the HLA region  
224 from the ‘sliding window analysis’, rs9268528 (chr. 6: 32 491 086bp). By excluding

225 other SNPs from the HLA region and SNPs appearing in more than one of the analyses  
226 (doubles and triples), it was possible to reduce the SNP set for the confirmation and  
227 replication step (step 2) from 105 to 61 SNPs. Individual genotyping was performed on  
228 the Illumina platform. For the DNA samples from the HNR-, KORA- and Popgen  
229 controls (n = 1 364), HumanHap550v3 genotyping BeadChips (Illumina Inc., San Diego,  
230 CA) were used. For all other DNA samples, Sequenom's Compact MALDI-ToF Mass  
231 Array system and iPLEX Gold reagents (Sequenom GmbH, Hamburg, Germany) in  
232 multiplex reactions were used. Primer sequences and Sequenom's standard assay  
233 conditions are available upon request. All primers were checked by MALDI-ToF. For  
234 quality reasons, a success rate of at least 95% was required for all analyzed SNPs. A  
235 95% call rate was required for each of the samples used in the confirmation and  
236 replication steps. In the confirmation step (step 2), DNA samples from eight AA cases  
237 and 11 controls failed to fulfil the quality criteria and were excluded from the analysis,  
238 and thus 721 AA cases and 645 controls remained for the analyses. In the independent  
239 replication step (step 2), four DNA case-samples failed to fulfil the quality criteria, and  
240 thus 450 cases and 1 364 controls remained for the analyses. Of the 61 selected SNPs,  
241 six SNPs (rs12493901, rs30117, rs41515, rs4777450, rs7246435, and rs9520256) were  
242 technical failures, and five SNPs (rs10123149, 11098149, rs6700586, rs7099812, and  
243 rs724841) had to be excluded as they did not reach the required call rate of  $\geq 95\%$  in one  
244 or more of the confirmation and/or replication samples. The SNP rs7334982 was not  
245 biallelic in the present samples and was therefore excluded from further analysis. Thus,  
246 a set of 49 SNPs remained for further analysis.

247 *Statistical analyses of individual genotyping data* The FAMHAP software package<sup>20</sup>  
248 was used for the association and haplotype analyses. The Armitage trend test was used

249 for the single marker analyses.<sup>21</sup> All SNPs met the following quality criteria: minor  
250 allele frequency > 1%,  $P_{HWE}$  in cases > 0.001, and  $P_{HWE}$  in controls > 0.05. The  
251 resulting  $P$ -values were corrected for multiple testing according to the number of SNPs  
252 successfully analyzed on the individual genotyping level ( $n = 49$ ).

253 Expression analyses We used the forward primer 5'-CCTTCAAACCGACGCATACT-  
254 3' and the reverse primer 5'-GCAGCCCACTCTTCTCTTGA-3' to analyse the  
255 expression of *SPATA5* in human hair follicle and skin samples (expected product size:  
256 197bp). As *SPATA5* expression has been proven in human kidney and lung  
257 (<http://www.genecards.org/cgi-bin/carddisp.pl?gene=SPATA5&search=SPATA5>),  
258 these tissues were included as positive controls. Total RNA was extracted from human  
259 hair follicles and skin using the the RNeasy® Micro Kit (Qiagen), and single strand  
260 cDNA was synthesized from a total of 400ng RNA using the Super Script III First  
261 Strand Synthesis System (Invitrogen). Single strand cDNA from kidney and lung was  
262 obtained from the Human Multiple-Tissue cDNA (MTC) Panel Human MTC™ Panel I  
263 (LOT Nr. 6060248; Clontech). A negative reverse transcription reaction (no enzyme)  
264 was included as a negative control.

265

## 266 **Results**

### 267 *Step 1: Pooling-based approach*

268 The genome-wide pooling-based approach involved three DNA-pools: (1) 729 AA  
269 cases (including the 224 cases with a positive family history), (2) 224 AA cases with a  
270 positive family history, and (3) 656 controls. Each pool was genotyped on five  
271 replicates of the Illumina Sentrix® HumanHap550v3 genotyping BeadChip. Pool 1 was  
272 successfully analyzed in all five replicates. Pool 2 was successfully analyzed in four

273 replicates. Two of the five control chips were excluded following quality control  
274 filtering.

275 Quality control measures and allele frequencies were estimated for all 504 931 SNPs.  
276 The quality control measures used were: (i) the coefficient of variation of each SNP in  
277 each pool, which reflected how close the SNP was to the replicates while taking into  
278 account allele frequencies and the number of replicates per pool; and (ii) correction  
279 factor  $k$  as an indicator of the closeness of the allele frequency estimates in the control  
280 pool to the allele frequencies in the CEPH sample from HapMap (for details see the  
281 Methods section).

282 Markers were excluded if the coefficient of variation was  $>1$  in at least one of three  
283 DNA-pools, or if the correction factor  $k$  was  $>4$  or  $<0.25$ . The remaining 487 932 SNPs  
284 were corrected for  $k$ , and SNPs with minor allele frequencies of  $<5\%$  in controls were  
285 filtered out. The remaining 468 389 SNPs were tested for association using modified Z-  
286 statistics.<sup>18</sup> Separate comparisons with controls were made for cases and familial cases.  
287 The results of the case-control comparison were further analyzed in a sliding window.  
288 The best markers from the top regions identified by the sliding window analysis<sup>19</sup> were  
289 selected for replication. Thus, three different analyses were performed: (I) 729 AA cases  
290 vs. 656 controls; (II) 224 AA cases with a positive family history vs. 656 controls; and  
291 (III) a sliding window analysis (Fig. 1). These analyses identified 31 SNPs with  $P$ -  
292 values  $< 5 \times 10^{-7}$  (supplement table 1), resulting in a total of 18 SNPs after the exclusion  
293 of duplicates and triplicates. Of these 18 SNPs, eight SNPs were localized in the HLA  
294 region, and ten SNPs were localized elsewhere in the genome. The best SNP of all three  
295 analyses was rs9952976 (chr. 18: 42,561,717bp) which had a  $P$ -value of  $6.48 \times 10^{-14}$  in  
296 analyses I and III (supplement table 1). The best SNP in analysis II was rs9275572,

297 which is localized in the HLA region ( $P = 1.87 \times 10^{-8}$ ). This SNP was also the best  
298 HLA-SNP in analyses I ( $P = 1.00 \times 10^{-11}$ ) and III ( $P = 5.67 \times 10^{-12}$ ) (supplement table  
299 1). When the best SNPs from each analysis were considered (see selection criteria in  
300 Materials and Methods), three SNPs (rs9275141 (chr. 6: 32 759 095); rs9275572 (chr.  
301 6: 32 786 977); and rs9952976 (chr. 18: 42 561 717)) appeared in all three analyses.  
302 Two of these SNPs (rs9275141 and rs9275572) are localized in the HLA region  
303 (supplement table 1).

304 *Step 2: Individual confirmation and independent replication*

305 The top 50 SNPs from analysis I, the top 20 SNPs from analysis II, and the top 35 SNPs  
306 from analysis III were selected for further analysis (Fig. 1). The elimination of  
307 duplicates and triplicates resulted in a total of 61 SNPs (see Materials and Methods and  
308 supplement table 1). To confirm the association findings of the 61 selected SNPs in the  
309 pooling approach, individual genotyping was performed in the previously pooled  
310 discovery sample of 729 AA cases and 656 controls. An independent replication step  
311 involved 454 AA cases and 1 364 controls. Following quality control, 49 SNPs  
312 remained for analysis (see Materials and Methods). With the exception of five SNPs,  
313 this analysis confirmed the pooling results at a nominal level of significance, and thus  
314 demonstrated the validity of the DNA pooling approach. The strongest association was  
315 found for the three HLA-SNPs (rs3115553, rs9268528, and rs9275572). The SNP  
316 rs9275572 showed the strongest association with  $P = 2.50 \times 10^{-10}$  (OR = 1.65 [1.41 -  
317 1.94]) (Table 1). These were the only SNPs to withstand correction for multiple testing  
318 using the previously suggested threshold of  $P = 5 \times 10^{-7}$ .<sup>22</sup> The remaining SNPs failed to  
319 show strong association. The best association finding outside of the HLA region was for  
320 rs2110597 (chr. 12: 12 832 280bp) with  $P = 1.42 \times 10^{-5}$  (OR = 1.44 [1.22 - 1.68]). The

321 best finding from the pooling-based analysis, rs9952976 (chr. 18: 42 561 717 bp),  
322 showed only borderline significance on the level of individual genotyping, with  $P =$   
323 0.034 (OR = 1.20 [1.01 - 1.43]). This was one of the weakest association findings in this  
324 confirmation step (Table 1).

325 In the independent replication step, the strongest association was again found for the  
326 three HLA SNPs. The SNP rs9275572 was the most strongly associated SNP ( $P = 7.94$   
327  $\times 10^{-11}$  (OR = 1.71 [1.46 - 2.01]). Twenty-seven SNPs showed the same risk alleles as in  
328 the discovery sample. Only one SNP outside of the HLA region (rs304650; chr. 4:  
329 124,303,368bp) showed significant association ( $P$ -value of 0.001 (OR = 1.31 [1.12 -  
330 1.53])). Following Bonferroni-correction for the number of SNPs tested ( $n = 49$ ), only  
331 the three HLA SNPs and rs304650 ( $P = 0.049$ ) remained significant (Table 1).

### 332 *Step 3: Follow-up analysis of the top finding*

333 A second independent sample of 537 cases and 657 controls was then used to  
334 investigate the association finding for rs304650 further (Fig. 1). Genotyping of  
335 rs304650 failed in one case and two controls, and thus 536 cases and 655 controls  
336 remained. In this analysis, rs304650 could not be replicated at a significant level ( $P =$   
337 0.127) (Table 2). However, the risk allele remained the same. A combined analysis was  
338 therefore performed using all AA cases and controls from the independent replication  
339 and follow-up steps (a total of 985 cases and 2,014 controls were successfully  
340 genotyped for this SNP) (data not shown). Here, rs304650 showed stronger association,  
341 with  $P = 3.43 \times 10^{-4}$  (OR = 1.24 [1.10 - 1.39]). After combining all cases and controls,  
342 we obtained a  $P$ -value of  $1.58 \times 10^{-5}$ ; OR = 1.23 [1.12 - 1.35]. Interestingly, *SPATA5*  
343 expression was observed in hair follicles and skin, which confirms the importance of  
344 this gene in terms of hair biology.

345

346 **Discussion**

347 The present GWA study of AA is the first to have used pooled DNA. The analysis was  
348 performed in several stages to avoid the high costs of performing a GWA study in large  
349 individual samples. Genotyping of DNA pools was performed on 15 Illumina  
350 HumanHap550 arrays of patients and controls. A limitation of pooling studies in  
351 comparison to individual genotyping approaches is that allele frequencies are estimates  
352 deriving from DNA-pools, which are inherently imprecise. In view of this and the fact  
353 that the generally used quality control measures cannot be applied, the pooling-based  
354 approach was used as the discovery step (step 1; Fig. 1). In the second step, the top  
355 SNPs were confirmed using individual genotyping of the previously pooled case and  
356 control samples and replicated through individual genotyping in an independent sample  
357 of cases and controls (step 2; Fig 1). In the third step, the best SNP was followed up in a  
358 further independent replication sample (Fig. 1).

359 The major histocompatibility complex on chromosome 6p21.3 was identified as a major  
360 risk locus for AA. Previous research by our group and others has implicated various  
361 HLA alleles in AA susceptibility. The best replicated findings have been for alleles of  
362 the DRB1 and DQB1 loci.<sup>5,6,8,23-25</sup> The present highly significant findings for variants in  
363 the HLA region demonstrate that the pooling-based strategy is a valid alternative to  
364 individual genotyping in complex disorders.<sup>26</sup> Although the pooling-based results for  
365 the HLA locus were not followed up systematically (step 1), the best three variants from  
366 the sliding window analysis were genotyped in the discovery and independent  
367 replication samples to confirm the initial pooling-based results (step 2). As expected, all  
368 three variants were confirmed on the individual genotyping level and reached genome-



369 wide significant *P*-values. This was also the case in the independent replication step.  
370 This indicates that the DNA pooling approach can reliably detect SNPs that have shown  
371 genome-wide significance in association studies. Furthermore, pooling detects highly  
372 significant results, and it is therefore very unlikely that any genes beyond the HLA  
373 region are more significant. However, our strategy carries a risk of false-negative  
374 findings in the case of smaller genetic effects. DNA pooling adds extra experimental  
375 error (e.g. pipetting for pool construction) to the allele frequency measurement that  
376 directly influences the power to detect small effect sizes.<sup>27</sup> Furthermore, only 61 top hits  
377 from the GWAS step were pursued in individual samples, with the great majority of  
378 nominally significantly associated markers having been excluded from the subsequent  
379 analyses. These are the two most likely explanations as to why the present study may  
380 have missed previously reported association findings<sup>12</sup> Therefore, the reliable detection  
381 of genes with smaller effects requires larger sample sizes and individual genotyping.  
382 The only other SNP to reach experiment-wide significance in the combined analysis  
383 was rs304650 in the *SPATA5* gene on chromosome 4q27-q28. Joint analysis of the  
384 replication samples used in steps 2 and 3, which included a total of 985 cases and 2,014  
385 controls who were successfully genotyped for this SNP, revealed a significant  
386 association between this variant and AA, with a *P*-value of  $3.43 \times 10^{-4}$  (OR = 1.24 [1.10  
387 - 1.39]). A joint analysis of all of the investigated samples revealed a significant  
388 association between this variant and AA, with a *P*-value of  $1.58 \times 10^{-5}$  (OR = 1.23 [1.12 -  
389 1.35]). The SNP rs304650 maps to an intronic region of the *SPATA5* transcript. At the  
390 time of writing, the functional aspects of this protein are unknown. However, one study  
391 identified *SPATA2*, another member of the spermatogenesis associated protein family,  
392 as a susceptibility gene for the autoimmune disorder psoriasis<sup>28</sup>.

393 Interestingly, although *SPATA5* was not among the eight loci with genome-wide  
394 significance reported by Petukhova et al., the authors reported 6 SNPs in *SPATA5* with  
395 a *P*-value < 10<sup>-4</sup> in their supplementary material. Although not very strong, there is  
396 evidence for LD between rs304650 and the Pethukova et al. SNPs (ranging between r<sup>2</sup> =  
397 0.29 for rs11735364 and r<sup>2</sup> = 0.46 for rs2201997). Although this might be viewed as  
398 supportive evidence, a more detailed workup of the region in very large samples is  
399 required to allow more definitive conclusions to be drawn. It is also interesting that the  
400 *SPATA5* gene is located only 320kb distal to rs7682241, a genome-wide significant  
401 marker in the study of Petukhova et al., which strongly suggests the involvement of the  
402 *IL2/IL21* gene locus. Data from the CEU HapMap sample, however, show that  
403 rs304650 is not in LD with the best variants of *the IL2/IL21* gene locus. Thus, the two  
404 loci probably confer their risk independently of each other. It remains theoretically  
405 possible, however, that the true causal variant for AA may be a functional variant that is  
406 in moderate LD to the variants reported in both analyses and which is located between  
407 the *IL2/IL21* and *SPATA5* regions.”

408

409

410

411

## 412 **Acknowledgements**

413 The authors thank all patients and controls for participating in this study. The Heinz-  
414 Nixdorf Recall cohort (HNR) was established with the support of the Heinz Nixdorf  
415 Foundation (Germany). Silke Redler is a past recipient of a BONFOR fellowship from  
416 the Medical Faculty of the University of Bonn. Markus M. Nöthen is recipient of a grant

417 from the Alfried Krupp von Bohlen und Halbach-Stiftung. Regina C. Betz is a past  
418 recipient of an Emmy-Noether fellowship and is a recipient of a Heisenberg  
419 Professorship (both German Research Foundation (DFG)).

420

421 **Conflict of interest**

422 The authors have no conflicts of interest to declare.

423

- 426 1. Safavi KH, Muller SA, Suman VJ, Moshell AN, Melton LJ, 3rd: Incidence of  
427 alopecia areata in Olmsted County, Minnesota, 1975 through 1989. *Mayo Clin*  
428 *Proc* 1995; **70**: 628-633.
- 429 2. Blaumeiser B, van der Goot I, Fimmers R et al: Familial aggregation of alopecia  
430 areata. *J Am Acad Dermatol* 2006; 54: 627-632.
- 431 3. Jackow C, Puffer N, Hordinsky M, Nelson J, Tarrand J, Duvic M: Alopecia  
432 areata and cytomegalovirus infection in twins: genes versus environment? *J Am*  
433 *Acad Dermatol* 1998; 38: 418-425.
- 434 4. Tobin DJ: Characterization of hair follicle antigens targeted by the anti-hair  
435 follicle immune response. *J Investig Dermatol Symp Proc* 2003; 8: 176-181.
- 436 5. Entz P, Blaumeiser B, Betz RC et al: Investigation of the HLA-DRB1 locus in  
437 alopecia areata. *Eur J Dermatol* 2006; 16: 363-367.
- 438 6. Colombe BW, Lou CD, Price VH: The genetic basis of alopecia areata: HLA  
439 associations with patchy alopecia areata versus alopecia totalis and alopecia  
440 universalis. *J Investig Dermatol Symp Proc* 1999; 4: 216-219.
- 441 7. Xiao FL, Yang S, Yan KL et al: Association of HLA class I alleles with  
442 alopecia areata in Chinese Hans. *J Dermatol Sci* 2006; 41: 109-119.
- 443 8. de Andrade M, Jackow CM, Dahm N, Hordinsky M, Reveille JD, Duvic M:  
444 Alopecia areata in families: association with the HLA locus. *J Investig Dermatol*  
445 *Symp Proc* 1999; 4: 220-223.
- 446 9. Betz RC, Konig K, Flaquer A et al: The R620W polymorphism in PTPN22  
447 confers general susceptibility for the development of alopecia areata. *Br J*  
448 *Dermatol* 2008; 158: 389-391.
- 449 10. Kemp EH, McDonagh AJ, Wengraf DA et al: The non-synonymous C1858T  
450 substitution in the PTPN22 gene is associated with susceptibility to the severe  
451 forms of alopecia areata. *Hum Immunol* 2006; **67**: 535-539.  
452
- 453 11. Martinez-Mir A, Zlotogorski A, Gordon D et al: Genomewide scan for linkage  
454 reveals evidence of several susceptibility loci for alopecia areata. *Am J Hum*  
455 *Genet* 2007; **80**: 316-328.

- 456 12 Petukhova L, Duvic M, Hordinsky M *et al*: Genome-wide association study in  
457 alopecia areata implicates both innate and adaptive immunity. *Nature* 2010; **466**:  
458 113-117.
- 459 13. Krawczak M, Nikolaus S, von Eberstein H, Croucher PJ, El Mokhtari NE,  
460 Schreiber S: PopGen: population-based recruitment of patients and controls for  
461 the analysis of complex genotype-phenotype relationships. *Community Genet*  
462 2006; 9: 55-61.
- 463 14. Wichmann HE, Gieger C, Illig T: KORA-gen--resource for population genetics,  
464 controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* 2005;  
465 67 Suppl 1: S26-30.
- 466 15. Schmermund A, Mohlenkamp S, Stang A *et al*: Assessment of clinically silent  
467 atherosclerotic disease and established and novel risk factors for predicting  
468 myocardial infarction and cardiac death in healthy middle-aged subjects:  
469 rationale and design of the Heinz Nixdorf RECALL Study. *Risk Factors,*  
470 *Evaluation of Coronary Calcium and Lifestyle. Am Heart J* 2002; 144: 212-218.
- 471 16. Papassotiropoulos A, Stephan DA, Huentelman MJ *et al*: Common Kibra alleles  
472 are associated with human memory performance. *Science* 2006; 314: 475-478.
- 473 17. Moskva V, Norton N, Williams N, Holmans P, Owen M, O'Donovan M:  
474 Streamlined analysis of pooled genotype data in SNP-based association studies.  
475 *Genet Epidemiol* 2005; 28: 273-282.
- 476 18. Abraham R, Moskva V, Sims R *et al*: A genome-wide association study for  
477 late-onset Alzheimer's disease using DNA pooling. *BMC Med Genomics* 2008;  
478 1: 44.
- 479 19. : Signal detection in genome-wide association studies. *Proc of Proc of the 6th*  
480 *St.Petersburg Workshop on Simularion*. pp. 301-307.
- 481 20. Herold C, Becker T: Genetic association analysis with FAMHAP: a major  
482 program update. *Bioinformatics* 2009; 25: 134-136.
- 483 21. Armitage P: Tests for linear trends in proportions and frequencies. *Biometrics*  
484 1955; 11: 375-386.
- 485 22. Consortium. TWTCC: Genome-wide association study of 14,000 cases of seven  
486 common diseases and 3,000 shared controls. *Nature* 2007; 447: 661-678.
- 487 23. Petukhova L, Sousa EC, Jr., Martinez-Mir A *et al*: Genome-wide linkage  
488 analysis of an autosomal recessive hypotrichosis identifies a novel P2RY5  
489 mutation. *Genomics* 2008.

490 24. Barahmani N, de Andrade M, Slusser JP et al: Human leukocyte antigen class II  
491 alleles are associated with risk of alopecia areata. *J Invest Dermatol* 2008; 128:  
492 240-243.

493 25. Colombe BW, Price VH, Khoury EL, Garovoy MR, Lou CD: HLA class II  
494 antigen associations help to define two types of alopecia areata. *J Am Acad*  
495 *Dermatol* 1995; 33: 757-764.

496 26. Pearson JV, Huentelman MJ, Halperin RF et al: Identification of the genetic  
497 basis for complex disorders by use of pooling-based genomewide single-  
498 nucleotide-polymorphism association studies. *Am J Hum Genet* 2007; 80: 126-  
499 139.

500 27. Bosse Y, Bacot F, Montpetit A *et al*: Identification of susceptibility genes for  
501 complex diseases using pooling-based genome-wide association scans. *Hum*  
502 *Genet* 2009; 125: 305-318.

503 28. Capon F, Bijlmakers MJ, Wolf N et al: Identification of ZNF313/RNF114 as a  
504 novel psoriasis susceptibility gene. *Hum Mol Genet* 2008; 17: 1938-1945.

505

506

507

508

509

510

511

512

513 **Figure legends**

514

515 **Fig. 1**

516 Overall workflow. The study was conducted in three steps: (a) pooling-based analyses  
517 using three different approaches ((I) all cases vs. all controls, (II) cases 'positive family  
518 history' vs. all controls and (III) a sliding window analysis; (b) confirmation of selected  
519 best pooling-based findings through individual genotyping in previously pooled case-  
520 and control samples; (c) independent replication and follow-up analyzes in additional  
521 samples of cases and controls.

522

523 **Fig. 2**

524 Expression analysis of mRNA of SPATA5. SPATA5 was found to be expressed in  
525 human hair follicle, skin, kidney, and lung (samples displayed from left to right). The  
526 final lane shows a negative control.

527

**Table 1: Association between alopecia areata and selected markers in the case-control confirmation- and case-control replication analyses**

SNP	Chr	Position <sup>1</sup>	Allele	Cases-control confirmation analysis					Cases-control replication analysis						
				MAF <sup>2</sup>		Risk			MAF <sup>2</sup>		Risk			P <sub>corr.</sub> <sup>4</sup>	OR (95% CI) <sup>5</sup>
				Cases	Controls	Allele	P	OR (95% CI) <sup>3</sup>	Cases	Controls	Allele	P			
rs3099624	1	11 419 581	T/C	0.419 (T)	0.477 (T)	C	0.002	1.27 (1.09 - 1.47)	0.451 (T)	0.459 (T)	C	0.7	1	1.03 (0.89 - 1.20)	
rs304303	1	43 950 657	G/T	0.241 (T)	0.311 (T)	G	5.79 x 10 <sup>-5</sup>	1.42 (1.20 - 1.69)	0.269 (T)	0.275 (T)	G	0.761	1	1.03 (0.87 - 1.22)	
rs4660260	1	43 967 940	T/C	0.245 (T)	0.313 (T)	C	9.22 x 10 <sup>-5</sup>	1.40 (1.18 - 1.66)	0.278 (T)	0.281 (T)	C	0.838	1	1.02 (0.86 - 1.20)	
rs12403551	1	101 636 803	G/A	0.472 (A)	0.492 (G)	G	0.06	1.15 (0.99 - 1.34)	0.472 (A)	0.478 (A)	G	0.97	1	1.00 (0.86 - 1.17)	
rs2345724	2	18 548 555	G/A	0.052 (G)	0.029 (G)	G	0.003	1.84 (1.23 - 2.75)	0.052 (G)	0.042 (G)	G	0.185	1	1.27 (0.90 - 1.80)	
rs2666138	2	36 184 661	C/T	0.163 (C)	0.116 (C)	C	5.92 x 10 <sup>-4</sup>	1.48 (1.19 - 1.84)	0.139 (C)	0.150 (C)	T	0.42	1	0.91 (0.74 - 1.13)	
rs13409979	2	184 062 055	A/G	0.131 (A)	0.103 (A)	A	0.024	1.31 (1.03 - 1.66)	0.109 (A)	0.110 (A)	G	0.983	1	1.00 (0.78 - 1.27)	
rs1692617	3	23 438 818	T/C	0.399 (T)	0.463 (T)	C	0.001	1.30 (1.11 - 1.51)	0.463 (T)	0.440 (T)	T	0.219	1	0.91 (0.78 - 1.06)	
rs6414541	3	173 403 331	C/T	0.119 (C)	0.092 (C)	C	0.023	1.33 (1.04 - 1.71)	0.101 (C)	0.116 (C)	T	0.257	1	0.87 (0.68 - 1.11)	
rs4130791	4	1 404 684	A/G	0.105 (A)	0.064 (A)	A	2.16 x 10 <sup>-4</sup>	1.71 (1.29 - 2.26)	0.107 (A)	0.099 (A)	A	0.436	1	1.10 (0.86 - 1.41)	
rs16873952	4	23 241 241	A/G	0.148 (A)	0.108 (A)	A	0.001	1.44 (1.15 - 1.81)	0.107 (A)	0.129 (A)	G	0.072	1	0.81 (0.63 - 1.02)	
rs17023881	4	96 581 206	T/C	0.060 (T)	0.037 (T)	T	0.007	1.66 (1.15 - 2.40)	0.052 (T)	0.047 (T)	T	0.438	1	1.15 (0.81 - 1.61)	
rs7657799	4	105 594 872	G/T	0.031 (G)	0.016 (G)	G	0.014	1.98 (1.16 - 3.38)	0.022 (G)	0.022 (G)	G	0.88	1	1.04 (0.62 - 1.74)	
rs7664318	4	123 818 574	G/A	0.177 (G)	0.125 (G)	G	1.57 x 10 <sup>-4</sup>	1.51 (1.22 - 1.87)	0.152 (G)	0.146 (G)	G	0.574	1	1.06 (0.86 - 1.31)	
rs304650	4	124 303 368	G/A	0.328 (G)	0.272 (G)	G	0.001	1.31 (1.11 - 1.54)	0.360 (G)	0.301 (G)	G	0.001	<b>0.049</b>	1.31 (1.12 - 1.53)	
rs9997120	4	147 312 539	T/C	0.234 (T)	0.289 (T)	C	0.001	1.33 (1.12 - 1.58)	0.240 (T)	0.266 (T)	C	0.162	1	1.13 (0.95 - 1.35)	
rs2125856	5	122 073 387	C/T	0.325 (C)	0.268 (C)	C	0.002	1.32 (1.11 - 1.55)	0.299 (C)	0.291 (C)	C	0.379	1	1.08 (0.91 - 1.27)	
rs3115553	6	32 353 805	A/G	0.137 (A)	0.217 (A)	G	<b>7.11 x 10<sup>-8</sup></b>	1.74 (1.42 - 2.12)	0.139 (A)	0.208 (A)	G	<b>6.27 x 10<sup>-8</sup></b>	<b>3.07 x 10<sup>-4</sup></b>	1.63 (1.32 - 2.01)	
rs9268528	6	32 491 086	G/A	0.477 (G)	0.381 (G)	G	<b>2.52 x 10<sup>-7</sup></b>	1.48 (1.27 - 1.73)	0.499 (A)	0.391 (G)	G	<b>6.82 x 10<sup>-9</sup></b>	<b>3.34 x 10<sup>-7</sup></b>	1.56 (1.34 - 1.82)	
rs9275572	6	32 786 977	A/G	0.304 (A)	0.420 (A)	G	<b>2.50 x 10<sup>-10</sup></b>	1.65 (1.41 - 1.94)	0.303 (A)	0.430 (A)	G	<b>7.94 x 10<sup>-11</sup></b>	<b>3.89 x 10<sup>-9</sup></b>	1.71 (1.46 - 2.01)	
rs916305	6	117 986 388	T/C	0.123 (T)	0.107 (T)	T	0.195	1.17 (0.93 - 1.49)	0.107 (T)	0.122 (T)	C	0.246	1	0.87 (0.68 - 1.10)	
rs505532	7	29 050 917	T/C	0.320 (T)	0.260 (T)	T	5.20 x 10 <sup>-4</sup>	1.34 (1.13 - 1.58)	0.300 (T)	0.276 (T)	T	0.166	1	1.12 (0.95 - 1.33)	
rs718121	8	3 146 968	T/C	0.331 (T)	0.408 (T)	C	4.73 x 10 <sup>-5</sup>	1.39 (1.19 - 1.63)	0.401 (T)	0.372 (T)	T	0.122	1	0.88 (0.76 - 1.03)	
rs2979742	8	25 451 514	T/C	0.332 (T)	0.374 (T)	C	0.02	1.20 (1.03 - 1.41)	0.360 (T)	0.349 (T)	T	0.54	1	0.95 (0.81 - 1.11)	
rs10503991	8	36 539 160	G/A	0.126 (G)	0.102 (G)	G	0.04	1.28 (1.01 - 1.62)	0.107 (G)	0.116 (G)	A	0.457	1	0.91 (0.72 - 1.16)	
rs4738296	8	74 020 093	C/A	0.041 (C)	0.018 (C)	C	8.62 x 10 <sup>-4</sup>	2.33 (1.43 - 3.79)	0.032 (C)	0.039 (C)	A	0.388	1	0.83 (0.55 - 1.26)	
rs7022183	9	13 006 817	C/T	0.092 (C)	0.066 (C)	C	0.013	1.44 (1.09 - 1.91)	0.070 (C)	0.065 (C)	C	0.626	1	1.08 (0.80 - 1.45)	
rs7036795	9	70 358 138	C/T	0.214 (C)	0.181 (C)	C	0.03	1.23 (1.02 - 1.49)	0.181 (C)	0.197 (C)	T	0.628	1	0.95 (0.78 - 1.16)	
rs10512241	9	97 977 483	A/G	0.346 (A)	0.405 (A)	G	0.002	1.28 (1.10 - 1.50)	0.350 (A)	0.377 (A)	G	0.142	1	1.13 (0.96 - 1.32)	
rs17479692	10	91 278 439	G/T	0.157 (G)	0.109 (G)	G	3.52 x 10 <sup>-4</sup>	1.52 (1.21 - 1.91)	0.108 (G)	0.135 (G)	T	0.068	1	0.80 (0.63 - 1.02)	
rs9419187	10	133 761 922	T/C	0.318 (T)	0.373 (T)	C	0.002	1.28 (1.09 - 1.50)	0.368 (T)	0.340 (T)	T	0.121	1	0.88 (0.76 - 1.03)	
rs16928055	11	35 870 147	C/T	0.151 (C)	0.104 (C)	C	2.54 x 10 <sup>-4</sup>	1.55 (1.22 - 1.95)	0.116 (C)	0.113 (C)	C	0.557	1	1.07 (0.85 - 1.36)	
rs1216476	11	99 915 556	G/A	0.230 (G)	0.177 (G)	G	6.13 x 10 <sup>-4</sup>	1.39 (1.15 - 1.68)	0.206 (G)	0.216 (G)	A	0.539	1	0.94 (0.78 - 1.14)	
rs11224294	11	99 954 372	C/T	0.098 (C)	0.063 (C)	C	9.34 x 10 <sup>-4</sup>	1.62 (1.22 - 2.15)	0.078 (C)	0.091 (C)	T	0.24	1	0.85 (0.64 - 1.12)	
rs11600229	11	130 998 481	G/A	0.042 (G)	0.057 (G)	A	0.068	1.38 (0.97 - 1.95)	0.052 (G)	0.039 (G)	G	0.091	1	0.74 (0.52 - 1.05)	
rs11223339	11	132 423 674	T/G	0.217 (T)	0.163 (T)	T	4.02 x 10 <sup>-4</sup>	1.42 (1.17 - 1.73)	0.199 (T)	0.203 (T)	G	0.783	1	0.97 (0.81 - 1.18)	
rs10791360	11	133 807 109	C/A	0.072 (C)	0.084 (C)	A	0.25	1.18 (0.89 - 1.56)	0.078 (C)	0.073 (C)	C	0.607	1	0.93 (0.70 - 1.23)	
rs2270797	12	128 248	A/G	0.054 (A)	0.031 (A)	A	0.003	1.81 (1.22 - 2.68)	0.029 (A)	0.038 (A)	G	0.242	1	0.77 (0.50 - 1.19)	
rs2110597	12	12 832 280	G/A	0.381 (G)	0.300 (G)	G	1.42 x 10 <sup>-5</sup>	1.44 (1.22 - 1.68)	0.362 (G)	0.332 (G)	G	0.101	1	1.14 (0.97 - 1.34)	
rs10506012	12	26 918 468	A/G	0.150 (A)	0.189 (A)	G	0.007	1.32 (1.08 - 1.61)	0.172 (A)	0.189 (A)	G	0.262	1	1.12 (0.92 - 1.37)	
rs12228387	12	117 886 926	A/G	0.059 (A)	0.036 (A)	A	0.005	1.70 (1.18 - 2.45)	0.054 (A)	0.060 (A)	G	0.532	1	0.90 (0.65 - 1.25)	
rs12430	17	71 453 550	T/C	0.114 (T)	0.086 (T)	T	0.017	1.36 (1.06 - 1.76)	0.103 (T)	0.102 (T)	T	0.821	1	1.03 (0.80 - 1.32)	
rs7228576	18	10 398 492	T/C	0.144 (T)	0.114 (T)	T	0.024	1.30 (1.04 - 1.63)	0.149 (T)	0.129 (T)	T	0.115	1	1.19 (0.96 - 1.48)	
rs9954649	18	18 867 268	G/A	0.384 (G)	0.331 (G)	G	0.004	1.26 (1.08 - 1.47)	0.340 (G)	0.356 (G)	A	0.387	1	0.93 (0.80 - 1.09)	
rs1431704	18	35 107 732	C/T	0.080 (C)	0.047 (C)	C	4.56 x 10 <sup>-4</sup>	1.79 (1.30 - 2.47)	0.069 (C)	0.059 (C)	C	0.281	1	1.18 (0.87 - 1.60)	
rs9952976	18	42 561 717	G/A	0.241 (G)	0.277 (G)	A	0.034	1.20 (1.01 - 1.43)	0.286 (G)	0.284 (G)	G	0.829	1	0.98 (0.83 - 1.16)	
rs11666141	19	15 295 382	C/T	0.227 (C)	0.163 (C)	C	3.18 x 10 <sup>-5</sup>	1.51 (1.24 - 1.83)	0.186 (C)	0.170 (C)	C	0.3	1	1.11 (0.91 - 1.35)	
rs2766671	20	51 617 909	T/C	0.126 (T)	0.107 (T)	T	0.123	1.20 (0.95 - 1.53)	0.120 (T)	0.112 (T)	T	0.479	1	1.09 (0.86 - 1.38)	
rs9982439	21	42 622 269	C/T	0.055 (C)	0.030 (C)	C	0.001	1.88 (1.27 - 2.78)	0.040 (C)	0.037 (C)	C	0.725	1	1.07 (0.73 - 1.58)	



<sup>1</sup> Position in base pairs based on NCBI build 36. <sup>2</sup> MAF = Minor allele frequency with minor allele in parentheses. <sup>3</sup> OR = odds ratio; CI = confidence interval. OR-calculation based on the risk allele deriving from the initial pooling confirmation step. <sup>4</sup> Nominal *P*-value was corrected for the number of SNPs finally investigated (n = 49) in the initial pooling-based confirmation step.

**Table 2: Follow-up analysis of rs304650 in an additional independent sample of 536 AA cases and 655 controls.**

SNP	Chr	Position <sup>1</sup>	Allele	MAF <sup>2</sup>		Risk		P	OR (95% CI) <sup>3</sup>
				Cases	Controls	Allele			
rs304650	4	124 303 368	G/A	0.347	0.316	<b>G</b>	0.127	1.15 (0.97 - 1.36)	

<sup>1</sup> Position in base pairs based on NCBI build 36. <sup>2</sup> MAF = Minor allele frequency with minor allele in parentheses. <sup>3</sup> OR = odds ratio; CI = confidence interval. OR-calculation based on the risk allele deriving from the initial pooling confirmation step.



