

ARTICLE

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

Lina M Forstbauer^{1,20}, Felix F Brockschmidt^{1,2,20}, Valentina Moskvina³, Christine Herold⁴, Silke Redler¹, Alexandra Herzog¹, Axel M Hillmer⁵, Christian Meesters^{4,6}, Stefanie Heilmann^{1,2}, Florian Albert¹, Margrieta Alblas^{1,2}, Sandra Hanneken⁷, Sibylle Eigelshoven⁷, Kathrin A Giehl⁸, Dagny Jagielska^{1,9}, Ulrike Blume-Peytavi⁹, Natalie Garcia Bartels⁹, Jennifer Kuhn^{10,11,12}, Hans Christian Hennies^{10,11,12}, Matthias Goebeler¹³, Andreas Jung¹³, Wiebke K Peitsch¹⁴, Anne-Katrin Kortüm¹⁵, Ingrid Moll¹⁵, Roland Kruse¹⁶, Gerhard Lutz¹⁷, Hans Wolff⁷, Bettina Blaumeiser¹⁸, Markus Böhm¹⁹, George Kirov³, Tim Becker^{4,6}, Markus M Nöthen^{1,2} and Regina C Betz^{*,1}

Alopecia areata (AA) is a common hair loss disorder, which is thought to be a tissue-specific autoimmune disease. Previous research has identified a few AA susceptibility genes, most of which are implicated in autoimmunity. To identify new genetic variants and further elucidate the genetic basis of AA, we performed a genome-wide association study using the strategy of pooled DNA genotyping (729 cases, 656 controls). The strongest association was for variants in the HLA region, which confirms the validity of the pooling strategy. The selected top 61 single-nucleotide polymorphisms (SNPs) were analyzed in an independent replication sample (454 cases, 1364 controls). Only one SNP outside of the HLA region (rs304650) showed significant association. This SNP was then analyzed in a second independent replication sample (537 cases, 657 controls). The finding was not replicated on a significant level, but showed the same tendency. A combined analysis of the two replication samples was then performed, and the SNP rs304650 showed significant association with $P=3.43 \times 10^{-4}$ (OR=1.24 (1.10–1.39)). This SNP maps to an intronic region of the *SPATA5* (spermatogenesis-associated protein 5) gene on chromosome 4. The results therefore suggest the *SPATA5* locus is a new susceptibility locus for AA.

European Journal of Human Genetics (2012) 20, 326–332; doi:10.1038/ejhg.2011.185; published online 26 October 2011

Keywords: alopecia areata; genome-wide association study; autoimmunity; hair loss

INTRODUCTION

Alopecia areata (AA) is a common hair loss disorder, which affects approximately 1–2% of the general population. It affects both sexes and all age groups.¹ Affected individuals present with a non-scarring, circumscribed hair loss, which has a sudden onset and a recurrent course. The scalp is the most commonly affected site, although all hair-bearing areas of the skin may be involved. Episodes of hair loss typically start with isolated hairless patches. These extend centrifugally and may coalesce. AA is divided into three main clinical types on the basis of the degree of hair loss and the sites affected: (i) patchy AA; (ii) AA totalis, affecting the whole scalp; and (iii) AA universalis, affecting the whole body. Individuals with AA may experience complete remission, a chronic course, or progression towards AA totalis or AA universalis. Familial AA also occurs, and recurrence

risks of 5–6% have been reported for the children of affected individuals.^{2,3} The pattern of familiarity and a limited twin study, which reports a concordance rate in monozygotic twins of 55%,³ suggest that multiple genetic factors and environmental factors are involved in the pathology of AA. Nevertheless, the etiopathogenesis of AA remains poorly understood. One hypothesis is that AA is a tissue-specific autoimmune disease of the hair follicle.⁴ This is supported by reports of association between AA and specific HLA alleles.^{5–8} Association has also been reported between AA and the *W620* variant of the *PTPN22* gene (protein tyrosine phosphatase, nonreceptor type 22), which has been implicated in several autoimmune disorders.^{9,10} To date, only one systematic genome-wide linkage study of AA has been performed in humans. This identified four susceptibility loci on chromosomes 6, 10, 16, and 18, respectively.¹¹ The pattern of

¹Institute of Human Genetics, University of Bonn, Bonn, Germany; ²Department of Genomics, Life and Brain Center, University of Bonn, Bonn, Germany; ³MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, UK; ⁴German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany; ⁵Genome Technology and Biology Group, Genome Institute of Singapore, Singapore, Singapore; ⁶Institute for Medical Biometry, Informatics and Epidemiology, University of Bonn, Bonn, Germany; ⁷Department of Dermatology, University of Düsseldorf, Düsseldorf, Germany; ⁸Department of Dermatology, University of Munich, Munich, Germany; ⁹Clinical Research Center for Hair and Skin Science, Department of Dermatology and Allergy, Charité-Universitätsmedizin Berlin, Berlin, Germany; ¹⁰Division of Dermatogenetics, Cologne Center for Genomics, University of Cologne, Cologne, Germany; ¹¹Center for Molecular Medicine, University of Cologne, Cologne, Germany; ¹²Cologne Cluster of Excellence on Cellular Stress Responses in Aging-associated Diseases, University of Cologne, Cologne, Germany; ¹³Department of Dermatology, University of Giessen, Giessen, Germany; ¹⁴Department of Dermatology, University Medical Centre Mannheim, University of Heidelberg, Mannheim, Germany; ¹⁵Department of Dermatology and Venerology, University Hospital Hamburg-Eppendorf, Hamburg, Germany; ¹⁶Dermatological Practice, Paderborn, Germany; ¹⁷Hair and Nail, Wesseling, Germany; ¹⁸Department of Medical Genetics, University and University Hospital of Antwerp, Antwerp, Belgium; ¹⁹Department of Dermatology, University of Münster, Münster, Germany

*Correspondence: Professor RC Betz, Institute of Human Genetics, University of Bonn, Sigmund-Freud Strasse 25, D-53127 Bonn, Germany. Tel: +49 228 287 51023; Fax: +49 228 287 51011; E-mail: regina.betz@uni-bonn.de

²⁰These authors contributed equally to this work.

Received 9 May 2011; revised 5 August 2011; accepted 7 September 2011; published online 26 October 2011

familiarity and a limited twin study which reports a concordance rate in monozygotic twins of 55%.

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

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

MATERIALS AND METHODS

DNA samples

The analyses involved 1720 individuals with AA (cases) and 2677 controls. For cases, the inclusion criterion was a dermatologist-assigned diagnosis of AA. Patients with Down's syndrome or Turner's syndrome were excluded. The cases were recruited from: (i) outpatient clinics in Belgium (University Hospitals of Antwerp and Gent) and Germany (University Hospitals of Munich, Münster, Düsseldorf, Berlin, Bonn, Gießen, Hamburg and Mannheim); (ii) a private dermatology practice (Wesseling, Germany); and (iii) AA self-help support groups (Germany, The Netherlands). All cases were asked whether or not they had a positive family history of AA. This was defined as a history of at least one first- or second-degree relative with any form of the disorder. The controls were drawn from the general population, and were therefore not specifically screened for the absence of AA. They were healthy unrelated blood donors from the University Hospital of Bonn ($n=1313$) or participants in one of three population-based epidemiological studies: (1) PopGen¹³ ($n=490$); (2) KORA¹⁴ ($n=490$); and (3) HNR¹⁵ ($n=384$). All cases and controls were of central European origin. Ethical approval was obtained from the respective ethics committees, and all participants provided written informed consent before blood sampling. The DNA of patients and blood donors was extracted from peripheral blood leukocytes. This was achieved by salting out with saturated NaCl solution according to standard methods, or through the use of a Chemagic Magnetic Separation Module I (Chemagen, Baesweiler, Germany) in accordance with the manufacturer's instructions. The DNA was stored in liquid nitrogen until use. For the controls from the three population-based epidemiological studies, the genotypes were taken from previously generated GWA data sets. The PopGen and KORA data sets were generated within the German National Genome Research Network to serve as a national research resource. The HNR data set was generated as part of a collaboration to generate a set of universal controls for genetic studies.

DNA pooling

The DNA samples of 729 AA cases and 656 blood donor controls were selected at random. These samples were then used to generate a control pool and two AA case pools. The first AA case pool consisted of the DNA of all 729 AA cases (including the 224 AA cases with a positive family history). The second AA case pool consisted of the DNA of 'positive family history' AA patients only ($n=224$). To ensure equal amounts of DNA in each pool, each individual DNA sample was subjected to a series of dilution steps (above 100 ng/ μ l to 20 ng/ μ l, and then to a final concentration of 10 ng/ μ l). Each DNA sample was double-quantified using a NanoDrop device (PEQLAB Biotechnologie GmbH,

Erlangen, Germany) and adjusted to $\pm 10\%$ of the required concentration. For the 10 ng/ μ l dilution, a broader range of -60 to $+80\%$ was used to reflect the measurement error of the NanoDrop device at low DNA concentrations. Equimolar amounts of each DNA sample (100 ng) were pipetted manually into a large tube to form the pool. Each pool was then concentrated to 60 ng/ μ l using the Microcon YM-100 Centrifugal Filter Device (Millipore GmbH, Schwalbach, Germany), and then requantified with the NanoDrop device.

Genotyping of the pooled DNA and data analysis

The Illumina Sentrix HumanHap550v3 genotyping BeadChips were used. These contain more than 550 000 tag SNPs (Illumina Inc., San Diego, CA, USA). To avoid inter-experimental variation, each pool was genotyped on five chips in accordance with the manufacturer's recommendations for individual samples. SNP allele frequencies were estimated using data from the BeadArray Reader imaging (Illumina Inc.) and Illumina's genotyping software Bead Studio 2.0 (Illumina Inc.). Three analyses were performed on the basis of the allele frequency estimates obtained from the genotyping experiment: (i) 729 cases *versus* 656 controls; (ii) 224 cases with a positive family history *versus* 656 controls; and (iii) a sliding window analysis using the data set of 729 cases *versus* 656 controls, in accordance with recommendations described elsewhere.¹⁶

For each replicate, the following approximation of allele frequencies in cases and controls was performed on the basis of the raw data: $f_i = X_{raw} / (X_{raw} + Y_{raw})$, where X_{raw} and Y_{raw} are the intensities of the two dyes (Cy5 and Cy3) used to genotype SNPs on the Illumina platform, and i is the replicate number. The obtained allele frequency was then averaged over the number of replicates of each pool $f = (f_1 + \dots + f_M) / M$ (where M is the number of replicates), and corrected for unequal amplification using the formula $f_{corrected} = f / (f + k \cdot (1 - f))$.¹⁷ For each SNP, the correction factor k was estimated using the HapMap CEPH data (<http://hapmap.ncbi.nlm.nih.gov/>) and the formula $k = (f_{con} \cdot f_{con} \cdot f_{CEPH}) / (f_{CEPH} \cdot f_{con} \cdot f_{CEPH})$, where f_{con} is the allele frequency in controls obtained from the pooling experiment, and f_{CEPH} is the allele frequency calculated from the HapMap CEPH data.

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

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

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

A signal processing analysis technique was used to interpret the results of the case-control association analysis. This enables the detection of genetic association with disease, while taking into account the significance of several subsequent genetic markers in a sliding window. Moskvina *et al.*¹⁷ has described a theoretical approach to calculating the probability of at least one false alarm being flagged by the detection statistic under the Null hypothesis of no signal (association).¹⁴ This is equivalent to the probability of type I error when taking the number of comparisons in the window into account, and thus provides a 'genome-wide window-based' significance level.¹⁹

SNP selection, confirmation of pooling results and independent replication

A number of SNPs were selected to confirm the pooling results on an individual genotyping level. A replication step was performed in two independent case-control samples (for an overview, see Figure 1). The following SNPs were selected from the three pooling-based analyses: (i) the top 50 SNPs from the analysis '729 cases *versus* 656 controls'; (ii) the top 20 SNPs from the analysis '224 cases with a positive family history *versus* 656 controls'; and

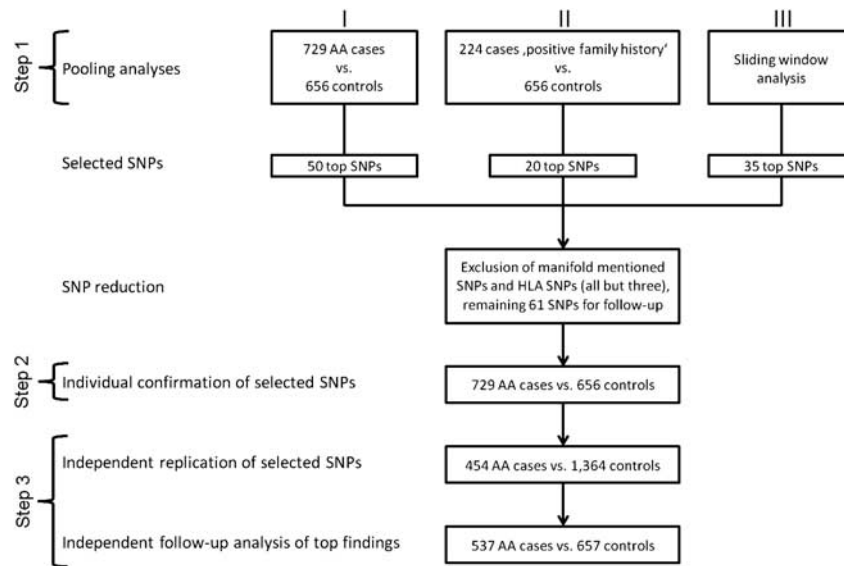


Figure 1 Overall workflow. The study was conducted in three steps: (a) pooling-based analyses using three different approaches (I) all cases vs all controls, (II) cases 'positive family history' vs all controls and (III) a sliding window analysis; (b) confirmation of selected best pooling-based findings through individual genotyping in previously pooled case- and control samples; (c) independent replication and follow-up analyses in additional samples of cases and controls.

(iii) the top 35 SNPs from the 'sliding window analysis'. All selected SNPs had shown P -values of $<1 \times 10^{-4}$ in the respective pooling analysis. As SNPs from the previously known HLA region were highly overrepresented, all but three of these SNPs were excluded. The three remaining HLA SNPs, rs3115553 (chr. 6: 32 353 805 bp), rs9275141 (chr. 6: 32 759 095 bp) and rs9275572 (chr. 6: 32 786 977 bp), were selected, as they were the best findings in the HLA region in the 'sliding window analysis'. These were used as positive controls in subsequent analyses. Regrettably, rs9275141 failed during the assay design of the Sequenom iPLEX reaction (Sequenom GmbH, Hamburg, Germany). It was therefore replaced by the next best SNP in the HLA region from the 'sliding window analysis', rs9268528 (chr. 6: 32 491 086 bp). By excluding other SNPs from the HLA region and SNPs appearing in more than one of the analyses (doubles and triples), it was possible to reduce the SNP set for the confirmation and replication step (step 2) from 105 to 61 SNPs. Individual genotyping was performed on the Illumina platform. For the DNA samples from the HNR-, KORA- and Popgen controls ($n=1364$), HumanHap550v3 genotyping Bead-Chips (Illumina Inc.) were used. For all other DNA samples, Sequenom's Compact MALDI-ToF Mass Array system and iPLEX Gold reagents (Sequenom GmbH) in multiplex reactions were used. Primer sequences and Sequenom's standard assay conditions are available upon request. All primers were checked by MALDI-ToF. For quality reasons, a success rate of at least 95% was required for all analyzed SNPs. A 95% call rate was required for each of the samples used in the confirmation and replication steps. In the confirmation step (step 2), DNA samples from eight AA cases and 11 controls failed to fulfill the quality criteria and were excluded from the analysis, and thus 721 AA cases and 645 controls remained for the analyses. In the independent replication step (step 2), four DNA case samples failed to fulfill the quality criteria, and thus 450 cases and 1 364 controls remained for the analyses. Of the 61 selected SNPs, six SNPs (rs12493901, rs30117, rs41515, rs4777450, rs7246435 and rs9520256) were technical failures, and five SNPs (rs10123149, 11098149, rs6700586, rs7099812 and rs724841) had to be excluded, as they did not reach the required call rate of $\geq 95\%$ in one or more of the confirmation and/or replication samples. The SNP rs7334982 was not biallelic in the present samples and was therefore excluded from further analysis. Thus, a set of 49 SNPs remained for further analysis.

Statistical analyses of individual genotyping data

The FAMHAP software package²⁰ was used for the association and haplotype analyses. The Armitage trend test was used for the single marker analyses.²¹ All

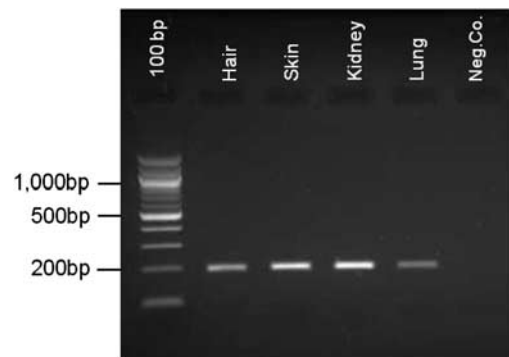


Figure 2 Expression analysis of mRNA of *SPATA5*. *SPATA5* was found to be expressed in human hair follicle, skin, kidney, and lung (samples displayed from left to right). The final lane shows a negative control.

SNPs met the following quality criteria: minor allele frequency $>1\%$, P_{HWE} in cases >0.001 and P_{HWE} in controls >0.05 . The resulting P -values were corrected for multiple testing according to the number of SNPs successfully analyzed on the individual genotyping level ($n=49$).

Expression analyses

We used the forward primer 5'-CCTTCAAACCGACGCATAC-3' and the reverse primer 5'-GCAGCCACTCTTCTCTGA-3' to analyse the expression of *SPATA5* in human hair follicle and skin samples (expected product size: 197 bp). As *SPATA5* expression has been proven in human kidney and lung (<http://www.genecards.org/cgi-bin/carddisp.pl?gene=SPATA5&search=SPATA5>), these tissues were included as positive controls. Total RNA was extracted from human hair follicles and skin using the RNeasy Micro Kit (Qiagen, Hilden, Germany), and single strand cDNA was synthesized from a total of 400 ng RNA using the Super Script III First Strand Synthesis System (Invitrogen, Karlsruhe, Germany). Single strand cDNA from kidney and lung was obtained from the human Multiple-Tissue cDNA Panel I (LOT Nr. 6060248; Clontech, Takara Bio Europe/Clontech, Saint-Germain-en-Laye, France). A negative reverse transcription reaction (no enzyme) was included as a negative control (Figure 2).

RESULTS

Step 1: Pooling-based approach

The genome-wide pooling-based approach involved three DNA-pools: (1) 729 AA cases (including the 224 cases with a positive family history), (2) 224 AA cases with a positive family history, and (3) 656 controls. Each pool was genotyped on five replicates of the Illumina Sentrix HumanHap550v3 genotyping BeadChip. Pool 1 was successfully analyzed in all five replicates. Pool 2 was successfully analyzed in four replicates. Two of the five control chips were excluded following quality control filtering.

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

Markers were excluded if the coefficient of variation was > 1 in at least one of three DNA pools, or if the correction factor k was > 4 or < 0.25 . The remaining 487 932 SNPs were corrected for k , and SNPs with minor allele frequencies of $< 5\%$ in controls were filtered out. The remaining 468 389 SNPs were tested for association using modified Z -statistics.¹⁸ Separate comparisons with controls were made for cases and familial cases. The results of the case-control comparison were further analyzed in a sliding window. The best markers from the top regions identified by the sliding window analysis¹⁹ were selected for replication. Thus, three different analyses were performed: (I) 729 AA cases versus 656 controls; (II) 224 AA cases with a positive family history versus 656 controls; and (III) a sliding window analysis (Figure 1). These analyses identified 31 SNPs with P -values $< 5 \times 10^{-7}$ (Supplementary Table 1), resulting in a total of 18 SNPs after the exclusion of duplicates and triplicates. Of these 18 SNPs, 8 SNPs were localized in the HLA region, and 10 SNPs were localized elsewhere in the genome. The best SNP of all three analyses was rs9952976 (chr. 18: 42 561 717 bp), which had a P -value of 6.48×10^{-14} in analyses I and III (Supplementary Table 1). The best SNP in analysis II was rs9275572, which is localized in the HLA region ($P=1.87 \times 10^{-8}$). This SNP was also the best HLA-SNP in analyses I ($P=1.00 \times 10^{-11}$) and III ($P=5.67 \times 10^{-12}$; Supplementary Table 1). When the best SNPs from each analysis were considered (see selection criteria in Materials and methods section), three SNPs (rs9275141 (chr. 6: 32 759 095 bp); rs9275572 (chr. 6: 32 786 977 bp); and rs9952976 (chr. 18: 42 561 717 bp)) appeared in all three analyses. Two of these SNPs (rs9275141 and rs9275572) are localized in the HLA region (Supplementary Table 1).

Step 2: Individual confirmation and independent replication

The top 50 SNPs from analysis I, the top 20 SNPs from analysis II, and the top 35 SNPs from analysis III, were selected for further analysis (Figure 1). The elimination of duplicates and triplicates resulted in a total of 61 SNPs (see Materials and methods section and Supplementary Table 1). To confirm the association findings of the 61 selected SNPs in the pooling approach, individual genotyping was performed in the previously pooled discovery sample of 729 AA cases and 656 controls. An independent replication step involved 454 AA cases and 1364 controls. Following quality control, 49 SNPs remained for analysis (see Materials and methods section). With the exception of five SNPs, this analysis confirmed the pooling results at a nominal level of significance, and thus demonstrated the validity of the DNA

pooling approach. The strongest association was found for the three HLA-SNPs (rs3115553, rs9268528 and rs9275572). The SNP rs9275572 showed the strongest association with $P=2.50 \times 10^{-10}$ (OR=1.65 (1.41–1.94); Table 1). These were the only SNPs to withstand correction for multiple testing using the previously suggested threshold of $P=5 \times 10^{-7}$.²² The remaining SNPs failed to show strong association. The best association finding outside of the HLA region was for rs2110597 (chr. 12: 12 832 280 bp) with $P=1.42 \times 10^{-5}$ (OR=1.44 (1.22–1.68)). The best finding from the pooling-based analysis, rs9952976 (chr. 18: 42 561 717 bp), showed only borderline significance on the level of individual genotyping, with $P=0.034$ (OR=1.20 (1.01–1.43)). This was one of the weakest association findings in this confirmation step (Table 1).

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

Step 3: Follow-up analysis of the top finding

A second independent sample of 537 cases and 657 controls was then used to investigate the association finding for rs304650 further (Figure 1). Genotyping of rs304650 failed in one case and two controls, and thus, 536 cases and 655 controls remained. In this analysis, rs304650 could not be replicated at a significant level ($P=0.127$; Table 2). However, the risk allele remained the same. A combined analysis was therefore performed using all AA cases and controls from the independent replication and follow-up steps (a total of 985 cases and 2014 controls were successfully genotyped for this SNP; data not shown). Here, rs304650 showed stronger association, with $P=3.43 \times 10^{-4}$ (OR=1.24 (1.10–1.39)). After combining all cases and controls, we obtained a P -value of 1.58×10^{-5} ; OR=1.23 (1.12–1.35). Interestingly, *SPATA5* expression was observed in hair follicles and skin, which confirms the importance of this gene in terms of hair biology (Figure 2).

DISCUSSION

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

The major histocompatibility complex on chromosome 6p21.3 was identified as a major risk locus for AA. Previous research by our group and others has implicated various HLA alleles in AA susceptibility. The best replicated findings have been for alleles of the DRB1 and DQB1 loci.^{5,6,8,23,24} The present highly significant findings for variants in the

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

SNP	Chr	Position ^b	Alleles	Cases-control confirmation analysis					Cases-control replication analysis					
				Cases	Controls	Risk allele	P	OR (95% CI) ^c	Cases	Controls	Risk allele	P	P _{corr.} ^d	OR (95% CI) ^c
				MAF ^a					MAF ^a					
rs099624	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 × 10 ⁻⁵	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 × 10 ⁻⁵	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 × 10 ⁻⁴	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 × 10 ⁻⁴	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 × 10 ⁻⁴	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	0.049	1.31 (1.12–1.53)
rs997120	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	7.11 × 10 ⁻⁸	1.74 (1.42–2.12)	0.139 (A)	0.208 (A)	G	6.27 × 10 ⁻⁶	3.07 × 10 ⁻⁴	1.63 (1.32–2.01)
rs9268528	6	32 491 086	G/A	0.477 (G)	0.381 (G)	G	2.52 × 10 ⁻⁷	1.48 (1.27–1.73)	0.499 (A)	0.391 (G)	G	6.82 × 10 ⁻⁹	3.34 × 10 ⁻⁷	1.56 (1.34–1.82)
rs9275572	6	32 786 977	A/G	0.304 (A)	0.420 (A)	G	2.50 × 10 ⁻¹⁰	1.65 (1.41–1.94)	0.303 (A)	0.430 (A)	G	7.94 × 10 ⁻¹¹	3.89 × 10 ⁻⁹	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)
rs050532	7	29 050 917	T/C	0.320 (T)	0.260 (T)	T	5.20 × 10 ⁻⁴	1.34 (1.13–1.58)	0.300 (T)	0.276 (T)	T	0.166	1	1.12 (0.95–1.33)
rs1718121	8	3 146 968	T/C	0.331 (T)	0.408 (T)	C	4.73 × 10 ⁻⁵	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 × 10 ⁻⁴	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 × 10 ⁻⁴	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 × 10 ⁻⁴	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 × 10 ⁻⁴	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 × 10 ⁻⁴	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	A/G	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 × 10 ⁻⁴	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 × 10 ⁻⁵	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)

Table 1 (Continued)

SNP	Chr	Position ^b	Alleles	Cases-control confirmation analysis				Cases-control replication analysis				
				MAF ^a	Risk allele	P	OR (95% CI) ^c	MAF ^a	Risk allele	P	OR (95% CI) ^c	
rs7228576	18	10 398 492	T/C	0.144 (T)	0.114 (T)	0.024	1.30 (1.04–1.63)	0.149 (T)	0.129 (T)	0.115	1	1.19 (0.96–1.48)
rs9954649	18	18 867 268	G/A	0.384 (G)	0.331 (G)	0.004	1.26 (1.08–1.47)	0.340 (G)	0.356 (G)	0.387	1	0.93 (0.80–1.09)
rs1431704	18	35 107 732	C/T	0.080 (C)	0.047 (C)	4.56 × 10 ⁻⁴	1.79 (1.30–2.47)	0.069 (C)	0.059 (C)	0.281	1	1.18 (0.87–1.60)
rs9952976	18	42 561 717	G/A	0.241 (G)	0.277 (G)	0.034	1.20 (1.01–1.43)	0.286 (G)	0.284 (G)	0.829	1	0.98 (0.83–1.16)
rs11666141	19	15 295 382	C/T	0.227 (C)	0.163 (C)	3.18 × 10 ⁻⁵	1.51 (1.24–1.83)	0.186 (C)	0.170 (C)	0.3	1	1.11 (0.91–1.35)
rs2766671	20	51 617 909	T/C	0.126 (T)	0.107 (T)	0.123	1.20 (0.95–1.53)	0.120 (T)	0.112 (T)	0.479	1	1.09 (0.86–1.38)
rs9982439	21	42 622 269	C/T	0.055 (C)	0.030 (C)	0.001	1.88 (1.27–2.78)	0.040 (C)	0.037 (C)	0.725	1	1.07 (0.73–1.58)

Abbreviations: CI, confidence interval; MAF, minor allele frequency; OR, odds ratio.
^aOR, calculation based on the risk allele deriving from the initial pooling confirmation step.
^bPosition in base pairs based on NCBI build 36.
^cMinor allele in parentheses.
^dNominal *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 ^b	Allele	MAF ^a		Risk		OR (95% CI) ^c
				Cases	Controls	Allele	P	
rs304650	4	124 303 368	G/A	0.347	0.316	G	0.127	1.15 (0.97–1.36)

Abbreviations: CI, confidence interval; MAF, minor allele frequency; OR, odds ratio.
^aOR, calculation based on the risk allele deriving from the initial pooling confirmation step.
^bPosition in base pairs based on NCBI build 36.
^cMinor allele in parentheses.

HLA region demonstrate that the pooling-based strategy is a valid alternative to individual genotyping in complex disorders.²⁵ Although the pooling-based results for the HLA locus were not followed up systematically (step 1), the best three variants from the sliding window analysis were genotyped in the discovery and independent replication samples to confirm the initial pooling-based results (step 2). As expected, all three variants were confirmed on the individual genotyping level and reached the genome-wide significant *P*-values. This was also the case in the independent replication step. This indicates that the DNA pooling approach can reliably detect SNPs that have shown genome-wide significance in association studies. Furthermore, pooling detects highly significant results, and it is therefore very unlikely that any genes beyond the HLA region are more significant. However, our strategy carries a risk of false-negative findings in the case of smaller genetic effects. DNA pooling adds extra experimental error (eg, pipetting for pool construction) to the allele frequency measurement that directly influences the power to detect small effect sizes.²⁶ Furthermore, only 61 top hits from the GWA study step were pursued in individual samples, with the great majority of nominally significantly associated markers having been excluded from the subsequent analyses. These are the two most likely explanations as to why the present study may have missed previously reported association findings.¹² Therefore, the reliable detection of genes with smaller effects requires larger sample sizes and individual genotyping.

The only other SNP to reach experiment-wide significance in the combined analysis was rs304650 in the *SPATA5* gene on chromosome 4q27–q28. Joint analysis of the replication samples used in steps 2 and 3, which included a total of 985 cases and 2014 controls, who were successfully genotyped for this SNP, revealed a significant association between this variant and AA, with a *P*-value of 3.43 × 10⁻⁴ (OR=1.24 (1.10–1.39)). A joint analysis of all of the investigated samples revealed a significant association between this variant and AA, with a *P*-value of 1.58 × 10⁻⁵ (OR=1.23 (1.12–1.35)). The SNP rs304650 maps to an intronic region of the *SPATA5* transcript. At the time of writing, the functional aspects of this protein are unknown. However, one study identified *SPATA2*, another member of the spermatogenesis-associated protein family, as a susceptibility gene for the autoimmune disorder psoriasis.²⁷

Interestingly, although *SPATA5* was not among the eight loci with genome-wide significance reported by Petukhova *et al*,¹² the authors reported six SNPs in *SPATA5*, with a *P*-value < 10⁻⁴ in their Supplementary Material. Although not very strong, there is evidence for LD between rs304650 and the Petukova *et al*¹² SNPs (ranging between *r*²=0.29 for rs11735364 and *r*²=0.46 for rs2201997). Although this might be viewed as supportive evidence, a more detailed workup of the region in very large samples is required to allow more definitive conclusions to be drawn. It is also interesting that the *SPATA5* gene is located only 320 kb distal to rs7682241, a genome-wide significant marker in the study of Petukhova *et al*,¹²

which strongly suggests the involvement of the *IL2/IL21* gene locus. Data from the CEU HapMap sample, however, show that rs304650 is not in LD with the best variants of the *IL2/IL21* gene locus. Thus, the two loci probably confer their risk independently of each other. It remains theoretically possible, however, that the true causal variant for AA may be a functional variant that is in moderate LD to the variants reported in both analyses, and which is located between the *IL2/IL21* and *SPATA5* regions.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank all patients and controls for participating in this study. The Heinz-Nixdorf Recall cohort (HNR), 1975 through 1989. The Heinz Nixdorf Foundation (Germany). Silke Redler is a past recipient of a BONFOR fellowship from the Medical Faculty of the University of Bonn. Markus M Nöthen is recipient of a grant from the Alfred Krupp von Bohlen und Halbach-Stiftung. Regina C Betz is a past recipient of an Emmy-Noether fellowship and is a recipient of a Heisenberg Professorship (both German Research Foundation (DFG)).

- 1 Safavi KH, Muller SA, Suman VJ, Moshell AN, Melton 3rd LJ: Incidence of alopecia areata in Olmsted County, Minnesota, 1975 through 1989. *Mayo Clin Proc* 1995; **70**: 628–633.
- 2 Blaumeiser B, van der Goot I, Fimmers R *et al*: Familial aggregation of alopecia areata. *J Am Acad Dermatol* 2006; **54**: 627–632.
- 3 Jackow C, Puffer N, Hordinsky M, Nelson J, Tarrand J, Duvic M: Alopecia areata and cytomegalovirus infection in twins: genes versus environment? *J Am Acad Dermatol* 1998; **38**: 418–425.
- 4 Tobin DJ: Characterization of hair follicle antigens targeted by the anti-hair follicle immune response. *J Invest Dermatol Symp Proc* 2003; **8**: 176–181.
- 5 Entz P, Blaumeiser B, Betz RC *et al*: Investigation of the HLA-DRB1 locus in alopecia areata. *Eur J Dermatol* 2006; **16**: 363–367.
- 6 Colombe BW, Lou CD, Price VH: The genetic basis of alopecia areata: HLA associations with patchy alopecia areata versus alopecia totalis and alopecia universalis. *J Invest Dermatol Symp Proc* 1999; **4**: 216–219.
- 7 Xiao FL, Yang S, Yan KL *et al*: Association of HLA class I alleles with alopecia areata in Chinese Hans. *J Dermatol Sci* 2006; **41**: 109–119.
- 8 de Andrade M, Jackow CM, Dahm N, Hordinsky M, Reveille JD, Duvic M: Alopecia areata in families: association with the HLA locus. *J Invest Dermatol Symp Proc* 1999; **4**: 220–223.
- 9 Betz RC, König K, Flaquer A *et al*: The R620W polymorphism in PTPN22 confers general susceptibility for the development of alopecia areata. *Br J Dermatol* 2008; **158**: 389–391.
- 10 Kemp EH, McDonagh AJ, Wengraf DA *et al*: The non-synonymous C1858T substitution in the PTPN22 gene is associated with susceptibility to the severe forms of alopecia areata. *Hum Immunol* 2006; **67**: 535–539.
- 11 Martínez-Mir A, Zlotogorski A, Gordon D *et al*: Genomewide scan for linkage reveals evidence of several susceptibility loci for alopecia areata. *Am J Hum Genet* 2007; **80**: 316–328.
- 12 Petukhova L, Duvic M, Hordinsky M *et al*: Genome-wide association study in alopecia areata implicates both innate and adaptive immunity. *Nature* 2010; **466**: 113–117.
- 13 Krawczak M, Nikolaus S, von Eberstein H, Croucher PJ, El Mokhtari NE, Schreiber S: PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genet* 2006; **9**: 55–61.
- 14 Wichmann HE, Gieger C, Illig T: KORA-gen—resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* 2005; **67** (Suppl 1): S26–S30.
- 15 Schmermund A, Mohlenkamp S, Stang A *et al*: Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: rationale and design of the Heinz Nixdorf RECALL Study. Risk Factors, Evaluation of Coronary Calcium and Lifestyle. *Am Heart J* 2002; **144**: 212–218.
- 16 Pappasotiropoulos A, Stephan DA, Huentelman MJ *et al*: Common Kibra alleles are associated with human memory performance. *Science* 2006; **314**: 475–478.
- 17 Moskvina V, Norton N, Williams N, Holmans P, Owen M, O'Donovan M: Streamlined analysis of pooled genotype data in SNP-based association studies. *Genet Epidemiol* 2005; **28**: 273–282.
- 18 Abraham R, Moskvina V, Sims R *et al*: A genome-wide association study for late-onset Alzheimer's disease using DNA pooling. *BMC Med Genomics* 2008; **1**: 44.
- 19 Moskvina V: Signal detection in genome-wide association studies. In: Ermakov SM, Melas VB, Pepelyshev AN (eds). *Proceedings of the 6th St Petersburg Workshop on Simulation*; June 28–July 4 2009; St Petersburg: VVM com. Ltd., 2009, pp 301–307.
- 20 Herold C, Becker T: Genetic association analysis with FAMHAP: a major program update. *Bioinformatics* 2009; **25**: 134–136.
- 21 Armitage P: Tests for linear trends in proportions and frequencies. *Biometrics* 1955; **11**: 375–386.
- 22 Consortium. TWTC: Genome-wide association study of 14 000 cases of seven common diseases and 3000 shared controls. *Nature* 2007; **447**: 661–678.
- 23 Barahmani N, de Andrade M, Slusser JP *et al*: Human leukocyte antigen class II alleles are associated with risk of alopecia areata. *J Invest Dermatol* 2008; **128**: 240–243.
- 24 Colombe BW, Price VH, Khoury EL, Garovoy MR, Lou CD: HLA class II antigen associations help to define two types of alopecia areata. *J Am Acad Dermatol* 1995; **33**: 757–764.
- 25 Pearson JV, Huentelman MJ, Halperin RF *et al*: Identification of the genetic basis for complex disorders by use of pooling-based genomewide single-nucleotide-polymorphism association studies. *Am J Hum Genet* 2007; **80**: 126–139.
- 26 Bosse Y, Bacot F, Montpetit A *et al*: Identification of susceptibility genes for complex diseases using pooling-based genome-wide association scans. *Hum Genet* 2009; **125**: 305–318.
- 27 Capon F, Bijlmakers MJ, Wolf N *et al*: Identification of ZNF313/RNF114 as a novel psoriasis susceptibility gene. *Hum Mol Genet* 2008; **17**: 1938–1945.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)