Genome-wide meta-analysis in alopecia areata resolves HLA associations and reveals two new susceptibility loci

Regina C. Betz, Lynn Petukhova, Stephan Ripke, Hailiang Huang, Androniki Menelaou, Silke Redler, Tim Becker, Stefanie Heilmann, Tarek Yamany, Madeliene Duvic, Maria Hordinsky, David Norris, Vera H. Price, Julian Mackay-Wiggan, Annemieke de Jong, Gina M. De Stefano, Susanne Moebus, Markus Böhm, Ulrike Blume-Peytavi, Hans Wolff, Gerhard Lutz, Roland Kruse, Li Bian, Christopher I. Amos, Annette Lee, Peter K. Gregersen, Bettina Blaumeiser, David Altschuler, Raphael Clynes, Paul I.W. de Bakker, Markus M. Nöthen, Mark J. Daly & Angela M. Christiano

Alopecia areata (AA) is a prevalent autoimmune disease with 10 known susceptibility loci. Here we perform the first meta-analysis of research on AA by combining data from two genome-wide association studies (GWAS), and replication with supplemented ImmunoChip data for a total of 3,253 cases and 7,543 controls. The strongest region of association is the major histocompatibility complex, where we fine-map four independent effects, all implicating human leukocyte antigen-DR as a key aetiologic driver. Outside the major histocompatibility complex, we identify two novel loci that exceed the threshold of statistical significance, containing ACOXL/BCL2L11(BIM) (2q13); GARP (LRRC32) (11q13.5), as well as a third nominally significant region SH2B3/LNK/ATXN2 (12q24.12). Candidate susceptibility gene expression analysis in these regions demonstrates expression in relevant immune cells and the hair follicle. We integrate our results with data from seven other autoimmune diseases and provide insight into the alignment of AA within these disorders. Our findings uncover new molecular pathways disrupted in AA, including autophagy/apoptosis, transforming growth factor beta/Tregs and JAK kinase signalling, and support the causal role of aberrant immune processes in AA.

1 Institute of Human Genetics, University of Bonn, Bonn D-53127, Germany. 2 Department of Dermatology, Columbia University, New York, New York 10032, USA. 3 Department of Epidemiology, Columbia University, New York, New York 10032, USA. 4 Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA. 5 Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02114, USA. 6 Department of Medical Genetics, University Medical Center Utrecht, Utrecht 3584 CG, The Netherlands. 7 German Center for Neurodegenerative Diseases, Bonn D-53175, Germany. 8 Institute for Medical Biometry, Informatics and Epidemiology, University of Bonn, Bonn D-53127, Germany. 9 Department of Genomics, Life and Brain Center, University Bonn, Bonn D-53127, Germany. 10 Department of Dermatology, MD Anderson Cancer Center, Houston, Texas 77030, USA. 11 Department of Dermatology, University of Minnesota, Minneapolis, Minnesota 55455, USA. 12 Department of Dermatology, University of Colorado, Denver, Colorado 80204, USA. 13 Department of Dermatology, University of California, San Francisco, San Francisco, California 94110, USA. 14 Department of Genetics and Development, Columbia University, New York, New York 10032, USA. 15 Institute of Medical Biometrics, Biometry, and Epidemiology, University Duisburg-Essen, Essen D-45122, Germany. 16 Department of Dermatology, University of Münster, Münster D-48149, Germany. 17 Department of Dermatology and Allergy, Clinical Research Center for Hair and Skin Science, Charité-Universitätsmedizin Berlin, Berlin D-10117, Germany. 18 Department of Dermatology, University of Munich, Munich D-80337, Germany. 19 Dermatological Practice, Hair and Nail, Wesseling D-50389, Germany. 20 Dermatological Practice, Paderborn D-33098, Germany. 21 Community and Family Medicine and Genetics, Dartmouth College, Hanover, New Hampshire 03755, USA. 22 The Feinstein Institute for Medical Research, Manhasset, New York 11030, USA. 23 Department of Medical Genetics, University of Antwerp, Antwerp BE-2650, Belgium. 24 Department of Medicine, Columbia University, New York, New York 10032, USA. 25 Department of Epidemiology, University Medical Center Utrecht, Utrecht 3584 CG, The Netherlands. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to A.M.C. (email: amc65@columbia.edu).
alopecia areata (AA) is one of the most prevalent autoimmune diseases, with a lifetime risk of 1.7% (ref. 1), and is the most common cause of hair loss in children. In AA, aberrant immune destruction is targeted to the hair follicle, resulting in non-scarring hair loss that typically begins as patches, which can increase in size and coalesce and may progress to cover the entire scalp (alopecia totalis), and body as well (alopecia universalis). Disease prognosis is unpredictable and highly variable. Its aetiologic basis has remained largely undefined, creating barriers to the development of effective therapeutic strategies and resulting in an enormous unmet medical need.

Our first GWAS in AA identified associations in eight regions of the genome, which were subsequently confirmed in independent candidate gene studies4–7. Associated loci outside the human leukocyte antigen (HLA) highlight particular immune response pathways and also implicate genes expressed in the hair follicle. For example, several regions contain genes with Treg functions, including IL2RA, IL2/IL21, CTLA4 and Eos, whereas ULBP3/ULBP6 implicate NKG2D-mediated cytotoxic T cells. Within the hair follicle, expression of STX17 suggests a role for end-organ autophagy, while PRDX5 implicates oxidative stress. A combined analysis of this GWAS and a subsequent replication study led to the identification of IL13/IL4 and KIAA0350/CLEC16A as new gene loci3.

Here we perform a meta-analysis to expand our sample size, and identify two new loci that exceed our threshold for genome-wide significance and a third locus that is nominally significant. We identify transcripts and/or protein for candidate genes at all three loci in disease-relevant tissues. We perform imputation and fine-mapping of the HLA, identifying four independent associations that implicate HLA-DRB1. Finally, a cross phenotype meta-analysis (CPMA) of our data with published results from seven other autoimmune diseases identify molecular pathways shared by AA and one or more other disorders.

Results

In this study, we have increased our cohort size and performed a combined analysis of two GWAS using Illumina Human660W- and Omni1-Quad BeadChips, analysing a total of 2,489 cases and 5,287 controls ascertainment in the US and Central Europe (Supplementary Table 1). Association analyses are performed with logistic regression. In a meta-analysis of these data, nine of the previously implicated regions exceeded statistical significance ($P < 5 \times 10^{-8}$), with STX17 achieving nominal significance (rs10124366; $P = 1.09 \times 10^{-5}$) (Fig. 1; Supplementary Data 1).

First, to resolve the major histocompatibility complex (MHC) association signal ($P = 4.91 \times 10^{-58}$ for the best single-nucleotide polymorphism (SNP), rs9275516), we used a published imputation and analysis protocol to perform fine-mapping (Supplementary Data 2)8. Conditional analysis revealed four independent variants located at the classical HLA-DRB1 and HLA-DRB1 genes. The most significant variant was amino-acid position 37 in HLA-DRB1 (omnibus $P$ value = $4.99 \times 10^{-73}$). Of the five possible amino acids at this position, Leu (odds ratio (OR) = 1.56), Tyr (OR = 1.54) and Phe (OR = 1.19) conferred a higher risk of AA, whereas the other residues conferred lower risk (OR for Asn = 0.42; OR for Ser = 0.74). Adjusting for the effects of HLA-DRB1 amino-acid position 37, we found an independent association due to an intronic SNP of HLA-DRA, rs9268657 (OR = 0.63, $P$ value = $4.84E^{-41}$). Functional annotation of this SNP and its close proxies ($R^2 > 0.9$) reveal that they influence expression levels of HLA-DRB1 (ref. 9). Adjusting for both HLA-DRB1 amino-acid position 37 and rs9268657, we identified another independent association for the classical allele HLA-DRB1*04:01 (OR = 1.64, $P$ value = $1.76 \times 10^{-12}$), confirming previous associations in candidate gene studies10,11. Finally, adjusting for HLA-DRB1 amino-acid position 37, rs9268657 and HLA-DRB1*04:01, amino-acid position 13 in HLA-DRB1 was also significant ($P$ value = $4.57 \times 10^{-16}$). Of the six possible alleles at this position, three increased risk (Tyr, OR = 1.41; Ser, OR = 1.35; Phe, OR = 1.09) and two are protective (His, OR = 0.57; Gly, OR = 0.50) and one demonstrated no effect (Arg, OR = 0.98) (Supplementary Table 2). Further rounds of conditional analyses yielded no additional significant results ($P > 2.11 \times 10^{-6}$).

Collectively, these four independent associations in the MHC implicate HLA-DR as the primary risk factor in AA, presumably through antigen presentation, similar to other immune-mediated diseases12. For example, HLA-DRB1 amino-acid positions 13 and 37 contribute to P4 and P9 peptide-binding pockets, respectively; the disease associations of HLA-DRB1*04:01 are thought to be driven by the shared epitope (amino-acid residues 70–74)12, which also occur within the peptide-binding cleft (Fig. 2). Polymorphic residues within peptide-binding pockets of HLA-DR influence binding affinities of peptides, and thus may shape the repertoire of autoantigens capable of triggering or perpetuating disease.

Next, we performed replication of SNPs in functionally relevant loci that achieved suggestive evidence for association in the meta-analysis ($5 \times 10^{-8} > P > 1 \times 10^{-5}$), utilizing an independent cohort of Central European ancestry. The Immunochip was used to genotype 318 cases and 1,688 controls and a Sequenom assay was used to genotype 85 SNPs not included on the Immunochip in a cohort of 764 cases and 568 controls. This analysis identified statistically significant associations in two novel genomic regions: chromosome 2q13 containing ACOXL and BCL2L11 (rs3789129, $P = 1.51 \times 10^{-8}$, OR$_R = 1.3$), and chromosome 11q13.5 containing C11orf30 and LRCC32 (rs2155219, $P = 1.25 \times 10^{-8}$, OR$_F = 1.2$) (Table 1; Supplementary Data 3).

The association signal at chromosome 2q13 is located within an intron of ACOXL, and the region of association extends to include BCL2L11 (Fig. 3). This region has been implicated in GWAS for two other autoimmune diseases: immunoglobulin-A nephropathy and primary sclerosing cholangitis13. ACOXL belongs to the acyl-coenzyme A oxidase gene family. While other family members play well-studied roles in peroxisomal beta-oxidation, very little is known about the function of this
BIM is widely expressed in multiple cell types, including immune-related, epithelial and hair follicle cells. We first performed reverse transcription (RT)–PCR analysis of BIM expression in immune cells and scalp hair follicles (SHFs). We detected BIM transcript in whole-peripheral blood mononuclear cells (PBMCs), including T, natural killer and B cells, and monocytes (MCs) (Fig. 4a). In plucked human SHFs, we observed expression of an alternative splice isoform, BIM-S, the most pro-apoptotic variant (Fig. 4a)\(^{20,21}\). To characterize protein localization within hair follicles, we performed immunofluorescence staining on both human and mouse hair follicles. In the human hair follicle, BIM is highly expressed in the bulb of the hair follicle, especially within the matrix cells, in a pattern strikingly restricted to early catagen, but not anagen or telogen (Fig. 4b). The hair bulb matrix is the principal location of differentiated, pigment-producing melanocytes, which are postulated to undergo apoptosis during catagen\(^{22}\). In mouse hair follicles, BIM localized to the apoptosing strand and lower portion of the catagen hair follicle, consistent with its expression in human hair. Remarkably, BIM expression was restricted to catagen hair follicles, where immunofluorescence staining performed on anagen or telogen hairs produced no signal (Fig. 4, upper panel). Finally, to investigate a possible role of BIM in AA pathogenesis, we used a mouse model that recapitulates genetic and molecular profiles of human AA\(^{23}\), C3H/HeJ spontaneously develops AA, and we performed immunofluorescence staining on affected and unaffected skin. We observed a striking and widespread increase in BIM expression throughout affected skin and hair follicles, which was not restricted to catagen hair follicles and appeared further upregulated in lesional skin (Fig. 4c, lower panel). The regression stage of a normal cycling hair follicle, catagen, is characterized by apoptosis and cell death. We postulate that dysregulation of BIM in hair-matrix keratinocytes and/or hair follicle melanocytes contributes to the early entry into dystrophic catagen in AA hair follicles in active disease.

The second novel region that exceeded statistical significance at chromosome 11q13.5 (rs2155219, \(P = 1.25 \times 10^{-8}\)) contains three gene transcripts, including C11orf30, GUCY2E and GARP (LRRC32). GWAS have implicated this region in several autoimmune and inflammatory diseases (www.genome.gov/gwas-studies). GARP is expressed on activated Tregs and T and B cells, and augmenting transforming growth factor beta bioavailability\(^{23–26}\), functioning to induce FOXP3 expression and Treg differentiation. Expression of GARP on activated human Treg cells correlates with...
their increased suppressive activity, and GARP knockdown reduces their suppressive activity\textsuperscript{24,27}. Not surprisingly, we found that GARP is highly expressed in whole PBMCs, consistent with previous reports. Unexpectedly, we also detected a strong signal for GARP in plucked control SHFs, a site not previously described and perhaps atypical for the expression of a Treg protein (Fig. 4). This suggests a potential role for GARP in hair biology and implicates significance of its disruption in AA at the site of pathology for the phenotype.

A third novel region on chromosome 12q24.12 that achieved suggestive evidence for association ($P = 1.3 \times 10^{-7}$) is of interest, because several genes harboured within the region functionally align with other associated genes. This region has demonstrated associations with multiple autoimmune diseases, and contains 10 genes including SH2B adaptor protein 3 ($SH2B3$) and aldehyde dehydrogenase 2 family ($ALDH2$). $SH2B3$, also known as LNK, is a key negative regulator of cytokine signalling via receptor tyrosine kinases and JAK signalling. Two missense variants that confer increased cytokine production and enhanced signalling have been reported in the literature (rs3184504 and rs72650673)\textsuperscript{28,29}. We genotyped these two functional polymorphisms in a sample of 96 chromosomes from AA patients carrying the AA-associated risk allele (rs653178*G), and found that 93 chromosomes also carried the rs3184504*$T$ risk allele ($f = 0.97$). This increase in allele frequency over what would be expected among a sample of European Americans ($f = 0.51$ in Exome Variant Server ESP6500) indicates that there is strong linkage disequilibrium (LD) between the AA-associated risk variant and this functional polymorphism, and suggests that R262W could be contributing to AA pathogenesis. The rs72650673*$A$ allele was not found in our sample, consistent with its low allele frequency among European Americans ($f = 0.002$ in Exome Variant Server ESP6500). We found by RT–PCR analysis that LNK is highly expressed in whole PBMCs, as well as T cells, natural killer, B cells and MCs, but did not detect expression in plucked SHFs (Fig. 4). $ALDH2$ has been identified as a citrullinated antigen in rheumatoid arthritis patients\textsuperscript{30}. This gene has been studied in the skin, with expression localizing to epidermis, sebaceous glands and hair follicles, where it is hypothesized to reduce the accumulation of oxidative stress-induced aldehydes\textsuperscript{31}.

The strong association with HLA class II in AA points towards the involvement of CD4+ T cells in the pathogenesis.

Figure 3 | Detailed map of associated SNPs and gene locations for newly identified loci. Two regions in the genome exceeded statistical significance when the replication data were combined with the meta-analysis results ($N = 10,796$) and analysed with logistic regression. (a) Chromosome 2q13 includes ACOXL and BCL2L11 (BIM). (b) Chromosome 12q24.12 included C11orf30 and LRCC32 (GARP).
The importance of this T-cell subset is supported by the presence of perifollicular CD4+ T cells in addition to intrabulbar CD8+ T cells in human AA. Finally, we analysed these data in the context of other autoimmune diseases by performing CPMA using the 107 SNPs used in the original description of the analysis and data from seven other autoimmune diseases32. We found 50 SNPs that are associated at or near significance across two or more diseases, and clustering of these according to their association with disease identified five groups (Fig. 5a,b). Protein–protein interaction graphs of each SNP cluster demonstrate that for several of the groups, the proteins coded from these regions interact either directly or via an intermediary to a significant degree (Fig. 5c). Integration of AA into the CPMA demonstrates mechanistic alignment with coeliac disease, type 1 diabetes, rheumatoid arthritis, multiple sclerosis and Crohn’s disease, with most of the overlap coming from the first cluster of genes.

Collectively, this meta-analysis for the first time resolves HLA associations and demonstrates the pivotal aetiological role of HLA-DR. Furthermore, the identification of specific residues in HLA-DRβ1 that are over-represented among AA patients will allow us to better model peptide class II MHC interactions and thus predict autoantigens in AA. Associations outside HLA provide new evidence for the importance of Treg maintenance and immune response pathways in AA. Furthermore, autophagy and apoptosis are emerging as processes of aetiologic importance in AA. These insights will allow us to better understand the molecular taxonomy of autoimmune diseases and the alignment of AA within this class of disorders. Importantly, as GWAS help to resolve disease mechanisms and identify pathogenic pathways perturbed in AA and autoimmunity in general, these approaches advance the field towards precision medicine in autoimmunity.

Methods

Patient population. All participating studies were reviewed and approved by the Institutional Review Boards and ethics committees at Columbia University; MD Anderson Cancer Center; University of Minnesota, Minneapolis; University of Colorado, Denver; University of California, San Francisco; and the Universities of Bonn, Düsseldorf, Münster, Berlin, Hamburg and Munich, Germany; and Antwerp, Belgium, and were conducted in accordance with the Declaration of Helsinki Principles. All study subjects provided written informed consent.

For the previously published GWAS and the ImmunoChip samples, cases were ascertained through the National Alopecia Areata Registry, which recruits patients in...
the US through five clinical sites and confirms diagnosis. Control data for these samples were obtained from publically available data. There were two sources of samples: the New York Cancer Project33 and genotyped as part of previous studies34. Second, a data set was obtained from the Cancer Genetic Markers of Susceptibility (CGEMS) breast35 and prostate36 cancer studies (http://cgems.cancer.gov/data/). The ImmunoChip control data were obtained from the NIDDK inflammatory bowel disease consortium (http://medicine.yale.edu/intmed/ibdgc/index.aspx). The discovery QC and imputation
90. Additional genotyping was performed by standard PCR-based techniques. Metrics were employed, removing data for SNPs and samples with call rates less than 0.2 (Fhet, statistic, standardized values in plink), (iv) missing rate per SNP varied between 420,000 and 640,000. On average, the QC processes excluded 44 common to all platforms and successfully genotyped in each GWAS sample was extracted. These SNPs were then further pruned to remove LD (leaving no pairs extracted). These SNPs were then further pruned to remove LD (leaving no pairs
enriched. The varying pattern of disease association for each cluster suggests each group represents a distinct co-morbid mechanism. (c) Proteins coded within a 100-Mb window centred on each SNP within each cluster are depicted in protein–protein interaction maps. Three of the five clusters have significant protein inter-connectivity (permuted P < 0.05).

Discovery QC and imputation. Technical QC was performed with QC conducted on each data set separately using a common approach. The following QC parameters were applied: (i) missing rate per SNP < 0.05 (before sample removal below), (ii) missing rate per individual < 0.02, (iii) heterozygosity per individual ± 0.2 (Fhet, statistic, standardized values in plink), (iv) missing rate per SNP < 0.02 (after sample removal above), (iv) missing rate per SNP difference in cases and controls <0.02, (vi) Hardy–Weinberg equilibrium (in controls) P < 10−6, (vii) Hardy–Weinberg equilibrium (in cases) P < 10−10. Study sample sizes varied between 1,200 and 3,000 individuals. The number of SNPs per study after QC varied between 420,000 and 640,000. On average, the QC processes excluded 44 individuals per study (with a range of 13–107 individuals) and 13,000 SNPs per study (with a range of 3,000–20,000 SNPs). After QC, the GWAS data sets together comprised 2,489 cases and 5,287 controls and, for the next steps of the ‘genetic QC’ analysis, a set of 221,784 SNPs (with a range of 3,000–20,000 individuals per study (with a range of

Genotyping. DNA was extracted from peripheral blood leukocytes by salting out with saturated NaCl solution according to standard methods, or by using a Chemagic Magnetic Separation Module I (Chemagen, Baesweiler, Germany), in accordance with the manufacturer’s instructions. Whole-genome genotyping for the meta-analysis was performed on either Illumina HumanHap550 BeadChip or the Illumina Omni express, as detailed in Supplementary Table 1. The replication cohort was genotyped with the Illumina Immunochip. In addition, SNPs selected for replication, which were not present on the Immunochip, were genotyped on the MassArray system using a Sequenom Compact MALDI-TOF device and iPLEX Gold reagents (Sequenom, San Diego, CA) in multiplex reactions. Primer sequences and standard assay conditions are available upon request. Standard quality-control (QC) metrics were employed, removing data for SNPs and samples with call rates less than 90. Additional genotyping was performed by standard PCR-based techniques.
Association analysis. Downstream analysis as associated covariates. We decided to include principal components 1, 2, 3, 4, 5 and 6 for downstream analysis as associated covariates.

Discovery principal component analysis. Principal component estimation was done with the same collection of SNPs on the non-related subset of individuals. We estimated the first 20 principal components and tested each of them for phenotypic association (using logistic regression with study indicator variables included as covariates) and evaluated their impact on the genome-wide test statistics using Lambda-GC (the genomic control inflation factor based on the median $\chi^2$) after genome-wide association of the specified principal component. On the basis of this, we decided to include principal components 1, 2, 3, 4, 5 and 6 for downstream analysis as associated covariates.

Cross-phenotype meta-analysis. The protocol for this analysis was adopted from previously described work on cross-phenotype meta-analysis. Using Dataset S1 from the Supplementary Data of the original analysis, we added in AA association for the 107 SNPs originally investigated. For SNPs with no $P$ value, we used Broad Institutes SNAP to find a proxy SNP with $r^2 > 0.9$. (http://www.broadinstitute.org/mpg/snap/). The CPMA statistic was then recalculated for all 107 SNPs using code (found here: http://csg.sph.umich.edu/imputation/ software.html) implementing the original description. All SNPs meeting a significance threshold of less than 0.01 were included in clustering analysis. Clustering was done by first binning $P$ values into four groups based on magnitude and clustering using the Cluster package. Cumulative association statistics were calculated in R using Fisher’s omnibus test. Protein–protein interaction analysis was performed using DAPPLE, a publicly available web application (http://www.broadinstitute.org/mpg/dapple/). Settings for DAPPLE included the 1000 genome assembly, 5,000 permutations and 50-kb regulatory region upstream and downstream. All statistical analyses were performed in R version 3.0.1.

References


Acknowledgements

We are deeply indebted to the many patients and their family members who participated in this study. We are thankful to Drs Lawrence Shapiro and Tatyana Gindin for guidance on the structural effects of HLA associations. We thank Spandon V. Shah, Holly Jiang, Matthew Ding and Steven Chiu for their assistance and contributions in isolating the immune cells by FACS and Jane E. Cerise for her bioinformatic expertise. We are most grateful for the support of the National Alopecia Areata Foundation for funding the initial studies, and Ms Vicki Kalabokes and her staff at NAAF for their efforts on our behalf. The US patient cohort was collected and maintained by the National Alopecia Areata Registry (N01-AR62279) (to M.D.). Some controls were drawn from the Heinz Nixdorf Recall Study cohort, which was established with the support of the Heinz Nixdorf Foundation. R.C.B. and M.M.N. are members, T.B. is an associate member, of the DFG-funded Excellence Cluster ImmunoSensation. R.C.B. is a recipient of a Heisenberg Professorship of the German Research Foundation (DFG). This work was supported in part by the DFG grant BE 2346/5-1, as well as by local funding (BONFOR) to R.C.B., Vernieuwingsimpuls VIDI Award from the Netherlands Organization for Scientific Research for project 016.126.354 to P.I.W.d.B., and USPHS NIH/NIAMS grants RO1AR52579 and RO1AR56016 (to A.M.C.).

Author contributions


Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications and all reporting requirements for granting agencies.

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/