A Large-Scale Genetic Analysis Reveals a Strong Contribution of the HLA Class II Region to Giant Cell Arteritis Susceptibility

David Carmona, F.; Mackie, Sarah L.; Martin, Jose-Ezequiel; Taylor, John C.; Vaglio, Augusto; Eyre, Stephen; Bossini-Castillo, Lara; Castaneda, Santos; Cid, Maria C.; Hernandez-Rodriguez, Jose

Published in:
American Journal of Human Genetics

DOI:
10.1016/j.ajhg.2015.02.009

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2015

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
We conducted a large-scale genetic analysis on giant cell arteritis (GCA), a polygenic immune-mediated vasculitis. A case-control cohort, comprising 1,651 case subjects with GCA and 15,306 unrelated control subjects from six different countries of European ancestry, was genotyped by the Immunochip array. We also imputed HLA data with a previously validated imputation method to perform a more comprehensive analysis of this genomic region. The strongest association signals were observed in the HLA region, with rs477515 representing the highest peak ($p = 4.05 \times 10^{-40}$, OR = 1.73). A multivariate model including class II amino acids of HLA-DRB1 and HLA-DQA1 and one class I amino acid of HLA-B explained most of the HLA association with GCA, consistent with previously reported associations of classical HLA alleles like HLA-DRB1*04. An omnibus test on polymorphic amino acid positions highlighted DRB1*13 ($p = 4.08 \times 10^{-44}$) and HLA-DQA1*04 ($p = 4.02 \times 10^{-46}$), 56, and 76 (both $p = 1.84 \times 10^{-45}$) as relevant positions for disease susceptibility. Outside the HLA region, the most significant loci included PTPN22 (rs2476601, $p = 1.73 \times 10^{-46}$, OR = 1.38), LRC32 (rs10160518, $p = 4.39 \times 10^{-6}$, OR = 1.20), and REL (rs11567447, $p = 1.10 \times 10^{-5}$, OR = 1.63). Our study provides evidence of a strong contribution of HLA class I and II molecules to susceptibility to GCA. In the non-HLA region, we confirmed a key role for the functional rs2476601 variant and proposed other putative risk loci for GCA involved in Th1, Th17, and Treg cell function.

Introduction

Giant cell arteritis (GCA [MIM 187360]) is a chronic and polygenic immune-mediated disease of unknown etiology that is the most common form of vasculitis in individuals over the age of 50 in Western countries. It is characterized by inflammatory damage of large- and medium-sized arteries, particularly the extracranial branches of the carotid artery, which can lead to severe complications such as blindness or cerebrovascular events.

During the last decade, genetic association studies have described several genes that are associated with predisposition to GCA, including genes of immune/inflammatory pathways and genes of the human leukocyte antigen (HLA) class I and II regions. The HLA-DRB1*04 alleles seem to be the most consistently associated genetic risk factors for this form of vasculitis.

One of the most successful platforms to identify immune-related risk variants is the Human Immuno DNA Analysis BeadChip Kit (known as the Immunochip), a custom Illumina Infinium High-Density array developed by the Immunochip Consortium for immunogenetics gene mapping. The Immunochip allows a dense analysis of 196,524 SNPs, rare variants, and insertion/deletion (indel) polymorphisms, located within 186 known susceptibility loci for autoimmune and inflammatory disorders. The use of the Immunochip has substantially increased the number of established genetic risk factors for multiple immune-mediated diseases, including Takayasu arteritis (another large-vessel vasculitis [MIM 207600]), celiac disease (MIM 212750), rheumatoid arthritis (RA [MIM 180300]), autoimmune thyroid disease (MIM 275000 and 140300), psoriasis (MIM 177900), primary biliary cirrhosis (MIM 109720), juvenile idiopathic arthritis (MIM 604302), primary sclerosing cholangitis (MIM 613806), narcolepsy (MIM 161400), ankylosing spondylitis (MIM 106300), atopic dermatitis (MIM 603165), and systemic sclerosis (SSc [MIM 181750]). The use of the same platform in all the above studies has facilitated the identification of common aetiological pathways among those disorders.

Considering the above, we decided to carry out a large-scale genetic analysis of GCA in a well-sized case-control cohort with the Immunochip genotyping platform. Additionally, taking advantage of the high coverage that this array has in the HLA region, we performed a comprehensive analysis of the HLA region by using a novel imputation method to obtain imputed types of SNPs, classical HLA alleles, and polymorphic amino acid positions.

Subjects and Methods

Study Population

Six independent case-control sample collections of European ancestry, from Spain (763 GCA-affected individuals and 1,517 unaffected controls), UK (251 GCA-affected individuals and 8,612 unaffected controls), North America (USA and Canada; 205 GCA-affected individuals and 1,641 unaffected controls), Italy (238 GCA-affected individuals and 1,270 unaffected controls), Norway (99 GCA-affected individuals and 374 unaffected controls), and Germany (95 GCA-affected individuals and 1,892 unaffected controls), were included in this study. The procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) of all participant centers, and written informed consent was obtained from all individuals. All cases fulfilled the 1990 American College of Rheumatology classification criteria for GCA, and the diagnosis was additionally confirmed by either a biopsy of the temporal artery (95.35%) or arterial imaging (4.65%). The most relevant clinical phenotypes of the case cohort are shown in Table S1.

Genotyping

Genomic DNA was extracted from blood samples by standard methods. The genotyping was performed on the Illumina iScan system with the Immunochip platform, as per Illumina protocols. Two different centers were involved in the genotyping. The sample sets from Spain, Italy, Norway, and Germany were genotyped by the Genomics and Genotyping Unit of the Pfizer-University of Granada-Junta de Andalucı´a Centre for Genomics and Oncological Research (GENYO, Granada, Spain) and those from UK and North America were genotyped by the Centre for Musculoskeletal Research (University of Manchester, Manchester, UK). Because of that, we used the same genotyping and calling procedures to control for possible batch effects as follows: (1) genotype calling was performed with the Genotyping Module (v.1.8.4) of the GenomeStudio Data Analysis software using the NCBI build 36 (hg18) mapping (Illumina manifest file Immununo_BeadChip_11419691_B.bpm); (2) low-quality SNPs
of every batch were removed if they had call rates < 0.98 or cluster separation < 0.4; and (3) the strand orientation of every batch was set accordingly with the TOP orientation of Illumina to avoid A/T and G/C mismatches during the merging of the sets.

Quality Control
Different quality filters were applied to the Immunochip raw data from each cohort independently prior to the statistical analyses by PLINK v.1.07.24 SNPs with a genotyping call rate lower than 98% and those that were not in Hardy-Weinberg equilibrium (HWE; \( p < 0.001 \)) were removed from the analysis. Similarly, subjects were excluded if fewer than 90% of SNPs were called. In addition, principal-component (PC) analyses were performed to identify and exclude outliers based on their ethnicity by PLINK and the gcta64 and R-base under GNU Public license v.2. With this software, we calculated and plotted the three first PCs using the “null” SNPs of the Immunochip (i.e., not associated with autoimmune diseases) for each subject, and those who deviated more than 4 SDs from the cluster centroids were considered outliers. PC analysis was also used to confirm that all subjects in our filtered study cohorts overlapped with the CEU cohort of the HapMap project (Figure S1). Furthermore, one subject per duplicate pair and per pair of first-degree relatives was also removed via the Genome function with a Ph-HAT threshold 0.5 in PLINK. A total of 1,651 GCA case subjects and 15,306 unrelated control subjects remained after applying the threshold 0.5 in PLINK. A total of 1,651 GCA case subjects and one subject per duplicate pair and per pair of first-degree relatives were also discarded.

Imputation of Non-HLA Regions
We performed SNP genotype imputation of the most associated loci outside the HLA region. For that, we isolated the genotyping data of 5 Mb regions centered in the lead SNPs and obtained imputed genotypes with the software IMPUTE v.2.29 The 1000 Genomes Phase 3 was used as reference panel.30 The probability threshold for merging genotypes was established at 0.9. Imputed data were subsequently subjected to stringent quality filters in PLINK, i.e., individuals who generated genotypes at <90% were removed from the datasets, and SNPs with call rates <98% and those that deviated from HWE in control subjects (\( p < 0.001 \)) were also discarded.

Statistical Analyses
CaTS Power Calculator for Genetic Studies, which implements the methods described in Skol et al.,31 was used to estimate the statistical power of our study (Table S3).

The statistical analyses were performed with PLINK and R. To test for association, we compared the variation frequencies of case and control subject by logistic regression on the best-guess genotypes (>0.9 probability) assuming an additive model with the three first PCs, the gender, and the country of origin as covariates. A minimal difference in the results was observed after adjusting for five or ten PCs. Inverse variance weighted fixed effects meta-analysis was also used to evaluate the consistency of the results of the non-HLA region. For the HLA region, we tested SNPs, classical HLA alleles, and all possible combinations of amino acid residues per position. Statistical significance was established by comparison of the deviance model to the null model as previously described.19 Cochran’s Q and I² tests were used to measure the heterogeneity of the ORs across studies.

The omnibus association test was also performed to determine the influence of the polymorphic amino acid positions in the disease susceptibility, as described.21 We established a null generalized linear model, which included the first three PCs, the country of origin, and gender as variables. Then, an alternative model was built for each position, including the previously described variables and all the possible alleles in the analyzed positions. Finally, both models were compared with a likelihood ratio test. This gives rise to a \( n – 1 \) degrees of freedom test, where \( n \) is the number of alternative alleles at that position. Additionally, we also conducted conditional analyses controlling the amino acid positions by the most associated positions with a putative functional effect in our set (conditioning factors). In these conditional analyses, the conditioning factors were included in both the null and the alternative models. By these means, we could test the association of all residues in the analyzed positions but including the conditioning factors as covariates in the models.

The haplotypes of the associated amino acid positions were also evaluated by additive logistic regression. Amino acid combinations with a frequency < 1% in control groups were excluded from this analysis.
To identify independent effects, dependency analyses by step-wise logistic regression were carried out with conditioning by the most associated signals, as described. The statistical significance was set at $p < 5 \times 10^{-8}$. Odds ratios (OR) and 95% confidence intervals (CI) were also calculated. The Manhattan plots were obtained with an in-house modification of the R script written by Stephen Turner, and the 3D models of the HLA molecules were performed with the UCSF Chimera software. The online tool of LocusZoom v.1.1 was used to plot the results of the imputed regions.

**Analysis of Enrichment of Rheumatoid-Arthritis-Associated Variants**

Considering that previous candidate gene studies and the results reported here (see below) suggested a possible overlap between part of the genetic background of RA and GCA, we carried out an analysis of enrichment of confirmed RA-associated alleles in our GCA cohort. With this aim, we calculated in PLINK a genetic score predicting RA risk for each participant in our study by obtaining a weighted mean of genotype dosage across 52 RA-associated variants. These include the 5 amino acids conferring higher risk in the amino acid model proposed by Raychaudhuri et al. and 47 non-HLA SNPs (22 exact ones and 25 proxies with $r^2 > 0.9$ in the CEU population of the 1000 Genomes project) out of the 101 reported in Okada et al. The remaining SNPs were not used to calculate the scores either because they were not included in our GCA dataset or because they had a reduced statistical significance in the European population ($p < 1 \times 10^{-6}$).

The effect estimates used to weight the scores were the log of the ORs reported for the European population in Table S1 of Okada et al. and the per-allele ORs (i.e., assuming an additive genetic model) reported in the Table S4 of Raychaudhuri et al. Table S4 summarizes the main aspects of the variants included in the analysis.

To analyze the possible difference between the score distribution in case and control subjects, a null generalized linear model including the first three PCs, country of origin, and gender as variables was compared against an alternative model including the same variables and the RA risk scores by the means of a likelihood ratio test in R.

**Results**

Association signals at the genome-wide level of significance were observed only in the combined analysis (Figure 1) and in the independent analyses of the cohorts with higher statistical power, i.e., Spain, UK, and North America (Figure S2). In all cases, those signals were located within the HLA region at chromosome 6 (Table S5).

**Deep Interrogation of the HLA Region**

To narrow down the HLA association with GCA, we inferred SNPs, classical HLA alleles, and polymorphic amino acid positions. The overall accuracy reached after comparing four-digit types with the corresponding imputed data was 99.48% for HLA-DQA1, 98.45% for HLA-DQB1, 89.94% for HLA-DRB1, and 88.11% for HLA-B (Table S6 and Figure S3).

After the imputation, high association peaks were observed in HLA-DRB1*04, HLA-DQA1*03, and HLA-DQB1*03 alleles (Table 1), particularly HLA-DRB1*04:04 ($p = 2.12 \times 10^{-23}$, OR = 2.28), HLA-DQA1*03:01 ($p = 1.38 \times 10^{-35}$, OR = 1.85), and HLA-DQB1*03:02 ($p = 3.93 \times 10^{-28}$, OR = 1.90). Consequently, different amino acids included in those alleles were also strongly associated with disease predisposition in the combined analysis (Table S5). However, the peak signal corresponded to a SNP (rs477515, $p = 4.05 \times 10^{-40}$, OR = 1.73) in high linkage disequilibrium (LD; $r^2 = 0.43$, $D' = 1.00$) with a group of SNPs and amino acids of HLA-DRB1 in complete LD with one another that have a stronger effect size (OR = 1.92), including a His in position 13 ($p = 5.12 \times 10^{-38}$, OR = 1.92) located in the binding groove of the molecule. This SNP maps to the intergenic region between...
Table 1. Genome-wide Significant Associations of Classical HLA Alleles with Giant Cell Arteritis without Conditioning at Two- and Four-Digit Resolution

<table>
<thead>
<tr>
<th>Classical HLA Allele</th>
<th>Spain GCA</th>
<th>Spain CTRL</th>
<th>UK GCA</th>
<th>UK CTRL</th>
<th>North America GCA</th>
<th>North America CTRL</th>
<th>Italy GCA</th>
<th>Italy CTRL</th>
<th>Norway GCA</th>
<th>Norway CTRL</th>
<th>Germany GCA</th>
<th>Germany CTRL</th>
<th>Meta-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRB1*04</td>
<td>0.2104</td>
<td>0.1276</td>
<td>0.3247</td>
<td>0.1944</td>
<td>0.2854</td>
<td>0.1651</td>
<td>0.1155</td>
<td>0.0736</td>
<td>0.3182</td>
<td>0.2233</td>
<td>0.2263</td>
<td>0.1453</td>
<td>6.78 × 10⁻³⁸</td>
</tr>
<tr>
<td>DQA1*03</td>
<td>0.2156</td>
<td>0.1381</td>
<td>0.3327</td>
<td>0.2069</td>
<td>0.3024</td>
<td>0.1758</td>
<td>0.1218</td>
<td>0.0776</td>
<td>0.3384</td>
<td>0.2326</td>
<td>0.2368</td>
<td>0.1564</td>
<td>1.38 × 10⁻³⁵</td>
</tr>
<tr>
<td>DQA1*01</td>
<td>0.3014</td>
<td>0.3767</td>
<td>0.2410</td>
<td>0.3885</td>
<td>0.2732</td>
<td>0.4007</td>
<td>0.2668</td>
<td>0.3622</td>
<td>0.2727</td>
<td>0.4412</td>
<td>0.3211</td>
<td>0.4382</td>
<td>3.84 × 10⁻²⁹</td>
</tr>
<tr>
<td>DQB1*03</td>
<td>0.3847</td>
<td>0.3174</td>
<td>0.4781</td>
<td>0.3434</td>
<td>0.4415</td>
<td>0.3501</td>
<td>0.4475</td>
<td>0.4193</td>
<td>0.4697</td>
<td>0.3262</td>
<td>0.4316</td>
<td>0.3380</td>
<td>5.15 × 10⁻¹⁵</td>
</tr>
</tbody>
</table>

Dependence Analysis

We then hypothesized that the HLA association with GCA might be explained by polymorphic amino acid residues, as reported for other immune-mediated diseases, and we carried out further logistic regression analysis of the HLA data controlling for the most associated amino acid residues. If the most associated signals were in high LD, we selected for those located in the binding pocket of their corresponding molecule.

A summary of the logistic regression analysis is shown in Figure 2. First, we included the above-mentioned HLA-DRβ1 His13 amino acid as covariate, which decreased considerably the statistical significance of the class II-associated signals (highest peak: rs3104407, p = 3.27 × 10⁻¹⁰, OR = 0.70, Figure 2B). Subsequently, the addition of the model of a glycine (Gly) in position 56 of the HLA-DQα1 chain explained the association at the genome-wide level of significance observed in the class II region, although some independent signals in class I (i.e., amino acids in positions 45 and 97 of HLA-B and the SNP rs3130944) still remained (highest peak: presence of Arg or Thr in position 97 of HLA-B, p = 1.36 × 10⁻⁸; Figure 2C). Finally, when a Thr in position 45 of the HLA-B molecule was included in the condition list, all the class I variants were non-significant (Figure 2D).

Hence, the model that best explained the HLA association with GCA included HLA-DRβ1 His13 (p = 5.12 × 10⁻³⁸, OR = 1.92), HLA-DQα1 Gly56 (p = 3.84 × 10⁻²⁹, OR = 0.61), and HLA-B Thr45 (p = 3.78 × 10⁻⁹, OR = 0.76) (Table S5). However, it should be noted that we based our analysis on the hypothesis that the leading drivers are most likely to be located in antigen-binding pockets of the HLA molecules and, therefore, other putative models (e.g., including the rs477515 genetic variant) might also explain our data equally well.

Omnibus Test

We also tested the possible influence of the polymorphic amino acid positions by means of an omnibus test (Table S7 and Figure S4). The most associated signals corresponded to the HLA-DQα1 positions 47, 56, and 76 (p = 4.02 × 10⁻⁴⁶, p = 1.84 × 10⁻⁴⁵, and p = 1.84 × 10⁻⁴⁵, respectively). Regarding the latter, DQα1 76 might harbor three amino acid residues (Val76, Met76, and Leu76) that are in complete LD with the three possible variants in DQα1 56 (Arg56, Gly56, and *56) and, therefore, they cannot be distinguished by statistical analysis alone. In relation to DQα1 47, four amino acids can be present at this position (Gln47, Arg47, Lys47, and Cys47). Two of them (Gln47 and Arg47) are in complete LD with amino acids of DQα1 56-76 and form two haplotypes, i.e., Gln47-Arg47-Lys47 (haplotype A, conferring risk) and Arg47-Gly56-Met76 (haplotype B, conferring protection) (Figure 3A, Table S8). The third and fourth possible DQα1 47 amino acids also form two other haplotypes with DQα1 *56-Leu (haplotype C, Lys47-*56-Leu76; and haplotype D, Cys47-*56-Leu76), but those
haplotypes did not reach statistical significance (Table S8). Position 13 of HLA-DRβ1 also showed high association (p = 4.08 × 10^{-10}). It was not possible to unambiguously establish casualty from HLA-DRβ1 positions 47, 56, and 76 due to the high LD between some of the amino acids at those positions (Table S8).

Considering the similarity in the association with RA susceptibility observed for the positions 11 and 13 of HLA-DRβ1, we conditioned these two positions one to another to analyze possible casualty in GCA. Contrary to that observed in RA, position 11 completely lost its statistical significance after conditioning on position 13 (conditioned p = 0.64), whereas a suggestive p value was observed for position 13 after conditioning on position 11 (conditioned p = 6.74 × 10^{-6}).

Table 2 summarizes the results of the associated amino acids within the positions comprising our proposed model to explain the HLA association with GCA.

Additionally, similar to that observed in the analysis of the specific amino acid residues, no statistically significant signals were detected after including the amino acids of HLA-DRβ1 13, HLA-DQα1 56, and HLA-B 45 as conditioning factors in the analyses (Figure S4).

Haplotype Analysis
Taking into account the high LD of the class II HLA genomic region, we decided to carry out a haplotype test combining all the possible amino acids present in the class II

---

Figure 2. Manhattan Plot Representation of the Step-wise Conditional Logistic Regression of the HLA Region
(A) Unconditioned test of the HLA region.
(B) Results of the HLA region after conditioning to HLA-DRβ1 His13.
(C) Results of the HLA region after conditioning to HLA-DRβ1 His13 and HLA-DQα1 Gly56.
(D) Results of the HLA region after conditioning to HLA-DRβ1 His13, HLA-DQα1 Gly56, and HLA-B Thr45.

The −log10 of the combined logistic regression test p values are plotted against their physical chromosomal position. A red/blue color gradient was used to represent the effect size of each analyzed variant (red for risk and blue for protection). The red line represents the genome-wide level of significance (p < 5 × 10^{-8}).
The classical HLA alleles in our dataset that contain those amino acids are shown. Asterisk (*) indicates amino acids involved in antigen binding.

A, p = 4.12 × 10⁻³⁷, OR = 1.90; DRβ1 His13 - DQx1 haplotype A - B Glu45, p = 9.09 × 10⁻²⁰, OR = 2.52; DRβ1 His13 - DQx1 haplotype A - B Met45, p = 1.03 × 10⁻¹⁸, OR = 2.78; DRβ1 His13 - DQx1 haplotype A - B Lys45, p = 3.05 × 10⁻¹⁴, OR = 1.86); and (3) haplotype B of DQx1 (Arg47-Gly56-Met76) was included in the haplotypes conferring higher protection (DRβ1 Ser13 - DQx1 haplotype B, p = 7.72 × 10⁻¹⁶, OR = 0.52; DRβ1 Ser13 - DQx1 haplotype B - Thr45, p = 9.16 × 10⁻⁸, OR = 0.39).
Analysis of the Non-HLA Region

No associations at the genome-wide significance level were detected when the non-HLA region was analyzed by logistic regression (Table 3). Two genetic variants of the protein tyrosine phosphatase non-receptor type 22 (PTPN22 [MIM 600716]) gene in almost complete LD ($r^2$ = 0.99) represented the highest non-HLA association signal with GCA (rs6679677, $p = 1.31 \times 10^{-6}$, OR = 1.39; rs2476601, $p = 1.73 \times 10^{-6}$, OR = 1.38). One of them (rs2476601) is a non-synonymous (p.Arg620Trp) functional variant that has been associated with a variety of immune-mediated diseases, including GCA and several autoimmune diseases, at a statistical significance ($p < 0.05$) between GCA and several autoimmune diseases.38 Although the signals were relatively weak, associations at the genome-wide significance level (Table S11 and Figure S5), e.g., nitric oxide synthase 2 inducible (NOS2 [MIM 163730]; rs2274894, $p = 1.58 \times 10^{-3}$, OR = 1.14), interleukin 6 (IL6 [MIM 147620]; rs10242595, $p = 3.70 \times 10^{-3}$, OR = 0.89), IL4 (MIM 147780); rs2243200, $p = 4.47 \times 10^{-3}$, OR = 6.86), interferon gamma (IFNG [MIM 147570]; rs2193046, $p = 5.08 \times 10^{-3}$, OR = 0.89), and IL10 (MIM 124092); rs74148796, $p = 7.84 \times 10^{-3}$, OR = 0.83).

In order to conduct a more detailed analysis of the top signals, we obtained imputed data of the PTPN22, LRRC32, and REL genomic regions. A total of 922 SNPs in the PTPN22 region, 462 in the LRRC32 region, and 1,158 in the REL region were included in the imputed datasets. However, because of the dense coverage of the fine-mapped loci in the Immunochip, all the imputed variants showed a lower statistical significance than PTPN22 rs2476601, LRRC32 rs10160518, and REL rs115674477 (Figure 4), consistent with previous studies.19

We checked the statistical significance of genes reported to be associated with GCA in candidate gene studies.5 Although the signals were relatively weak, associations at $p < 0.05$ were observed in most cases (Table S11 and Figure S5), e.g., nitric oxide synthase 2 inducible (NOS2 [MIM 163730]; rs2274894, $p = 1.58 \times 10^{-3}$, OR = 1.14), interleukin 6 (IL6 [MIM 147620]; rs10242595, $p = 3.70 \times 10^{-3}$, OR = 0.89), IL4 (MIM 147780); rs2243200, $p = 4.47 \times 10^{-3}$, OR = 6.86), interferon gamma (IFNG [MIM 147570]; rs2193046, $p = 5.08 \times 10^{-3}$, OR = 0.89), and IL10 (MIM 124092); rs74148796, $p = 7.84 \times 10^{-3}$, OR = 0.83).

Considering that a shared genetic component might influence the development of different autoimmune phenotypes,20,40 we also checked the results of specific known susceptibility SNPs for immune-mediated diseases in our dataset accordingly with the reported data in Table S5 of Ricaño-Ponce and Wijmenga.40 Associations at the nominal level ($p < 0.05$) between GCA and several autoimmune SNPs were observed (Table S12), with PTPN22 rs6679677/rs2476601 (associated with RA, type 1 diabetes [MIM 222100], and Crohn disease [MIM 266600]), and Crohn disease [MIM 266600], among others) and LRRC32 rs10160518/rs2155219 (associated with ulcerative colitis, UC [MIM 266600]) representing the strongest signals.
<table>
<thead>
<tr>
<th>Chr</th>
<th>rs ID</th>
<th>Position (GRCh36)</th>
<th>Locus</th>
<th>Minor/Major Allele</th>
<th>Spain GCA</th>
<th>CTRL</th>
<th>UK GCA</th>
<th>CTRL</th>
<th>North America GCA</th>
<th>CTRL</th>
<th>Italy GCA</th>
<th>CTRL</th>
<th>Norway GCA</th>
<th>CTRL</th>
<th>Germany GCA</th>
<th>CTRL</th>
<th>Meta-analysis</th>
<th>p Value</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>rs6679677</td>
<td>114105331</td>
<td>PTPN22</td>
<td>A/C</td>
<td>0.097</td>
<td>0.063</td>
<td>0.112</td>
<td>0.095</td>
<td>0.129</td>
<td>0.097</td>
<td>0.050</td>
<td>0.035</td>
<td>0.136</td>
<td>0.115</td>
<td>0.121</td>
<td>0.112</td>
<td>1.31 × 10^{-6}</td>
<td>1.39</td>
<td>(1.22–1.58)</td>
</tr>
<tr>
<td>1</td>
<td>rs2476601</td>
<td>114179091</td>
<td>PTPN22</td>
<td>A/G</td>
<td>0.098</td>
<td>0.064</td>
<td>0.112</td>
<td>0.095</td>
<td>0.129</td>
<td>0.098</td>
<td>0.050</td>
<td>0.036</td>
<td>0.131</td>
<td>0.115</td>
<td>0.121</td>
<td>0.111</td>
<td>1.73 × 10^{-6}</td>
<td>1.38</td>
<td>(1.21–1.58)</td>
</tr>
<tr>
<td>11</td>
<td>rs10160518</td>
<td>75974319</td>
<td>LRRC32</td>
<td>A/G</td>
<td>0.500</td>
<td>0.447</td>
<td>0.528</td>
<td>0.494</td>
<td>0.522</td>
<td>0.490</td>
<td>0.494</td>
<td>0.465</td>
<td>0.561</td>
<td>0.529</td>
<td>0.605</td>
<td>0.513</td>
<td>4.39 × 10^{-6}</td>
<td>1.20</td>
<td>(1.11–1.29)</td>
</tr>
<tr>
<td>11</td>
<td>rs2155219</td>
<td>75976842</td>
<td>LRRC32</td>
<td>C/A</td>
<td>0.499</td>
<td>0.446</td>
<td>0.528</td>
<td>0.494</td>
<td>0.520</td>
<td>0.490</td>
<td>0.494</td>
<td>0.465</td>
<td>0.561</td>
<td>0.531</td>
<td>0.605</td>
<td>0.514</td>
<td>6.19 × 10^{-6}</td>
<td>1.19</td>
<td>(1.11–1.29)</td>
</tr>
<tr>
<td>2</td>
<td>rs11567477</td>
<td>60965427</td>
<td>REL</td>
<td>G/A</td>
<td>0.039</td>
<td>0.029</td>
<td>0.032</td>
<td>0.023</td>
<td>0.039</td>
<td>0.016</td>
<td>0.034</td>
<td>0.017</td>
<td>0.046</td>
<td>0.036</td>
<td>0.032</td>
<td>0.019</td>
<td>1.10 × 10^{-5}</td>
<td>1.63</td>
<td>(1.31–2.03)</td>
</tr>
<tr>
<td>21</td>
<td>rs2236430</td>
<td>32895279</td>
<td>TCP10L-C21orf59</td>
<td>A/G</td>
<td>0.306</td>
<td>0.293</td>
<td>0.347</td>
<td>0.317</td>
<td>0.349</td>
<td>0.320</td>
<td>0.387</td>
<td>0.302</td>
<td>0.389</td>
<td>0.297</td>
<td>0.437</td>
<td>0.361</td>
<td>1.33 × 10^{-5}</td>
<td>1.20</td>
<td>(1.11–1.30)</td>
</tr>
<tr>
<td>11</td>
<td>rs2508740</td>
<td>75904830</td>
<td>C11orf30</td>
<td>G/A</td>
<td>0.375</td>
<td>0.425</td>
<td>0.383</td>
<td>0.396</td>
<td>0.359</td>
<td>0.395</td>
<td>0.366</td>
<td>0.399</td>
<td>0.364</td>
<td>0.436</td>
<td>0.321</td>
<td>0.388</td>
<td>2.06 × 10^{-5}</td>
<td>0.84</td>
<td>(0.78–0.91)</td>
</tr>
<tr>
<td>11</td>
<td>rs1892952</td>
<td>76042374</td>
<td>LRRC32</td>
<td>A/G</td>
<td>0.528</td>
<td>0.473</td>
<td>0.496</td>
<td>0.480</td>
<td>0.524</td>
<td>0.501</td>
<td>0.532</td>
<td>0.495</td>
<td>0.525</td>
<td>0.480</td>
<td>0.558</td>
<td>0.491</td>
<td>4.11 × 10^{-5}</td>
<td>1.17</td>
<td>(1.09–1.27)</td>
</tr>
<tr>
<td>6</td>
<td>rs11752919</td>
<td>28511582</td>
<td>ZSCAN23</td>
<td>G/A</td>
<td>0.347</td>
<td>0.407</td>
<td>0.363</td>
<td>0.409</td>
<td>0.363</td>
<td>0.359</td>
<td>0.263</td>
<td>0.280</td>
<td>0.313</td>
<td>0.365</td>
<td>0.332</td>
<td>0.339</td>
<td>5.22 × 10^{-5}</td>
<td>0.84</td>
<td>(0.78–0.92)</td>
</tr>
<tr>
<td>11</td>
<td>rs7926009</td>
<td>75845110</td>
<td>C11orf30</td>
<td>G/A</td>
<td>0.388</td>
<td>0.435</td>
<td>0.388</td>
<td>0.407</td>
<td>0.381</td>
<td>0.410</td>
<td>0.382</td>
<td>0.412</td>
<td>0.374</td>
<td>0.448</td>
<td>0.342</td>
<td>0.398</td>
<td>5.51 × 10^{-5}</td>
<td>0.85</td>
<td>(0.79–0.92)</td>
</tr>
<tr>
<td>11</td>
<td>rs10790958</td>
<td>127862477</td>
<td>ETS1</td>
<td>G/A</td>
<td>0.320</td>
<td>0.301</td>
<td>0.351</td>
<td>0.323</td>
<td>0.354</td>
<td>0.295</td>
<td>0.267</td>
<td>0.226</td>
<td>0.404</td>
<td>0.338</td>
<td>0.384</td>
<td>0.330</td>
<td>5.71 × 10^{-5}</td>
<td>1.19</td>
<td>(1.09–1.29)</td>
</tr>
<tr>
<td>10</td>
<td>rs587198</td>
<td>6571155</td>
<td>PRKCQ</td>
<td>G/A</td>
<td>0.447</td>
<td>0.401</td>
<td>0.484</td>
<td>0.453</td>
<td>0.459</td>
<td>0.451</td>
<td>0.490</td>
<td>0.461</td>
<td>0.525</td>
<td>0.433</td>
<td>0.516</td>
<td>0.454</td>
<td>5.72 × 10^{-5}</td>
<td>1.17</td>
<td>(1.09–1.27)</td>
</tr>
<tr>
<td>18</td>
<td>rs1788110</td>
<td>65719816</td>
<td>CD226</td>
<td>C/G</td>
<td>0.321</td>
<td>0.337</td>
<td>0.257</td>
<td>0.318</td>
<td>0.334</td>
<td>0.337</td>
<td>0.334</td>
<td>0.411</td>
<td>0.298</td>
<td>0.369</td>
<td>0.295</td>
<td>0.320</td>
<td>6.51 × 10^{-5}</td>
<td>0.85</td>
<td>(0.78–0.92)</td>
</tr>
<tr>
<td>10</td>
<td>rs10749570</td>
<td>90028013</td>
<td>C10orf59</td>
<td>A/C</td>
<td>0.277</td>
<td>0.343</td>
<td>0.309</td>
<td>0.335</td>
<td>0.327</td>
<td>0.344</td>
<td>0.319</td>
<td>0.324</td>
<td>0.301</td>
<td>0.366</td>
<td>0.358</td>
<td>0.360</td>
<td>7.27 × 10^{-5}</td>
<td>0.84</td>
<td>(0.78–0.92)</td>
</tr>
<tr>
<td>14</td>
<td>rs10149689</td>
<td>80485553</td>
<td>C14orfT45-TSHR</td>
<td>G/A</td>
<td>0.417</td>
<td>0.458</td>
<td>0.462</td>
<td>0.449</td>
<td>0.395</td>
<td>0.456</td>
<td>0.410</td>
<td>0.467</td>
<td>0.399</td>
<td>0.447</td>
<td>0.358</td>
<td>0.404</td>
<td>7.36 × 10^{-5}</td>
<td>0.85</td>
<td>(0.79–0.92)</td>
</tr>
<tr>
<td>11</td>
<td>rs3817637</td>
<td>275920</td>
<td>NLRC6-Athl1</td>
<td>A/G</td>
<td>0.091</td>
<td>0.124</td>
<td>0.084</td>
<td>0.114</td>
<td>0.122</td>
<td>0.115</td>
<td>0.090</td>
<td>0.134</td>
<td>0.152</td>
<td>0.128</td>
<td>0.090</td>
<td>0.121</td>
<td>8.67 × 10^{-5}</td>
<td>0.77</td>
<td>(0.68–0.88)</td>
</tr>
<tr>
<td>11</td>
<td>rs10160382</td>
<td>75804862</td>
<td>PKRIR-C11orf30</td>
<td>G/A</td>
<td>0.321</td>
<td>0.357</td>
<td>0.331</td>
<td>0.354</td>
<td>0.322</td>
<td>0.351</td>
<td>0.332</td>
<td>0.363</td>
<td>0.328</td>
<td>0.385</td>
<td>0.284</td>
<td>0.343</td>
<td>9.60 × 10^{-5}</td>
<td>0.85</td>
<td>(0.78–0.92)</td>
</tr>
</tbody>
</table>
Polygenic Risk Score Analysis
Taking into account that some of the most associated variants with GCA are confirmed RA risk factors, we calculated polygenic risk scores predictive for RA susceptibility in our samples and tested whether there were differences in the distribution of the scores between case and control subjects (Figure 5). Three different scores were constructed per sample: (1) with all the RA-associated variants, (2) including only the non-HLA SNPs, and (3) considering only the HLA amino acids conferring higher susceptibility to RA. In the three analyses, the mean score was higher in the GCA set that in control subjects. The most significant differences between case and control subjects were observed when the scores were calculated with all the RA variants (likelihood p value = 9.32 x 10^{-20}; Figures 5A and 5B). The difference of the score distribution between case and control subjects was also highly significant when only the five RA-risk amino acids were considered to calculate the scores (likelihood p value = 3.71 x 10^{-23}; Figures 5E and 5F). Although a less significant p value was obtained with the scores calculated with the non-HLA alleles, there was still some evidence of an effect (likelihood p value = 5.57 x 10^{-6}; Figures 5C and 5D). However, these results must be interpreted with caution, because they suggest only that RA and GCA case subjects are genetically more similar to each other than each is to healthy control subjects. Further analyses are needed to obtain a better estimation of the shared genetic component between this type of vasculitis and other immune-mediated diseases.

Discussion
This study was performed with a large GCA cohort of individuals from different European ancestries that cover the whole gradient in prevalence for populations of European descent. Our data show strong evidence for substantial involvement of HLA class II region in the pathophysiology of GCA. We confirmed the previous associations of GCA

Figure 4. Regional Plots of the Three Most Associated Loci with GCA outside the HLA Region in the Overall Meta-analysis after Imputation
(A) Protein tyrosine phosphatase non-receptor type 22 (PTPN22) region.
(B) v-rel avian reticuloendotheliosis viral oncogene homolog (REL) region.
(C) Leucine-rich repeat containing 32 (LRRC32) region.
(D) Results for the LRRC32 region after conditioning for the lead SNP (rs10160518).
Red arrows point to the lead variants (highlighted in violet).
with HLA-DRB1*04 alleles (both 04:01 and 04:04). We also identified HLA-DQA1 as an independent novel susceptibility factor for GCA, specifically the classical alleles DQA1*0101, DQA1*0102, and HLA-DQA1*03:01. The level of statistical significance observed within the HLA region in our study highlights the importance of the immune system in the development of GCA. Our results are consistent with the hypothesis of dysregulated interactions between the vessel wall and immune cells as responsible for the development of large-vessel vasculitides.

Novel imputation methods are allowing the identification of specific amino acid positions of HLA molecules associated with immune-mediated diseases, including RA, SSC, and Behçet disease (BD [MIM 109650]), thus increasing our understanding of the complex HLA associations with different autoimmune processes. With this state-of-the-art methodology, we built a model including class II amino acid positions (HLA-DRb1 13 and HLA-DQa1 47, 56, 76) and one class I amino acid position (HLA-B 45) that explained most of the HLA association with GCA in our study cohort. Some of these amino acid positions are located in the binding groove of their corresponding HLA molecules and are directly involved in the peptide binding (Figure 6).

According to our results, the presence of a Thr in position 45 of HLA-B might be protective of the development of GCA. Interestingly, classical HLA alleles containing Thr45, such as HLA-B*51:01, have been associated with a delay in the onset of AIDS (MIM 609423)-defining diseases and a better control of HIV infection. In addition, HLA-B*51 (particularly 51:01) has been identified as the strongest known genetic risk factor for BD, another type of vasculitis, although HLA-B 45 does not seem to be a relevant position for this association. However, no significant risk effects were observed for any of the other amino acids (Met, Glu, and Lys) of this position in our sample set, despite the strong influence that this position seems to have on the features of the bound peptides.

In contrast to the data on DQ, HLA-DRb1 His13 is included in the previously identified GCA-associated alleles HLA-DRB1*04:01 and HLA-DRB1*04:04. This amino acid is in tight LD with the top most associated position with RA (HLA-DRB1 Val11). In the study of RA, it was not possible to unambiguously assign causality to one position at the exclusion of the other, indicating that HLA-DRB1 His13 cannot be ruled out as the major contributor to the HLA association with RA. Indeed, the two more associated amino acids with RA and GCA risk at those positions (HLA-DRB1 His13 and HLA-DRB1 Val11) showed also a high LD in our dataset ($r^2 = 0.94$), and it could be possible that HLA-DRB1 11 represents a
causal position for GCA. However, the effect sizes observed for both the HLA-DRβ1 Val11 residue (p = 2.22 × 10^{-32}, OR = 1.80) and the HLA-DRβ1 11 position (p = 5.06 × 10^{-37}) were clearly lower than those observed for HLA-DRβ1 His13 (5.12 × 10^{-38}, OR = 1.92) and position 13 of HLA-DRβ1 (p = 4.08 × 10^{-43}) in our study. In addition, the effect of position 11 in GCA susceptibility was completely eliminated after conditioning on position 13 (p = 0.64), whereas a suggestive association with position 13 remained after conditioning on position 11 (p = 6.74 × 10^{-6}).

Remarkably, candidate gene studies in GCA suggested that this vasculitis shares some of its genetic component with RA. The data clearly support this hypothesis, because two out of the three top signals outside the HLA region in our Immunochip study, PTPN22 and REL, are important susceptibility factors for RA. In addition, PRR5Q, another RA risk gene of the REL pathway, is also among the most significant loci in our study. The above is consistent with the clear difference that we observed in the polygenic risk scores predictive for RA risk between GCA case and control subjects.

Furthermore, the autoimmune disease-associated PTPN22 SNP rs2476601/p.Arg620Trp, which is a central negative regulator of both B and T cell receptor signaling, was the top signal in the non-HLA analysis of GCA. There is controversy about the suitability of using the strict genome-wide significance threshold (5 × 10^{-8}) in Immunochip studies, because the design of this custom array was not based on SNPs tagging the whole genome but on the fine-mapping of immune genes. We estimated that the total number of independent signals in our filtered dataset was 23,791 (indep-pairwise option of PLINK with values 50 0.2) and, therefore, an appropriate Bonferroni-based statistical threshold for our study could be considered 2.10 × 10^{-6}. Similarly, with the Genetic type 1 Error Calculator (GEC) software, which implements a previously validated method to calculate the threshold for statistical significance required to control the genome-wide type 1 error rate at 0.05, 1.77 × 10^{-6} was obtained as significance threshold for our study. Hence, with this criterion, PTPN22 rs2476601 would represent a confirmed associated variant in GCA. This same genetic variant was recently identified by our group as a susceptibility factor for this type of vasculitis through a candidate gene strategy, via using a GCA sample collection that partially overlaps (48%) with the one analyzed here. Therefore, this study confirms PTPN22 as the non-HLA gene with greatest effect susceptibility to GCA described to date, as it occurs in RA.

Regarding REL, it should be noted that rs115674477 appears as an isolated singleton in our study. Considering its low frequency (MAF < 0.03), reasonable doubts might arise about the reliability of this association. However, the consistency of the effect sizes observed for this SNP across the different populations suggests that it might represent a real association signal. REL encodes a member of the NF-κB family of transcription factors known as c-Rel, which is involved in T cell and antigen-presenting cell function. It has been demonstrated that c-Rel regulates the expression of the Th1-cell-promoting cytokine IL-12 and the Th17-cell-promoting cytokine IL-23 in dendritic cells and macrophages, respectively. Furthermore, c-Rel is also required for the expression of FoxP3, a master regulator in the development and function of regulatory T (Treg) cells. Interestingly, the second non-HLA hit in our study, LRRC32 (also known as glycoprotein A repetition predominant protein [GARP]), is highly expressed in activated Treg cells and it could also control the expression of FoxP3. It has been reported that the surface localization of LRRC32 alters the surface expression of T cell activation markers and that LRRC32+ Treg cells mediate more suppressive signals than LRRC32− Treg cells. Genetic variants of LRRC32 have been associated through genome-wide association studies with UC, atopic dermatitis, and allergic rhinitis (MIM 607154). In UC and allergic rhinitis, the associated variant corresponded with rs215521, one of the two linked LRRC32 SNPs that are more significantly associated with GCA in our study. Considering that current knowledge points to Th1, Th17, and Treg cells, together with dendritic cells and macrophages, as the most relevant cell types in GCA pathophysiology, we propose that both LRRC32 and REL might represent true risk factors for this type of vasculitis, albeit not reaching the genome-wide level of significance.

In summary, through a comprehensive screening of immune loci, we have derived a testable model of different HLA amino acid positions (HLA-DRβ1 13, HLA-DQα1 47, 56, 76, and HLA-B 45) that explains most of the HLA association with GCA. However, it is important to note that, despite the clear biological implication of our model, there was insufficient statistical power in the current dataset to definitively confirm these positions as the true causal drivers of the HLA association with GCA, considering the high level of LD within this
Supplemental Data

Supplemental Data include full affiliations for the Spanish GCA Group, 5 figures, and 12 tables and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2015.02.009.

Consortia


Acknowledgments

The authors thank Sofía Vargas, Sonia García, Gema Robledo, Steve Martin, Lubna-Haroon Raashid, and Edward Flynn for their excellent technical assistance and the participants for kindly accepting their collaboration. Steve Rich, Stephen Suna Onengut-Gumuscu, and Wei-Min Chen are thanked for genotyping and quality control. Marta Conde-Jaldón is thanked for assisting in HLA typing. The Norwegian Systemic Vasculitis and Connective Tissue Disease Registry (NOSVAR) at Oslo University Hospital is acknowledged for providing data on the Norwegian cases. F.D.C., J.M., and M.A.G.-G. were supported by Instituto de Salud Carlos III (ISCIII), Spain, through the RETICS Program RD12/0009/0004 (RIER). M.C.C., J.H.-R., and S.P.-G. were supported by Ministerio de Economía y Competitividad, Spain (SAF 11/30073). S.L.M. received grant funding from Research into Ageing, a Wellcome Trust/AMS Starter Grant for Clinical Lecturers, the Leeds Teaching Hospitals Charitable Trustees, and the Mason Medical Research Foundation and is funded by a NIH Research Clinician Scientist Award. A.W.M., J.H.B., and J.C.T. were supported by the NIHR-Leeds Musculoskeletal Biomedical Research Unit and the Ann Wilks Memorial Fund. The Vasculitis Clinical Research Consortium (VCRC) received support from the United States National Institute of Arthritis and Musculoskeletal and Skin Diseases (U54AR057319), the National Center for Research Resources (U54 RR019497), the Office of Rare Diseases Research, and the National Center for Advancing Translational Science. The VCRC is part of the Rare Diseases Clinical Research Network (RDCRN). A.P.D. was funded by Agder Medforsk (non-profit research organization stationed in Southern Norway). T.W. was funded by the grant DFG KFO TP03.

Received: November 7, 2014
Accepted: February 12, 2015
Published: March 26, 2015

Web Resources

The URLs for data presented herein are as follows:

1000 Genomes, http://browser.1000genomes.org
GEC: Genetic Type I Error Calculator, http://statgenpro.psychiatry.hku.hk/gec/
HLA Nomenclature, http://hla.alleles.org
IMPUTE2, http://mathgen.stats.ox.ac.uk/impute/impute_v2.html

References

8. Hinks, A., Cobb, J., Marion, M.C., Prahalad, S., Sudman, M., Bowes, J., Martin, P., Comeau, M.E., Sajuthi, S., Andrews, R., et al.; Boston Children's JIA Registry; British Society of Paediatric and Adolescent Rheumatology (BSPAR) Study Group; Childhood Arthritis Prospective Study (CAPS); Childhood Arthritis Response to Medication Study (CHARMS); German Society for Pediatric Rheumatology (GKR); JIA Gene Expression Study; NIAMS JIA Genetic Registry; TREAT Study; United Kingdom Juvenile Idiopathic Arthritis Genetics Consortium (UKJAGGC) (2013). Dense genotyping of immune-related disease regions identifies 14 new susceptibility loci for juvenile idiopathic arthritis. Nat. Genet. 45, 664–669.


