A genome-wide association study suggests the HLA Class II region as the major susceptibility locus for IgA vasculitis

Raquel López-Mejías1, F. David Carmona2,3, Santos Castañeda4, Fernanda Genre1, Sara Remuzgo-Martínez1, Belén Sevilla-Perez2, Norberto Ortego-Centeno5, Javier Llorca6, Begoña Ubilla1, Verónica Mijares1, Trinitario Pina3, José A. Miranda-Filloy7, Antonio Navas Parejo8, Diego de Argila9, Maximiliano Aragüés9, Esteban Rubio10, Manuel León Luque10, Juan María Blanco-Madrigal11, Eva Galindez-Aguirregoikoia12, David Jayne13, Ricardo Blanco1, Javier Martín2 & Miguel A. González-Gay3,13,14

The genetic component of Immunoglobulin-A (IgA) vasculitis is still far to be elucidated. To increase the current knowledge on the genetic component of this vasculitis we performed the first genome-wide association study (GWAS) on this condition. 308 IgA vasculitis patients and 1,018 healthy controls from Spain were genotyped by Illumina HumanCore BeadChips. Imputation of GWAS data was performed using the 1000 Genomes Project Phase III dataset as reference panel. After quality control filters and GWAS imputation, 285 patients and 1,006 controls remained in the datasets and were included in further analysis. Additionally, the human leukocyte antigen (HLA) region was comprehensively studied by imputing classical alleles and polymorphic amino acid positions. A linkage disequilibrium block of polymorphisms located in the HLA class II region surpassed the genome-wide level of significance (OR = 0.56, 95% CI = 0.46–0.68). Although no polymorphic amino acid positions were associated at the genome-wide level of significance, P-values of potential relevance were observed for the positions 13 and 11 of HLA-DRB1 (P = 6.67E-05, P = 1.88E-05, respectively). Outside the HLA, potential associations were detected, but none of them were close to the statistical significance. In conclusion, our study suggests that IgA vasculitis is an archetypal HLA class II disease.

Immunoglobulin-A (IgA) vasculitis, also known as Henoch-Schoenlein purpura (HSP), is the most common type of primary small-sized blood vessel leukocytoclastic vasculitis in children, although it may also develop in adults1. Although the classic clinical triad of IgA vasculitis consists of palpable purpura (involving the lower extremities),

1Epidemiology, Genetics and Atherosclerosis Research Group on Systemic Inflammatory Diseases, Rheumatology Division, Hospital Universitario Marqués de Valdecilla, IDIVAL, Santander, Spain. 2Instituto de Parasitología y Biomedicina ‘López-Neyra’, CSIC, PTS Granada, Granada, Spain. 3Departamento de Genética e Instituto de Biotecnología, Universidad de Granada, Granada, Spain. 4Rheumatology Department, Hospital Universitario La Princesa, IIS-IPrincesa, Madrid, Spain. 5Epidemiology and Computational Biology Department, School of Medicine, University of Cantabria, and CIBER Epidemiología y Salud Pública (CIBERESP), IDIVAL, Santander, Spain. 6Division of Rheumatology, Hospital Universitario Lucus Augusti, Lugo, Spain. 7Nephrology Department, Hospital Universitario San Cecilio, Granada, Spain. 8Department of Medicine, University of Cambridge, Cambridge, UK. 9School of Medicine, University of Cantabria, Santander, Spain. 10Rheumatology Department, Hospital Universitario Virgen del Rocío, Sevilla, Spain. 11Rheumatology Department, Hospital Universitario de Basurto, Bilbao, Spain. 12Department of Medicine, University of Cambridge, Cambridge, UK. 13School of Medicine, University of Cantabria, Santander, Spain. 14Cardiovascular Pathophysiology and Genomics Research Unit, School of Physiology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa. Raquel López-Mejías and F. David Carmona contributed equally to this work. Ricardo Blanco, Javier Martín and Miguel A. González-Gay jointly supervised this work. Correspondence and requests for materials should be addressed to R.L.-M. (email: rlopezmejias78@gmail.com)
joints and the gastrointestinal tract, renal complications may also develop in affected individuals. In this regard, the outcome of IgA vasculitis patients is related to the presence of glomerulonephritis, which may lead to chronic renal failure.

IgA vasculitis has a multifactorial etiology in which both environmental and genetic factors seem to contribute to the predisposition and clinical phenotype of the disease. However, the genetic component of this type of vasculitis remains poorly understood, as only a few candidate gene studies have been performed to date.

Unlike the candidate gene approach, genome-wide association studies (GWAS) imply a hypothesis-free analysis of hundreds of thousands of single-nucleotide polymorphisms (SNPs) across the whole genome. This strategy has proven to be a powerful tool to unravel the genetic component of complex diseases during the last decade, including primary vasculitides such as Takayasu Arteritis, Behçet disease, and antineutrophil cytoplasmic antibody (ANCA)-associated vasculitis (AAV).

This study aimed at conducting the first GWAS of IgA vasculitis using the largest series of IgA vasculitis patients of European ancestry ever assessed for a genetic study.

Patients and Methods

Study population. A series of 308 patients diagnosed with IgA vasculitis and 1,018 unaffected and unrelated controls were genotyped in this study. The total number of individuals that passed the quality control (QC) filters mentioned below and were finally included in further analysis was 1,291 (285 and 1,006 for IgA vasculitis patients and controls, respectively). All subjects were from Spain and had European ancestry. IgA vasculitis condition was diagnosed accordingly with both the guidelines included in Michel et al. and the American College of Rheumatology classification criteria for this form of vasculitis. A description of the main clinical features of the IgA vasculitis patients and controls analyzed after QC filters is shown in Supplementary Table S1.

Genotyping and quality controls. Genomic DNA was extracted from peripheral blood samples using standard methods. Genotyping was conducted using the GWAS platform “Infinium HumanCore Beadchip” in an iScan System (Illumina, Inc) and following the manufacturer’s protocol.

Raw data were subjected to the following QC filters using PLINK v.1.07: (1) SNPs with cluster separation <0.4, call rates <0.98, minor allele frequencies (MAF) <0.01, and those deviating from Hardy-Weinberg equilibrium (HWE; P < 0.001) were excluded; (2) samples with call rates <0.95, and those with identity by descent >0.4 were also removed. Sex chromosomes were not analyzed. The number of IgA vasculitis patients and controls that remained after each QC filter is shown in Supplementary Table S2.

Imputation of GWAS data. SNP genotype imputation throughout the genome was performed after initial QC using the 1000 Genomes Project (1KG) Phase III dataset as reference panel (www.1000genomes.org) and the software IMPUTE v.2. For that, we set the strand orientation, chromosome position, and SNP nomenclature accordingly with the build 37 (HG19) of the 1KG using PLINK. Singletons were removed. Finally, possible population sub-stratification was controlled by principal component (PC) analyses using PLINK and the gcta64 and R-base software under GNU Public license v2. To identify outliers, we calculated and plotted the ten first PCs of each individual, and those deviating >4 standard deviations from the cluster centroid were excluded. PC analysis for the first three PCs for each individual are plotted.
The total number of SNPs that passed the QC and were finally analyzed was 1,909,910 (2,581,927 and 2,185,351 for IgA vasculitis patients and controls, respectively). The number of polymorphisms that remained after each QC filter is shown in Supplementary Table S2.

### Human leukocyte antigen (HLA) imputation.

Considering that IgA vasculitis is an immune-mediated condition, a more comprehensive analysis of the HLA region was conducted. With that aim, we extracted the extended HLA region (29,000,000 to 34,000,000 bp in chromosome 6) from the non-imputed data and imputed SNPs, classical HLA alleles at two- and four-digits, and polymorphic amino acid positions as described in Supplementary Table S2.

#### Table 1. Signals within HLA associated with IgA susceptibility at the GWAS significance level P \( < 5 \times 10^{-8} \) after imputation of GWAS data.

| SNP          | Position in chr 6 (GRCh37) | Reference allele | \( P \)      | OR (CI 95%)         |
|--------------|----------------------------|-----------------|-------------|---------------------|
| rs9275260    | 32.661.575                 | C               | 3.42E-09    | 0.56 [0.46–0.68]   |
| rs9275259    | 32.661.572                 | C               | 3.42E-09    | 0.56 [0.46–0.68]   |
| rs9275284    | 32.663.073                 | C               | 4.30E-09    | 0.56 [0.46–0.68]   |
| rs9275285    | 32.663.080                 | A               | 4.30E-09    | 0.56 [0.46–0.68]   |
| rs9275286    | 32.663.143                 | T               | 4.30E-09    | 0.56 [0.46–0.68]   |
| rs9275288    | 32.663.203                 | A               | 4.30E-09    | 0.56 [0.46–0.68]   |
| rs9275292    | 32.663.289                 | C               | 4.30E-09    | 0.56 [0.46–0.68]   |
| rs9275244    | 32.660.881                 | G               | 4.92E-09    | 0.56 [0.46–0.68]   |
| rs6500063    | 32.663.610                 | C               | 5.20E-09    | 0.56 [0.46–0.68]   |
| rs2495522    | 32.664.722                 | A               | 5.25E-09    | 0.56 [0.46–0.68]   |
| rs9275279    | 32.662.843                 | G               | 5.32E-09    | 0.56 [0.46–0.68]   |
| rs9275281    | 32.662.920                 | G               | 5.32E-09    | 0.56 [0.46–0.68]   |
| rs4248168    | 32.659.743                 | G               | 5.66E-09    | 0.56 [0.47–0.68]   |
| rs9275224    | 32.659.878                 | A               | 5.66E-09    | 0.56 [0.47–0.68]   |
| rs4713580    | 32.659.994                 | C               | 5.46E-09    | 0.56 [0.47–0.68]   |
| rs4713581    | 32.660.237                 | G               | 5.46E-09    | 0.56 [0.47–0.68]   |
| rs9275225    | 32.660.262                 | G               | 5.46E-09    | 0.56 [0.47–0.68]   |
| rs50002704   | 32.659.279                 | T               | 5.67E-09    | 0.56 [0.46–0.68]   |
| rs4713581    | 32.660.023                 | T               | 6.08E-09    | 0.56 [0.47–0.68]   |
| rs4713583    | 32.660.153                 | T               | 6.08E-09    | 0.56 [0.47–0.68]   |
| rs9275228    | 32.660.347                 | G               | 6.12E-09    | 0.56 [0.47–0.68]   |
| rs9275227    | 32.660.337                 | C               | 6.12E-09    | 0.56 [0.47–0.68]   |
| rs9275295    | 32.663.391                 | A               | 6.19E-09    | 0.56 [0.46–0.68]   |
| rs9275277    | 32.662.677                 | G               | 7.33E-09    | 0.57 [0.47–0.69]   |
| rs9275276    | 32.662.676                 | T               | 7.33E-09    | 0.57 [0.47–0.69]   |
| rs9275245    | 32.660.943                 | A               | 7.99E-09    | 0.57 [0.47–0.69]   |
| rs50002708   | 32.659.357                 | T               | 8.00E-09    | 0.57 [0.47–0.69]   |
| rs50002707   | 32.659.337                 | T               | 8.00E-09    | 0.57 [0.47–0.69]   |
| rs67836634   | 32.662.128                 | G               | 8.89E-09    | 1.76 [1.45–2.13]   |
| rs9275222    | 32.659.516                 | T               | 1.28E-08    | 1.75 [1.44–2.12]   |
| rs6457617    | 32.663.851                 | C               | 1.44E-08    | 0.57 [0.47–0.70]   |
| rs6457620    | 32.663.999                 | G               | 1.44E-08    | 0.57 [0.47–0.70]   |
| rs50002702   | 32.659.158                 | G               | 1.47E-08    | 0.57 [0.47–0.70]   |
| rs9275226    | 32.660.311                 | C               | 1.49E-08    | 0.57 [0.47–0.70]   |
| rs9275230    | 32.660.442                 | A               | 1.49E-08    | 0.57 [0.47–0.70]   |
| rs50002705   | 32.659.319                 | C               | 1.60E-08    | 0.57 [0.47–0.70]   |
| rs7413581    | 32.660.051                 | T               | 1.63E-08    | 0.57 [0.47–0.70]   |
| rs7411304    | 32.660.170                 | T               | 1.65E-08    | 0.58 [0.47–0.70]   |
| rs9275246    | 32.661.003                 | C               | 2.11E-08    | 0.58 [0.48–0.70]   |
| rs7413587    | 32.659.535                 | G               | 2.56E-08    | 0.58 [0.48–0.70]   |
| rs9275247    | 32.661.015                 | T               | 2.87E-08    | 0.58 [0.48–0.70]   |
| rs9275231    | 32.660.505                 | C               | 2.81E-08    | 1.73 [1.42–2.09]   |

Table 1. Signals within HLA associated with IgA susceptibility at the GWAS significance level P \( < 5 \times 10^{-8} \) after imputation of GWAS data. HLA: Human leukocyte antigen; IgA: Immunoglobulin-A; GWAS: genome-wide association study; SNP: single nucleotide polymorphism; chr: chromosome; OR: odds ratio; CI: confidence interval.
four digits genotyping data of the HLA class I and II molecules. Imputed HLA data were also filtered with PLINK with the following thresholds: success call rate >0.95 for alleles and amino acids, deviation from HWE (P < 0.001) for SNPs, and >0.95 total call rate for individuals. Information of a total of 7,179 SNPs, 423 classical HLA alleles (126 at two-digit and 297 at four-digit resolution) of the HLA-A, HLA-B, HLA-C, HLA-DRB1, HLA-DQB1, HLA-DQA1, HLA-DPA1, and HLA-DPB1 genes, and 1,275 amino acidic variants of the HLA system remained after the filters.

Statistical analyses. An estimation of the statistical power of the final cohort (285 IgA vasculitis patients/1,006 healthy controls) was obtained with CaTS Power Calculator for Genetic Studies software (Supplementary Table S3).

To test for association, we compared the genotype frequencies of every SNP between cases and controls by logistic regression on the best-guess genotypes assuming an additive model in PLINK. The ten first PCs were included as covariates. In the case of the HLA region, we tested SNPs, classical HLA alleles and all possible combinations of amino acid residues per position. A likelihood ratio test of amino acid positions was also conducted, as described.

P-values, odds ratios (OR), and 95% confidence intervals (CI) were then calculated. The statistical threshold was set at the genome-wide level of significance (P < 5E-08). In the HLA analysis, despite not interrogating the whole genome but a specific region of chromosome 6, we decided to maintain the statistical threshold at the genome-wide level of significance (P < SE-08) to avoid possible false positive results.

Results

Figure 1 summarizes the overall results of the study. Several association signals in high linkage disequilibrium (LD, r^2 > 0.8) at the genome-wide level of significance were disclosed within the HLA region at chromosome 6. The strongest signal corresponded to a disequilibrium block of polymorphisms (OR = 0.56, 95% CI = 0.46–0.68) (Table 1), which we refer to as rs9275260, that maps to an intergenic region in HLA class II between HLA-DQA1 and HLA-DQB1. To confirm these results, we obtained direct genotypes of the Spanish cohort using a TaqMan probe for rs9275260. The overall concordance reached after comparing TaqMan types with the corresponding imputed data was 99.84%. Outside the HLA, some potential signals located in different intronic and intergenic regions were observed (Fig. 1), but none of them reached the statistical level of significance (P < SE-08) to avoid possible false positive results.

We tried to narrow down the HLA association with IgA vasculitis by inferring SNPs, classical HLA alleles, and polymorphic amino acid positions using as reference the T1DGC panel. Accordingly, association signals at the genome-wide level of significance were disclosed (Fig. 2A). The genetic variant rs9275224 represented the strongest peak (P = 5.74E-09, OR = 0.56, 95% CI = 0.46–0.68) (Supplementary Table S5). The polymorphism rs9275260 (and SNPs in high linkage disequilibrium with it) observed in the genome-wide data analysis was not detected in the analysis of the HLA region since 1KG Phase III dataset was not used as reference panel in...
this analysis. Nevertheless, rs9275224 was in complete LD ($r^2 = 1$) with rs9275260 (and, consequently, with all the SNPs of the same disequilibrium block) observed in the genome-wide data analysis, meaning that these polymorphisms represent the same signal. Although no polymorphic amino acid positions were associated at the genome-wide significance level, P-values of potential relevance were observed for the HLA-DRB1 positions 13 and 11 (P = 6.67E-05 and P = 1.88E-05, respectively) (Supplementary Table S6). Conditional logistic regression analyses of the HLA data indicated that rs9275224 explained most of the HLA-associated variants in HLA class II (Fig. 2B). Regarding HLA class I, a potential signal in HLA-B was observed (rs2523650, P = 1.10E-05, OR = 1.59, 95% CI = 1.29–1.96).

Discussion

This study represents the first GWAS of IgA vasculitis. Consistent with the results obtained in a former study⁴, our data suggest the involvement of HLA class II region in the pathophysiology of IgA vasculitis, thus supporting the high relevance of the immune system in the development of this disease and suggesting that IgA vasculitis may be related to other class II vasculitides such as giant cell arteritis (GCA)¹² or AAV³⁷. The strongest signal mapped to the HLA-DQA1/DQB1 region, which is in high LD with the HLA-DRB1 gene. Consequently, the associated polymorphisms may be tagging a putative aetiologic variant at HLA-DRB1. Regarding polymorphic amino acid positions, none of the signals reached the genome-wide level of significance. Nevertheless, likewise rheumatoid arthritis²⁹ and GCA¹², the HLA-DRB1 positions 13 and 11 were amongst the strongest signals, which support the notion that IgA vasculitis may share immunopathogenic pathways with these conditions. On the other hand, after performing the conditional analysis on the HLA data, a potential signal that maps to HLA-B was observed, although it did not reach the genome-wide level of significance. This result could be indicating a potential effect of HLA class I in the pathogenesis of IgA vasculitis, as previously proposed¹. In addition, no consistent associations with IgA vasculitis susceptibility were detected outside the HLA region, probably due to an insufficient statistical power to detect risk variants with a moderate effect.

Vasculitides constitute a heterogeneous group of diseases that often have overlapping clinical and pathological manifestations⁴⁸. Nevertheless, differences between them in molecular terms have been described¹. In this regard, the results derived from our study classify IgA vasculitis as a HLA class II condition linking it to GCA and AAV. Nonetheless, it is important to keep in mind that the number of cases recruited in our study was not high and replication was not carried out. Because of that, further confirmatory studies in independent populations should be performed to validate our data.

In summary, our results suggest that IgA vasculitis is an archetypal HLA class II disease.

References

1. Gonzalez-Gay, M. A. & Garcia-Porrua, C. Epidemiology of the vasculitides. Rheum Dis Clin North Am. 27, 729–49 (2001).
2. Calvo-Rio, V. et al. Henoch-Schönlein purpura in northern Spain: clinical spectrum of the disease in 417 patients from a single center. Medicine (Baltimore). 93, 106–13 (2014).
3. Wyatt, R. J. & Julian, B. A. IgA nephropathy. N Engl J Med. 368, 2402–14 (2013).
4. López-Mejías, R. et al. Association of HLA-B*41:02 with Henoch-Schönlein Purpura (IgA Vasculitis) in Spanish individuals irrespective of the HLA-DRB1 status. Arthritis Res Ther. 17, 102 (2015).
5. López-Mejías, R. et al. HLA-DRB1 association with Henoch-Schönlein purpura. Arthritis Rheumatol. 67, 823–7 (2014).
6. McCarthy, M. I. et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet. 9, 356–69 (2008).
7. Carmona, F. D., Martin, J. & González-Gay, M. A. Genetics of vasculitis. Curr Opin Rheumatol. 27, 10–7 (2015).
8. Michel, B. A., Hunder, G. G., Bloch, D. A. & Calabrese, L. H. Hypersensitivity vasculitis and Henoch-Schönlein purpura: a comparison between the 2 disorders. J Rheumatol. 19, 721–8 (1992).
9. Mills, J. A. et al. The American College of Rheumatology 1990 criteria for the classification of Henoch-Schönlein purpura. Arthritis Rheum. 33, 1114–21 (1990).
10. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 81, 559–75 (2007).
11. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 5, e1000529 (2009).
12. Carmona, F. D. et al. A large-scale genetic analysis reveals a strong contribution of the HLA class II region to giant cell arteritis susceptibility. Am J Hum Genet. 96, 565–80 (2015).
13. Raychaudhuri, S. et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat Genet. 44, 291–6 (2012).
14. Iba, X. et al. Imputing amino acid polymorphisms in human leukocyte antigens. PLoS One 8, e64683 (2013).
15. Mayes, M. D. et al. Immunochip analysis identifies multiple susceptibility loci for systemic sclerosis. Am J Hum Genet. 94, 47–61 (2014).
16. Ortiz-Fernández, L. et al. Genetic Analysis with the Immunochip Platform in Behçet Disease. Identification of Residues Associated in the HLA Class I Region and New Susceptibility Loci. PLoS One 11, e0161305 (2016).
17. Lyons, P. A. et al. Genetically distinct subsets within ANCA-associated vasculitides. N Engl J Med. 367, 214–23 (2012).
18. Jennette, J. C. et al. 2012 revised International Chapel Hill Consensus Conference Nomenclature of Vasculitides. Arthritis Rheum. 65, 1–11 (2013).

Acknowledgements

We wish to thank all the patients with IgA vasculitis and controls for their participation in this study. We want to specially thank Patricia Fuentevilla Rodriguez, María Del Camino Villa Llamazares and Maria Eugenia CuadradoMantecón for their technical assistance. This study was supported by European Union FEDER funds and “Fondo de Investigaciones Sanitarias” (grant PI12/00193) from Instituto de Salud Carlos III (ISCIII, Health Ministry, Spain). RL-M was supported by the Miguel Servet I programme of the Spanish Ministry of Economy and Competitiveness through the grant RYC-2014-16458. FG was recipient of a Sara Borrell postdoctoral fellowship from the “Instituto Carlos III de Salud” at the Spanish Ministry
of Health (Spain) (CD15/00095). SR-M and BU were supported by funds from the RETICS Program (RIER) (RD16/0012/0009 and RD12/0009/0013, respectively).

**Author Contributions**
R.L.-M. and F.D.C. conceptualized and designed the study, drafted the initial manuscript, and approved the final manuscript as submitted. S.C., F.G., S.R.-M., B.S.-P., N.O.-C. and J.L. carried out the initial analyses, reviewed and revised the manuscript, and approved the final manuscript as submitted. B.U. and V.M. performed the initial analyses, drafted the manuscript and read and approved the final version of the manuscript. T.P., J.A.M.-F., A.N.P., D.A., M.A., E.R., M.L.L., J.M.B.-M., E.G.-A. and D.J. Jayne carried out the initial analyses, reviewed and revised the manuscript, and approved the final manuscript as submitted. R.B., J.M. and M.A.G.-G. designed the data collection instruments, and coordinated and supervised data collection, critically reviewed the manuscript, and approved the final manuscript as submitted.

**Additional Information**

**Supplementary information** accompanies this paper at doi:10.1038/s41598-017-03915-2

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

[Open Access] This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017