A genome-wide association study identifies risk alleles in plasminogen and P4HA2 associated with giant cell arteritis

F. David Carmona[1,2,*], Augusto Vaglio[3], Sarah L. Mackie[4], José Hernández-Rodríguez[5], Paul A. Monach[6], Santos Castañeda[7], Roser Solans[8], Inmaculada C. Morado[9], Javier Narváez[10], Marc Ramentol-Sintas[8], Colin T. Pease[4], Bhaskar Dasgupta[11], Richard Watts[12], Nader Khalidi[13], Carol A. Langford[14], Steven Ytterberg[15], Luigi Boiardi[16], Lorenzo Beretta[17], Marcello Govoni[18], Giacomo Emmi[19], Francesco Bonatti[20], Marco A. Cimmino[21], Torsten Witte[22], Thomas Neumann[23], Julia Holle[24], Verena Schönau[25], Laurent Sailler[26], Thomas Papo[27], Julien Haroche[28], Alfred Mahr[29], Luc Mouthon[30], Øyvind Molberg[31], Andreas P. Diamantopoulos[32], Alexandre Voskuyl[33], Elisabeth Brouwer[34], Thomas Daikeler[35], Christoph T. Berger[36], Eamonn S. Molloy[37], Lorraine O'Neill[37], Daniel Blockmans[38], Benedicte A. Lie[39], Paul Mclaren[40], Timothy J. Vyse[41], Cisca Wijmenga[42], Yannick Allanore[43], Bobby P.C. Koeleman[44], Jennifer H. Barrett[4], María C. Cid[5], Carlo Salvarani[16], Peter A. Merkel[45], Ann W. Morgan[4], Miguel A. González-Gay[46], Javier Martín[1]

[1] Instituto de Parasitología y Biomedicina 'López-Neyra', CSIC, PTS Granada, Granada 18016, Spain.

[2] Departamento de Genética e Instituto de Biotecnología, Universidad de Granada, Granada, Spain.

[3] Unit of Nephrology, University Hospital of Parma, Parma 43126, Italy.

[4] School of Medicine, University of Leeds and NIHR-Leeds Musculoskeletal Biomedical Research Unit, Leeds Teaching Hospitals NHS Trust, Leeds LS7 4SA, UK.

[5] Vasculitis Research Unit, Department of Autoimmune Diseases, Hospital Clínic University of Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona 08036, Spain.

[6] Section of Rheumatology, Boston University, Boston, MA 02118, USA.

[7] Department of Rheumatology, Hospital de la Princesa, IIS-IP, Madrid 28006, Spain.

[8] Autoimmune Systemic Diseases Unit, Department of Internal Medicine, Hospital Vall d'Hebron, Autonomous University of Barcelona, Barcelona 08035, Spain.

[9] Department of Rheumatology, Hospital Clínico San Carlos, Madrid 28040, Spain.

[10] Department of Rheumatology, Hospital Universitario de Bellvitge-IDIBELL, L'Hospitalet de Llobregat, Barcelona 08907, Spain.

[11] Department of Rheumatology, Southend University Hospital NHS Foundation Trust, Westcliff-on-Sea SS0 0RY, UK.

[12] Department of Rheumatology, Ipswich Hospital NHS Trust, Ipswich IP4 5PD, UK.

[13] Division of Rheumatology, McMaster University, Hamilton, ON L8N 1Y2, Canada.

[14] Division of Rheumatology, Cleveland Clinic, Cleveland, OH 44195, USA.

[15] Division of Rheumatology, Mayo Clinic, Rochester, MN 55905, USA.

[16] Rheumatology Unit, Department of Internal Medicine, Azienda Ospedaliera Arcispedale Santa Maria Nuova, Istituto di Ricovero e Cura a Carattere Scientifico, Reggio Emilia 42123, Italy.

[17] Referral Center for Systemic Autoimmune Diseases, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico di Milano, Milan 20122, Italy.

[18] Department of Rheumatology, Azienda Ospedaliero Universitaria S. Anna, University of Ferrara, Ferrara 44121, Italy.

[19] Department of Experimental and Clinical Medicine, University of Florence, Florence 50134, Italy.

[20] Department of Clinical and Experimental Medicine, Medical Genetics Unit, University of Parma, Parma 43124, Italy.

[21] Research Laboratory and Academic Division of Clinical Rheumatology, Department of Internal Medicine, University of Genova, Genova 16132, Italy.

[22] Hannover Medical School, Hannover 30625, Germany.
[23] Klinik für Innere Medizin III, University-Hospital Jena, Jena 07743, Germany.
[24] Vasculitis Clinic, Klinikum Bad Bramstedt & University Hospital of Schleswig Holstein, Bad Bramstedt 24576, Germany.
[25] Department of Rheumatology and Immunology, Universitätsklinikum Erlangen, Erlangen 91054, Germany.
[26] Service de Médecine Interne, Hopital de Purpan, CHU de Toulouse, Toulouse 31300, France.
[27] Service de Médecine Interne, Hôpital Bichat, Université Paris-Diderot, Paris 75018, France.
[28] Department of Internal Medicine & French Reference Center for Rare Auto-immune & Systemic Diseases, Assistance Publique-Hôpitaux de Paris (AP-HP), Pitié-Salpêtrière Hospital, Paris 75013, France.
[29] Department of Internal Medicine, Hospital Saint-Louis ECSTRA Team, Epidemiology and Biostatistics, Sorbonne Paris Cité Research Center UMR 1153, Inserm, University Paris Diderot, Paris 75010, France.
[30] Department of Internal Medicine, Cochin Hospital, National Referral Center for Rare Autoimmune and Systemic Diseases, AP-HP, Université Paris Descartes, Paris 75014, France.
[31] Department of Rheumatology, Oslo University Hospital, Oslo 0424, Norway.
[32] Department of Rheumatology, Hospital of Southern Norway Trust, Kristiansand 4604, Norway.
[33] Amsterdam Rheumatology and Immunology Center, VU University Medical Center, Amsterdam 1007 MB, The Netherlands.
[34] Department of Rheumatology and Clinical Immunology, University of Groningen, University Medical Center Groningen 9713 GZ, Netherlands.
[35] Rheumatology Department of Internal Medicine, University Hospital, Basel 4056, Switzerland.
[36] Translational Immunology and Medical Outpatient Clinic, Departments of Biomedicine and Internal Medicine, University Hospital, Basel 4056, Switzerland.
[37] Department of Rheumatology, Centre for Arthritis and Rheumatic Diseases, St Vincent's University Hospital, Dublin Academic Medical Centre, Dublin 4, Ireland.
[38] Department of General Internal Medicine, University Hospital Gasthuisberg, Leuven 3000, Belgium.
[39] Department of Medical Genetics, University of Oslo and Oslo University Hospital, Oslo 0450, Norway.
[40] Swiss Institute of Bioinformatics, Lausanne 1015, Switzerland.
[41] Divisions of Genetics and Molecular Medicine and Immunology, Infection and Inflammatory Disease, King's College London, London SE1 9RT, UK
[42] Department of Genetics, University Medical Center Groningen, Groningen 9713 AV, The Netherlands.
[43] Paris Descartes University, Rheumatology A department, Sorbonne Paris Cité, Cochin Hospital, Paris 75014, France.
[44] Department of Medical Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht 3584 CX, The Netherlands.
[45] Vasculitis Center, Division of Rheumatology, University of Pennsylvania, Philadelphia, PA 19104, USA.
[46] Department of Rheumatology, Hospital Universitario Marqués de Valdecilla, IDIVAL, University of Cantabria, Santander 39008, Spain.

***Correspondence to**: F. David Carmona, PhD. Instituto de Parasitología y Biomedicina 'López-Neyra'. Consejo Superior de Investigaciones Científicas, Parque Tecnológico Ciencias de la Salud. Avenida del Conocimiento s/n, 18016-Armilla (Granada), Spain. E-mail: dcarmona@ipb.csic.es

ABSTRACT

Giant cell arteritis (GCA) is the most common form of vasculitis in individuals older than 50 years in Western countries. To shed light onto the genetic background influencing susceptibility for GCA, we performed a genome-wide association screening in a well-powered study cohort. After imputation, 1,844,133 genetic variants were analysed in 2,134 cases and 9,125 unaffected controls from ten independent populations of European ancestry. Our data confirmed HLA class II as the strongest associated region (independent signals: rs9268905, P = 1.94E-54, per-allele OR = 1.79; and rs9275592, P = 1.14E-40, OR = 2.08). Additionally, *PLG* and *P4HA2* were identified as GCA risk genes at the genome-wide level of significance (rs4252134, P = 1.23E-10, OR = 1.28; and rs128738, P = 4.60E-09, OR = 1.32, respectively). Interestingly, we observed that the association peaks overlapped with different regulatory elements related to cell types and tissues involved in the pathophysiology of GCA. PLG and P4HA2 are involved in vascular remodelling and angiogenesis, suggesting a high relevance of these processes for the pathogenic mechanisms underlying this type of vasculitis.

Introduction

During the last decade, genome-wide association studies (GWAS), in which common genetic variation across the whole genome is interrogated in a hypothesis-free fashion, were a breakthrough in biomedical research methodology and have led to the identification of thousands of robust genetic associations within a wide spectrum of complex human diseases [1]. However, some diseases of low prevalence have received less attention due to the difficulty in recruiting well-powered study cohorts, even though rare diseases result in a significant disease burden. An example is giant cell arteritis (GCA [MIM 187360]), the most common form of vasculitis in Western countries in people over 50 years old. GCA is characterised by chronic inflammation of large arteries, such as the aorta, the carotid arteries and its extracranial branches, which may lead to severe clinical sequelae if not treated promptly, including visual loss, scalp and tongue necrosis, aortic dissection/rupture or cerebral infarction [2-4]. Although the genetic component of GCA has previously been investigated following a candidate gene approach, most of the described genetic associations were based on underpowered analyses and usually failed to be replicated in independent populations [5].

Recently, an international collaborative effort involving different European and North American research consortia has made possible the conduct of more powerful studies, including an Immunochip study, that have identified firm risk signals for GCA predisposition, such as HLA molecules and key genes of the immune response like protein tyrosine phosphatase, non-receptor type 22 (*PTPN22* [MIM 600716]) and interleukin 17A (*IL17A* [MIM 603149]) [6-8]. Taking advantage of the large sample collection that this collaboration has enabled, comprising ten independent populations of European ancestry that cover the whole gradient for prevalence of the disease, we performed an agnostic genetic study in GCA at the genome-wide level.

Subjects and Methods

Study population

This study included a total of 2,134 GCA cases and 9,125 unaffected controls from ten independent populations of European ancestry: Spain (805 cases and 1,323 controls), United Kingdom (352 cases and 2,965 controls), Italy (271 cases and 960 controls), North America (176 cases and 1,181 controls from The United States of America and Canada), Germany (160 cases and 667 controls), France (114 cases and 488 controls), Norway (104 cases and 121 controls), The Netherlands (69 cases and 638 controls), Switzerland (46 cases and 500 controls) and Ireland (37 cases and 282 controls). The diagnosis of GCA was established according to the 1990 American College of Rheumatology classification criteria for this disease [9]. In addition, the diagnosis was subsequently confirmed by either a biopsy of the temporal artery (89.83%) or arterial imaging (10.17%) consistent with GCA. A detailed description of the main clinical characteristics of the different case cohorts is provided in **Table S1**. All participants signed an informed consent form before being enrolled in the study. The procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) of all participant centres.

Genotyping and quality controls

Genomic DNA from peripheral blood samples of all GCA cases as well as the Spanish and Irish controls were screened using the GWAS platform "Infinium® HumanCore Beadchip" in an iScan System and the Genotyping Module (v.1.9) of the GenomeStudio software (Illumina, Inc). The genotyping was conducted in the Genomics and Genotyping Unit of the Pfizer-University of Granada-Junta de Andalucía Centre for Genomics and Oncological Research (GENYO, Granada, Spain), following the manufacturer's protocol. Single-nucleotide polymorphisms (SNP) with a cluster separation <0.4 were removed after the calling.

Genotyping data from the remaining control population were obtained from 9 cohorts of geographically matched European ancestries included in previous studies [10-17] (**Table S2**).

All studies were subjected to stringent quality-control measures separately based on the same analytical protocol using PLINK v.1.07 [18]. Specifically, we filtered out SNPs with call rates <0.98, minor allele frequencies (MAF) <0.01, and those that deviated from Hardy-Weinberg equilibrium (HWE; P<0.001 in both cases and controls). Similarly, samples with less than 95% of successfully called SNPs, and one subject per pair of first-degree relatives (identity by descent > 0.4) were removed. Sex chromosomes were also excluded from the analysis.

Finally, to check the consistency of the results, we re-genotyped the Spanish samples for the associated signals using predesigned TaqMan® 5' SNP genotyping assays (assay IDs: C__16222465_10, C__25614474_20, and C___2397211_10) in a 7900HT Fast Real-Time PCR System (Applied Biosystems, Foster City, California, USA), and compared the TaqMan types with the corresponding imputed data.

Imputation methods

After applying the QC filters, whole-genome SNP genotype imputation was carried out with IMPUTE v.2 [19] using the combined 1000 Genome Project Phase III (1KGPh3) data as reference panel, which includes 2,504 individuals [20]. In brief, for each individual dataset, the strand orientation, chromosome position, and SNP identification were updated in PLINK to match the build 37 (HG19) of the 1KGPh3. Next, PLINK-formatted files were converted to the appropriate format with GTOOL and the genotyping data were split into chunks of 50,000 Mb covering whole-genome regions. Imputation was done separately for each study. To ensure quality of imputed SNPs, the most likely call was used for merging genotypes, but only if the most likely call exceeded a probability threshold of 0.9 (the genotype was set to missing otherwise). As the individual imputation of each case/control set could lead to stratification, imputed data were also

subjected to rigorous quality filters in PLINK, including removal of SNPs with call rates <0.98, MAFs <0.01, and those that were not in HWE (P<0.001). Moreover, after merging case/control sets, singleton SNPs and those showing strong evidence of discordance in the genotype distribution between cases and controls due to possible miscalling were removed using an in-house Perl script. Finally, principal component (PC) analyses were conducted to control for possible population stratification using PLINK and the GCTA64 and R-base software under GNU Public license v2. For that, we calculated and plotted the ten first PCs of each individual and those samples located at >4 standard deviations from the cluster centroids of each cohort were considered outliers and, consequently, excluded from the analyses.

The final numbers of individuals and SNPs that remained in the filtered datasets are shown in **Table S2**.

Statistical analyses

Table S3 shows the estimation of the overall statistical power of this study accordingly with the CaTS Power Calculator for Genetic Studies software, which implements the methods described in Skol *et al*. assuming an additive genetic model [21].

PLINK was used to conduct all the case/control analyses. First, the genotype frequencies of all markers were compared between cases and controls of every individual dataset by logistic regression on the best-guess genotypes (>0.9 probability) assuming an additive model and using the ten first study-specific PCs and the gender as covariates. Next, a combined analysis of all studies was performed using the inverse variance weighted meta-analysis under a fixed effects model. In order to maximise the total number of SNPs analysed without compromising the consistency of the results, we tested all SNPs present in the largest GCA sample set (Spain) and in one or more additional studies (*i.e*. ranging from 2 to 10 studies but including always the Spanish set). A total of 1,844,133 were evaluated in the meta-analysis. To identify independent effects across associated regions, dependency analyses at the cohort

level genome-wide scans were conducted in PLINK by step-wise logistic regression with adjustment for the most associated signals in the common set of SNPs, following by inverse variance weighted meta-analysis under a fixed effects model. The heterogeneity of the odds ratios (OR) across the different studies was estimated using both I^2 and Cochran's Q tests. The Manhattan plots were obtained with an in-house R script, and the zooms of the associated regions were obtained with LocusZoom v1.1 [22].

Functional annotation of associated variants

We evaluated the putative functional implications of the GCA risk *loci* by implementing our data with publicly available functional annotation data.

Despite the high efficiency of the imputation process, it was not possible to obtain imputed types for every known polymorphism. Therefore, in a first step, we identified all the SNP taggers (r^2>0.8) of the associated signals of our GWAS using the 1KGPh3 data for the European populations and PLINK. All taggers were considered equally as candidates for prioritising casualty or hypothesising possible molecular causes of the observed associations in the subsequent bioinformatic approaches. Then, we explored whether the taggers of each GWAS signal had possible functional implications. The tools PolyPhen-2 [23] and Combined Annotation Dependent Depletion (CADD) [24] were used to evaluate possible damaging effects on the protein sequence of coding non-synonymous SNPs. Regarding the intronic and intergenic variants, we explored whether they lay within known or predicted regulatory DNA elements (including regions of DNAase hypersensitivity, binding sites of transcription factors, promoter regions, chromatin marks, etc.) and whether they had predictive effects on clinical phenotypes using the online tools for exploring annotations of the noncoding genome RegulomeDB [25] and HaploReg v4.1 [26]. Source of these databases includes public datasets from the projects Gene Expression Omnibus (GEO), the Roadmap Epigenomics, the Encyclopedia of DNA Elements (ENCODE), and 1KGPh3, as well as published literature.

In addition, to provide an illustrative picture of the current knowledge on the GCA genetics, we conducted a molecular pathway enrichment analysis, considering both previously suggested GCA genes and those showing an association with disease susceptibility in this study, using the tool for that purpose of the Gene Ontology (GO) reference genome project [27; 28], powered by the Protein Analysis Through Evolutionary Relationships (PANTHER) Classification System [29]. Biological pathways showing *P*-values lower than 0.05 after Bonferroni correction were considered associated with the disease. Finally, predictive protein-protein relationships amongst these same genes were also tested using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database [30]. Only candidate genes that showed statistically significant signals after correction for multiple testing (which was performed with a previously validated method to control the genome-wide type 1 error rate at 0.05 [31]) were included in the above analyses.

Results

Testing for association with disease susceptibility

Three genomic regions contained association signals at the genome-wide level of significance in the overall meta-analysis: human leukocyte antigen (HLA) class II, plasminogen (*PLG* [MIM 173350]), and prolyl 4-hydroxylase subunit alpha 2 (*P4HA2* [MIM 600608]) (**Table 1**, **Figure 1** and **Table S4**).

Within the HLA region, the top hit was rs9268905 (P = 1.94E-54, OR = 1.79, 95% CI 1.67-1.93), located between *HLA-DRA* [MIM 142860] and *HLA-DRB1* [MIM 142857]. Dependence analyses suggested a possible independent effect on disease susceptibility within the *HLA-DQA1* [MIM 146880] / *HLA-DQA2* [MIM 613503] genomic region after controlling for the rs9268905 signal (lead SNP = rs9275592, adjusted P = 1.78E-10, OR = 1.48, 95% CI 1.31-1.66). No association at the genome-wide level of significance remained within the HLA after conditioning on both rs9268905 and rs9275592 (**Table S5** and **Figure S1**).

Outside the HLA region, rs4252134, located in an intron of *PLG* at chromosome 6, represented the most strongly associated variant (P = 1.23E-10, OR = 1.28, 95% CI 1.19-1.39, **Figure 2**). Although some heterogeneity was observed between studies (I^2=48.8), consistent OR directions of the minor allele (towards risk) were observed in all sets (**Table 1**, **Table S4** and **Figure S2**). Other SNPs within the gene (both intronic and exonic) also showed significant P-values, but their statistical significance was lost when conditioned on rs4252134 (**Table S6** and **Figure S3**). Finally, another intronic SNP of *P4HA2* at chromosome 5, rs128738, surpassed the statistical threshold (P = 4.60E-09, OR = 1.32, 95% CI 1.20-1.45). This genomic region also showed additional suggestive signals, but none of them was significant at the genome-wide level (**Figure 2**).

To confirm the above results, we obtained direct genotypes of the whole Spanish cohort using TaqMan probes for rs128738 and two perfect proxies (r2 = 1) of

rs4252134 and rs9268905 (rs4252125 and rs2395185, respectively). The overall concordance reached after comparing TaqMan types with the corresponding imputed data was 99.94% for rs128738, 99.81% for rs4252134/rs4252125, and 99.94% for rs9268905/rs2395185".

Functional annotations of proxies of the non-HLA hits

To prioritise variants that could drive the observed non-HLA associations, we identified all the SNPs in high linkage disequilibrium (LD, $r^2 > 0.8$) with the lead signals in the European populations of 1KGPh3 (**Tables S7, S8)**, and used public online annotation tools to evaluate their possible functional implications. One *PLG* polymorphism (rs4252125) was annotated as missense in the dbSNP database. However, this variant was predicted to be benign according to PolyPhen-2, and showed no evidence of being potentially pathogenic after scoring the deleteriousness with CADD (raw score = -0.63, PHRED-like scaled C-score = 0.104). Then, using RegulomeDB, we identified those SNPs overlapping with known and predicted DNA elements with a higher probability of regulatory effects (score ≤ 3) (**Tables S7, S8)**, and performed a detailed functional annotation of these tagger variants and the lead SNPs using HaploReg (**Table 2** and **Tables S9-S11**). Most of them overlapped with DNase hypersensitivity sites and histone marks enriched at promoters and enhancers (**Table 2**). Interestingly, some of these annotations were related to cell types and tissues involved in GCA pathophysiology. For example, *PLG* rs4252135 co-localised with DNase peaks in different immune cell lines and had a predicted enhancer chromatin state in lymphoblastoid cells (**Tables S9, S10**). A higher enrichment of promoter and enhancer epigenetic marks in these tissues was observed for the prioritised *P4HA2* variants (**Table S9**). Specifically, the lead signal rs128738 overlapped with enhancer histone marks in immune cells and with the imputed Transcription 3' Enhancer mark in the aorta (**Table S10**). Additionally, *P4HA2* rs156023 showed evidence of influencing

enhancer activity in hematopoietic stem cells and neutrophils as well as promoter activity in mononuclear cells and monocytes (**Tables S9, S10**).

In addition, key regulatory proteins (some of them related to the immune response) bound by ChIP-seq experiments as well as relevant motif disruptions were reported in most cases (**Table 2**).

Consequently, most prioritised SNPs also correlated with eQTL effects in peripheral blood monocytes and lymphoblastoid cells (**Table 2** and **Table S11**). Interestingly, rs101194 and rs152054 were reported to affect *P4HA2* expression in arterial tissues in the Genotype-Tissue Expression (GTEx) study [32], with rs101194 specifically acting as cis-eQTL in the aorta (**Table S11**).

Candidate genes and pathway analysis

We also checked the statistical significance in our GWAS of previously described GCA-associated genomic regions (± 100 Kbp 3' and 5' of the reported gene) through candidate gene and Immunochip studies [5; 8]. Significant associations after controlling for multiple testing were observed across most of the analysed regions, with the myeloperoxidase (*MPO* [MIM 606989]), tumor necrosis factor (*TNF* [MIM 191160]), interleukin 6 (*IL6* [MIM 147620]), nitric oxide synthase 2 (*NOS2* [MIM 163730]), and *PTPN22* regions harbouring the strongest non-HLA hits (*MPO* rs10853005, P = 7.19E-05, OR = 0.84; *TNF* rs4959077, P = 2.08E-04, OR = 1.42; *IL6* rs77741999, P = 2.17E-04, OR = 1.20; *NOS2* rs4255826, P = 5.75E-04, OR = 0.87; and *PTPN22* rs2476601, P = 7.88E-04, OR = 1.24) (**Tables S12**, **S13**). Subsequently, we accomplished a protein-protein interaction (PPI) and biological pathway enrichment analysis with those candidate hits showing statistically significant signals after correction for multiple testing in our GWAS (**Table S13**). The molecular network of the selected proteins had significantly more interactions than expected (number of nodes: 13, number of edges: 14, average node degree: 2.15, clustering coefficient: 0.851, expected number of edges: 3, PPI enrichment P = 1.13E-06; **Figure 3**). Interestingly, PLG represented a

relevant node showing evidence of interaction with different proteins (*e.g.* NOS3, IL-6, and TNF) (**Figure 3**). Regarding the functional enrichments of the network, the most significantly associated GO processes were those related to regulation of both cell-cell adhesion and the immune/inflammatory response (**Table S14**).

Discussion

This study represents an unbiased screening of genetic variation in GCA at the genome-wide level. GCA was one of the few types of vasculitis in which GWAS data were not available. Therefore, the results presented here may help to better understand the pathogenic mechanisms underlying this condition and its genetic similarities with other vasculitides. In this sense, our data reinforces the idea of GCA as an archetypal HLA class II disease mediated by an antigen-driven immune response [33], which is in contrast not only to Takayasu's arteritis (TAK [MIM 207600]), another large-vessel vasculitis, but also to other forms of vasculitis associated with class I molecules like Behçet's disease [MIM 109650] [34].

Two independent association signals with GCA predisposition were observed within the HLA region, one located between the *HLA-DRA* and *HLA-DRB1* genes and another one between *HLA-DQA1* and *HLA-DQA2*. This is consistent with the amino acid model that we proposed using imputed Immunochip data to explain the HLA class II association with GCA, which comprised the positions 13 and 56 of the DRβ1 and DQα1 molecules, respectively [8]. Indeed, the strongest hit in the Immunochip study (which had partial overlap of the sample collections with this one) was an SNP that tagged the model (rs477515) in high LD with the two independent HLA lead SNPs that we observed in our GWAS (rs9268923: r^2 = 0.84, D' = 0.95; rs3957146: r^2 = 0.23, D' = 0.99).

Regarding the non-HLA associations, several variants within the *PLG* gene were firmly associated with risk to develop GCA at the genome-wide level of significance in this study, although they all were in high LD and represented a single signal according to the dependence analysis. This gene encodes a secreted blood zymogen that can be converted through a complex conformational modification into two different active proteins, plasmin and angiostatin [35]. The plasminogen system has an important role in a wide spectrum of physiological processes, including wound healing, fibrinolysis,

angiogenesis, and lymphocyte recruitment and inflammation via production of cytokines and reactive oxygen species [36; 37], all of them relevant processes in GCA [38]. Considering the opposite roles of plasmin and angiostatin in the induction of pro-angiogenic processes [36; 37], we hypothesise that the *PLG* risk alleles could unbalance the metabolism of its encoded protein leading to the characteristic pro-inflammatory phenotypes of GCA, although there is a lack of experimental support for this assumption. Interestingly, anti-plasminogen antibodies have been correlated with systemic disease activity in ANCA-associated vasculitis [MIM 608710] [39], a type of vasculitis involving small- to medium-sized blood vessels that is also strongly associated with HLA class II molecules [40]. In addition, *PLG* has been shown to be a shared risk gene for coronary artery disease [MIM 608320] and periodontitis [MIM 260950], characterised by chronic inflammation [41], and multiple sclerosis [MIM 126200], which is also an immune-mediated condition [42]. Future studies aimed at improving the understanding of the zymogen activation may shed light into the *PLG* association with GCA.

P4HA2 represents the second non-HLA hit in our GWAS. This gene encodes an isoform of the alpha subunit of the collagen prolyl 4-hydroxylase, which catalyses the formation of 4-hydroxyproline from proline residues that is essential for collagen biosynthesis, as it is required for the proper three-dimensional folding of newly synthesized procollagen chains [43]. *P4HA2* is considered an important hypoxia response gene and its expression is regulated by hypoxia-inducible factor-1 (HIF-1) [44]. Other relevant HIF-1-induced genes include serpin family E member 1 (*SERPINE1* [MIM 173360], which is the principal inhibitor of the plasmin activation), and genes that have been previously associated with GCA risk through candidate gene studies, such as vascular endothelial growth factor A (*VEGFA* [MIM 192240], a potent endothelial growth factor), matrix metalloproteinase 9 (*MMP9* [MIM 120361], involved in the breakdown of extracellular matrix), and *IL6* (a pro-inflammatory cytokine) [5; 45; 46]. Indeed, the hypoxic induction of all these HIF-1 target genes (and, indirectly, the

inhibition of plasmin formation) is also related to the typical processes involved in the lesions of GCA individuals, *i.e.* fibrosis, inflammation, destruction of the internal elastic lamina, and vascular remodelling, with proliferation and migration of medial myofibroblasts and neoangiogenesis [38; 45]. The prioritised *P4HA2* SNPs proposed here correlated with cis-eQTLs in immune cells and arterial tissues. In particular, expression data indicate that rs101194 may influence *P4HA2* expression in whole blood, lymphoblastoid cells, and the aorta, which is one of the most severely affected vessels in GCA [32; 38; 47-49].

Overall, our results are consistent with the currently accepted understanding of the pathophysiology of GCA, in which vascular remodelling and angiogenesis, either under hypoxic conditions or by dysregulation of hypoxia-sensitive genes, are critical to the development of the clinical presentations [50]. Future analysis of the genetic overlap between GCA and other forms of vasculitis using GWAS data may help to elucidate whether these pathogenic processes are a common feature in vasculitides, and to identify other relevant pathways for the development of GCA.

To summarise, through the analysis of common variation across the whole genome, we have identified *PLG* and *P4HA2* as the main non-HLA genetic factors underlying GCA predisposition. Their crucial role in neoangiogenesis highlights the high relevance of this process in the pathogenic mechanisms leading to this form of vasculitis.

Supplemental Data

Supplemental Data include three figures and fourteen tables, and can be found with this article online at http://www.cell.com/AJHG.

Consortia

The members of the "*Spanish GCA Group*" are José Luis Callejas, Luis Caminal-Montero, Marc Corbera-Bellalta, Eugenio de Miguel, J. Bernardino Díaz López, María Jesús García-Villanueva, Carmen Gómez-Vaquero, Mercedes Guijarro-Rojas, Ana Hidalgo-Conde, Begoña Marí-Alfonso, Agustín Martínez Berriochoa, Aleida Martínez Zapico, Víctor Manuel Martínez-Taboada, José A. Miranda-Filloy, Jordi Monfort, Norberto Ortego-Centeno, Mercedes Pérez-Conesa, Sergio Prieto-González, Enrique Raya, Raquel Ríos Fernández, Julio Sánchez-Martín, Bernardo Sopeña, Laura Tío, and Ainhoa Unzurrunzaga.

The members of the "*UKGCA Consortium*" include Andrew Gough, John D. Isaacs, Michael Green, Neil McHugh, Lesley Hordon, Sanjeet Kamath, Mohammed Nisar, Yusuf Patel, Cee-Seng Yee, Robert Stevens, Pradip Nandi, Anupama Nandagudi, Stephen Jarrett, Charles Li, Sarah Levy, Susan Mollan, Abdel Salih, Oliver Wordsworth, Emma Sanders, Esme Roads, Anne Gill, Lisa Carr, Christine Routledge, Karen Culfear, Asanka Nugaliyadde, Lynne James, Jenny Spimpolo, Andy Kempa, Felicity Mackenzie, Rosanna Fong, Genessa Peters, Bridie Rowbotham, Zahira Masqood, Jane Hollywood, Prisca Gondo, Rose Wood, Steve Martin, Lubna Haroon Rashid, James I. Robinson, Mike Morgan, Louise Sorensen, and John Taylor.

The members of the "*Vasculitis Clinical Research Consortium*" are Simon Carette, Sharon Chung, David Cuthbertson, Lindsy J. Forbess, Ora Gewurz-Singer, Gary S. Hoffman, Curry L. Koening, Kathleen M. Maksimowicz-McKinnon, Carol A. McAlear, Larry W. Moreland, Christian Pagnoux, Philip Seo, Ulrich Specks, Robert F. Spiera, Antoine Sreih, Kenneth J. Warrington, and Michael Weisman.

Acknowledgements

The authors thank Sofía Vargas, Sonia García, Gema Robledo, Lubna Haroon Rashid and Steve Martin for technical support, and the participants for their collaboration. The Norwegian Systemic Vasculitis and Connective Tissue Disease Registry (NOSVAR) at Oslo University Hospital is acknowledged for providing data on the Norwegian cases.

FDC was supported by Instituto de Salud Carlos III (ISCIII), Spain, through the RETICS Program RD12/0009/0004 (RIER), and by the Ramón y Cajal programme of the Spanish Ministry of Economy and Competitiveness through the grant RYC-2014-16458. JM was founded by Junta de Andalucía through PAI group CTS180. 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).

The UKGCA Consortium has been funded by Research into Ageing and The Wellcome Trust and is currently supported by the National Institute of Health Research (Clinician Scientist Fellowship for SLM, Leeds Musculoskeletal Biomedical Research Unit (AWM and JHB) and Diagnostic Evidence Co-operative (AWM)), the Medical Research Council and the Ann Wilks Memorial Fund. This article presents independent research funded in part by the National Institute for Health Research (NIHR). The views expressed are those of the authors and not necessarily those of the UK National Health Service, the NIHR or the Department of Health.

Competing financial interests

The authors declare no competing financial interests of direct relevance to this manuscript.

Web Resources

The URLs for data presented herein are as follows:

1000 Genomes, http://www.1000genomes.org/

BEAGLE, http://faculty.washington.edu/browning/beagle/beagle.html

CADD, http://cadd.gs.washington.edu/

CaTS, http://www.sph.umich.edu/csg/abecasis/CaTS/

GEC, http://statgenpro.psychiatry.hku.hk/gec/

GEO, http://www.ncbi.nlm.nih.gov/geo/

GO, http://geneontology.org/

ENCODE, https://www.genome.gov/encode/

HaploReg, http://www.broadinstitute.org/mammals/haploreg/haploreg.php

IMPUTE2, http://mathgen.stats.ox.ac.uk/impute/impute_v2.html

LocusZoom, http://csg.sph.umich.edu/locuszoom/

NCBI, http://www.ncbi.nlm.nih.gov/

Online Mendelian Inheritance in Man (OMIM), http://omim.org/

PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/

PolyPhen-2, http://genetics.bwh.harvard.edu/pph2/

RegulomeDB, http://regulome.stanford.edu/

Roadmap Epigenomics, http://www.roadmapepigenomics.org/

STRING, http://string-db.org/

References

1. Manolio, T.A. (2013). Bringing genome-wide association findings into clinical use. Nature reviews Genetics 14, 549-558.

2. Jennette, J.C., Falk, R.J., Bacon, P.A., Basu, N., Cid, M.C., Ferrario, F., Flores-Suarez, L.F., Gross, W.L., Guillevin, L., Hagen, E.C., et al. (2013). 2012 revised International Chapel Hill Consensus Conference Nomenclature of Vasculitides. Arthritis and rheumatism 65, 1-11.

3. Gonzalez-Gay, M.A., Barros, S., Lopez-Diaz, M.J., Garcia-Porrua, C., Sanchez-Andrade, A., and Llorca, J. (2005). Giant cell arteritis: disease patterns of clinical presentation in a series of 240 patients. Medicine 84, 269-276.

4. Gonzalez-Gay, M.A., Vazquez-Rodriguez, T.R., Lopez-Diaz, M.J., Miranda-Filloy, J.A., Gonzalez-Juanatey, C., Martin, J., and Llorca, J. (2009). Epidemiology of giant cell arteritis and polymyalgia rheumatica. Arthritis and rheumatism 61, 1454-1461.

5. Carmona, F.D., Gonzalez-Gay, M.A., and Martin, J. (2014). Genetic component of giant cell arteritis. Rheumatology (Oxford) 53, 6-18.

6. Serrano, A., Marquez, A., Mackie, S.L., Carmona, F.D., Solans, R., Miranda-Filloy, J.A., Hernandez-Rodriguez, J., Cid, M.C., Castaneda, S., Morado, I.C., et al. (2013). Identification of the PTPN22 functional variant R620W as susceptibility genetic factor for giant cell arteritis. Annals of the rheumatic diseases 72, 1882-1886.

7. Marquez, A., Hernandez-Rodriguez, J., Cid, M.C., Solans, R., Castaneda, S., Fernandez-Contreras, M.E., Ramentol, M., Morado, I.C., Narvaez, J., Gomez-Vaquero, C., et al. (2014). Influence of the IL17A locus in giant cell arteritis susceptibility. Annals of the rheumatic diseases 73, 1742-1745.

8. Carmona, F.D., Mackie, S.L., Martin, J.E., Taylor, J.C., Vaglio, A., Eyre, S., Bossini-Castillo, L., Castaneda, S., Cid, M.C., Hernandez-Rodriguez, J., et al. (2015). A large-scale genetic analysis reveals a strong contribution of the HLA class II region to giant cell arteritis susceptibility. American journal of human genetics 96, 565-580.

9. Hunder, G.G., Bloch, D.A., Michel, B.A., Stevens, M.B., Arend, W.P., Calabrese, L.H., Edworthy, S.M., Fauci, A.S., Leavitt, R.Y., Lie, J.T., et al. (1990). The American College of Rheumatology 1990 criteria for the classification of giant cell arteritis. Arthritis and rheumatism 33, 1122-1128.

10. Barrett, J.H., Iles, M.M., Harland, M., Taylor, J.C., Aitken, J.F., Andresen, P.A., Akslen, L.A., Armstrong, B.K., Avril, M.F., Azizi, E., et al. (2011). Genome-wide association study identifies three new melanoma susceptibility loci. Nature genetics 43, 1108-1113.

11. Bentham, J., Morris, D.L., Cunninghame Graham, D.S., Pinder, C.L., Tombleson, P., Behrens, T.W., Martin, J., Fairfax, B.P., Knight, J.C., Chen, L., et al. (2015). Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. Nature genetics 47, 1457-1464.

12. Consortium, W.T.C.C. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447, 661-678.

13. Dubois, P.C., Trynka, G., Franke, L., Hunt, K.A., Romanos, J., Curtotti, A., Zhernakova, A., Heap, G.A., Adany, R., Aromaa, A., et al. (2010). Multiple common variants for celiac disease influencing immune gene expression. Nature genetics 42, 295-302.

14. Gregersen, P.K., Kosoy, R., Lee, A.T., Lamb, J., Sussman, J., McKee, D., Simpfendorfer, K.R., Pirskanen-Matell, R., Piehl, F., Pan-Hammarstrom, Q., et al.

(2012). Risk for myasthenia gravis maps to a (151) Pro-->Ala change in TNIP1 and to human leukocyte antigen-B*08. Annals of neurology 72, 927-935.

15. Radstake, T.R., Gorlova, O., Rueda, B., Martin, J.E., Alizadeh, B.Z., Palomino-Morales, R., Coenen, M.J., Vonk, M.C., Voskuyl, A.E., Schuerwegh, A.J., et al. (2010). Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. Nature genetics 42, 426-429.

16. Rotger, M., Dang, K.K., Fellay, J., Heinzen, E.L., Feng, S., Descombes, P., Shianna, K.V., Ge, D., Gunthard, H.F., Goldstein, D.B., et al. (2010). Genome-wide mRNA expression correlates of viral control in CD4+ T-cells from HIV-1-infected individuals. PLoS pathogens 6, e1000781.

17. Stahl, E.A., Raychaudhuri, S., Remmers, E.F., Xie, G., Eyre, S., Thomson, B.P., Li, Y., Kurreeman, F.A., Zhernakova, A., Hinks, A., et al. (2010). Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nature genetics 42, 508-514.

18. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics 81, 559-575.

19. Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS genetics 5, e1000529.

20. Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature 526, 68-74.

21. Skol, A.D., Scott, L.J., Abecasis, G.R., and Boehnke, M. (2006). Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nature genetics 38, 209-213.

22. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., and Willer, C.J. (2010). LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 26, 2336-2337.

23. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nature methods 7, 248-249.

24. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nature genetics 46, 310-315.

25. Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S., et al. (2012). Annotation of functional variation in personal genomes using RegulomeDB. Genome research 22, 1790-1797.

26. Ward, L.D., and Kellis, M. (2016). HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic acids research 44, D877-881.

27. (2015). Gene Ontology Consortium: going forward. Nucleic acids research 43, D1049-1056.

28. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature genetics 25, 25-29.

29. Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., and Narechania, A. (2003). PANTHER: a library of protein families and subfamilies indexed by function. Genome research 13, 2129-2141.

30. Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic acids research 43, D447-452.

31. Li, M.X., Yeung, J.M., Cherny, S.S., and Sham, P.C. (2012). Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. Human genetics 131, 747-756.

32. (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348, 648-660.

33. Carmona, F.D., Martin, J., and Gonzalez-Gay, M.A. (2016). New insights into the pathogenesis of giant cell arteritis and hopes for the clinic. Expert review of clinical immunology 12, 57-66.

34. Carmona, F.D., Martin, J., and Gonzalez-Gay, M.A. (2015). Genetics of vasculitis. Current opinion in rheumatology 27, 10-17.

35. Law, R.H., Abu-Ssaydeh, D., and Whisstock, J.C. (2013). New insights into the structure and function of the plasminogen/plasmin system. Current opinion in structural biology 23, 836-841.

36. Miles, L.A., and Parmer, R.J. (2013). Plasminogen receptors: the first quarter century. Seminars in thrombosis and hemostasis 39, 329-337.

37. Moser, T.L., Stack, M.S., Asplin, I., Enghild, J.J., Hojrup, P., Everitt, L., Hubchak, S., Schnaper, H.W., and Pizzo, S.V. (1999). Angiostatin binds ATP synthase on

the surface of human endothelial cells. Proceedings of the National Academy of Sciences of the United States of America 96, 2811-2816.

38. Ly, K.H., Regent, A., Tamby, M.C., and Mouthon, L. (2010). Pathogenesis of giant cell arteritis: More than just an inflammatory condition? Autoimmunity reviews, 635-645.

39. Hao, J., Wang, C., Gou, S.J., Zhao, M.H., and Chen, M. (2014). The association between anti-plasminogen antibodies and disease activity in ANCA-associated vasculitis. Rheumatology (Oxford) 53, 300-306.

40. Lyons, P.A., Rayner, T.F., Trivedi, S., Holle, J.U., Watts, R.A., Jayne, D.R., Baslund, B., Brenchley, P., Bruchfeld, A., Chaudhry, A.N., et al. (2012). Genetically distinct subsets within ANCA-associated vasculitis. The New England journal of medicine 367, 214-223.

41. Schaefer, A.S., Bochenek, G., Jochens, A., Ellinghaus, D., Dommisch, H., Guzeldemir-Akcakanat, E., Graetz, C., Harks, I., Jockel-Schneider, Y., Weinspach, K., et al. (2015). Genetic evidence for PLASMINOGEN as a shared genetic risk factor of coronary artery disease and periodontitis. Circulation Cardiovascular genetics 8, 159-167.

42. Sadovnick, A.D., Traboulsee, A.L., Bernales, C.Q., Ross, J.P., Forwell, A.L., Yee, I.M., Guillot-Noel, L., Fontaine, B., Cournu-Rebeix, I., Alcina, A., et al. (2016). Analysis of Plasminogen Genetic Variants in Multiple Sclerosis Patients. G3 (Bethesda) 6, 2073-2079.

43. Myllyharju, J. (2003). Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis. Matrix biology : journal of the International Society for Matrix Biology 22, 15-24.

44. Gilkes, D.M., Bajpai, S., Chaturvedi, P., Wirtz, D., and Semenza, G.L. (2013). Hypoxia-inducible factor 1 (HIF-1) promotes extracellular matrix remodeling under

hypoxic conditions by inducing P4HA1, P4HA2, and PLOD2 expression in fibroblasts. The Journal of biological chemistry 288, 10819-10829.

45. Deng, W., Feng, X., Li, X., Wang, D., and Sun, L. (2016). Hypoxia-inducible factor 1 in autoimmune diseases. Cellular immunology 303, 7-15.

46. Dimova, E.Y., Samoylenko, A., and Kietzmann, T. (2004). Oxidative stress and hypoxia: implications for plasminogen activator inhibitor-1 expression. Antioxidants & redox signaling 6, 777-791.

47. Stranger, B.E., Nica, A.C., Forrest, M.S., Dimas, A., Bird, C.P., Beazley, C., Ingle, C.E., Dunning, M., Flicek, P., Koller, D., et al. (2007). Population genomics of human gene expression. Nature genetics 39, 1217-1224.

48. Westra, H.J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E., et al. (2013). Systematic identification of trans eQTLs as putative drivers of known disease associations. Nature genetics 45, 1238-1243.

49. Lappalainen, T., Sammeth, M., Friedlander, M.R., t Hoen, P.A., Monlong, J., Rivas, M.A., Gonzalez-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G., et al. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. Nature 501, 506-511.

50. Kaiser, M., Younge, B., Bjornsson, J., Goronzy, J.J., and Weyand, C.M. (1999). Formation of new vasa vasorum in vasculitis. Production of angiogenic cytokines by multinucleated giant cells. The American journal of pathology 155, 765-774.

Figure legends.

Figure 1. Manhattan plot representation of the GWAS results. The -log10 of the inverse variance-weighted meta-analysis P-values are plotted against their physical chromosomal position. The red line represents the genome-wide level of significance (P < 5E-08). The most relevant associations are highlighted.

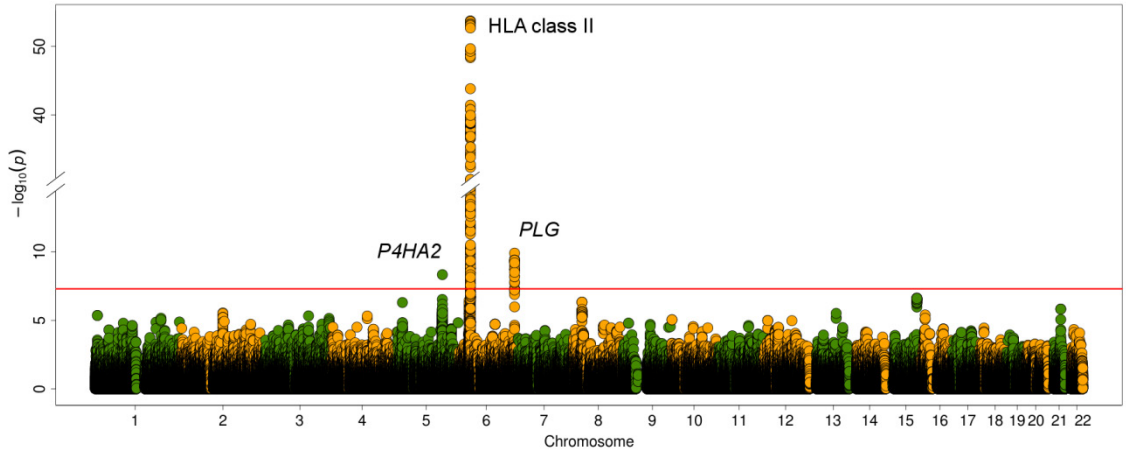


Figure 2. Regional plots of the associated *loci* with GCA outside the HLA Region in the overall meta-analysis. (A) Plasminogen (*PLG*) region. (B) prolyl 4-hydroxylase subunit alpha 2 (*P4HA2*) region. Lead variants are highlighted in violet.

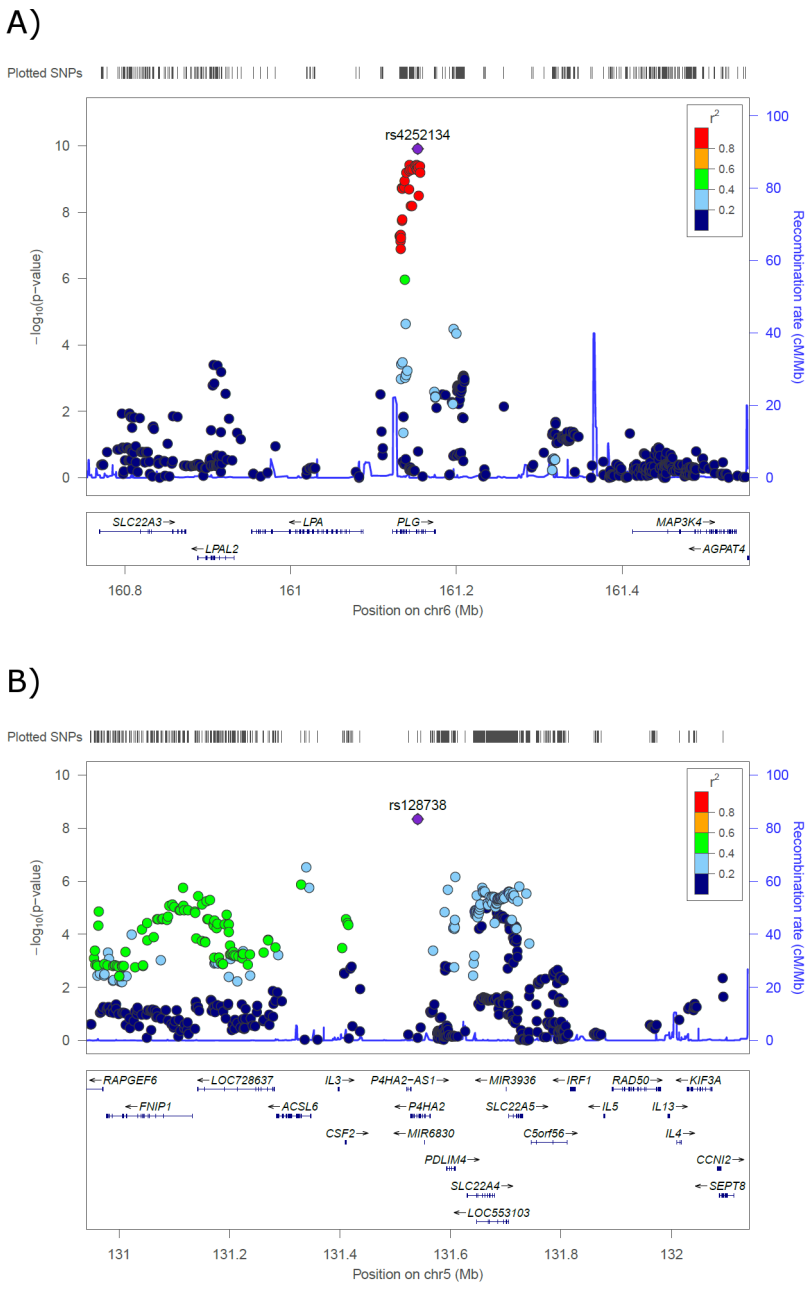


Figure 3. Interaction network formed for GCA risk *loci*. GWAS genes of this study and those previously identified through Immunochip and consistent candidate gene studies were included in the analysis. STRING database was used to look for both direct and indirect interactions amongst selected genes. The width of the blue lines indicates the reliability of each interaction.

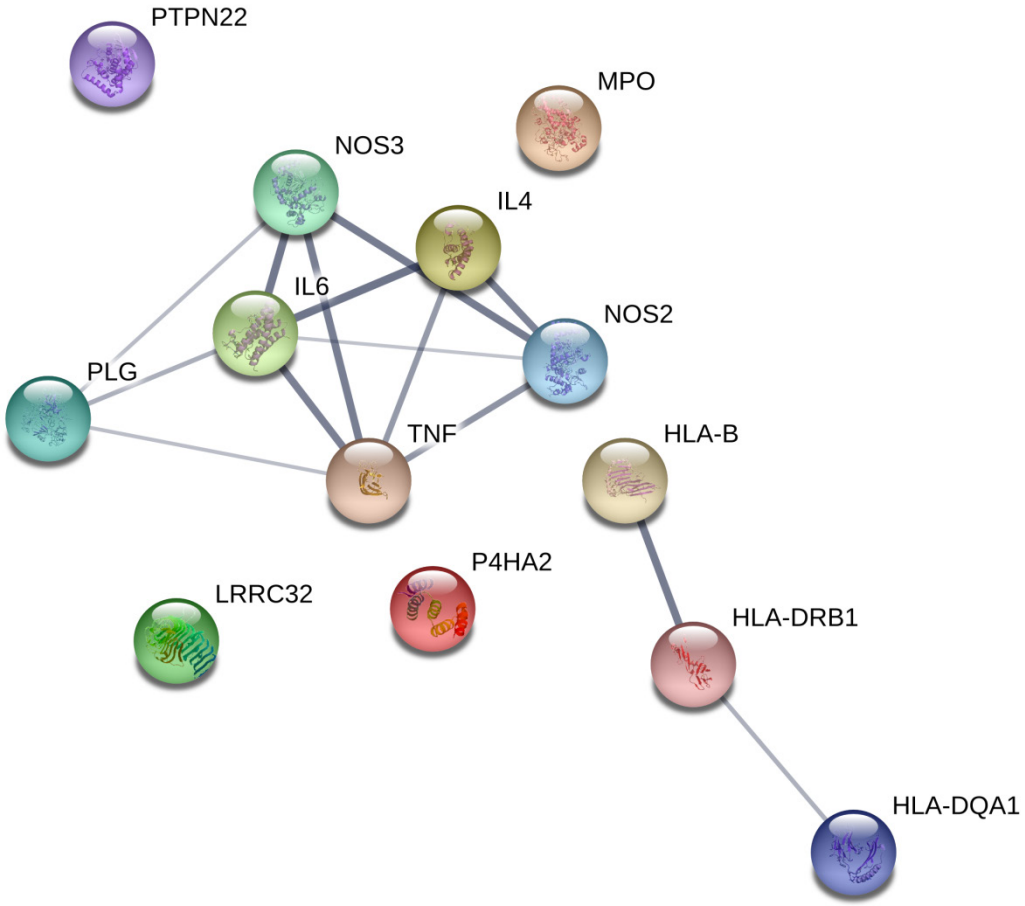


Table 1. Independent association signals with giant cell arteritis susceptibility at the genome-wide level of significance.

				META-ANALYSIS				MINOR ALLELE FREQUENCY (GCA/CTRL)[a]									
SNP	Location (GRCh37)	*Locus*	Change	P-value	OR [CI 95%]	P(Q)	I^2	Spain	United Kingdom	Italy	North America	Germany	France	Norway	Netherlands	Switzerland	Ireland
rs9268905	6:32432077	*HLA-DRA / HLA-DRB1*	C<G	**1.94E-54**	1.79 [1.67-1.93]	0.75	0	0.46/0.32	0.52/0.37	0.33/0.23	0.48/0.32	0.39/0.28	0.48/0.29	0.47/0.33	0.37/0.28	0.42/0.34	0.45/0.34
rs9275592	6:32680620	*HLA-DQA1 / HLA-DQA2*	T<G	**1.14E-40**	2.08 [1.87-2.32]	0.98	0	0.15/0.08	0.18/0.11	0.09/0.05	0.20/0.09	0.18/0.10	0.18/0.10	0.28/0.16	0.18/0.10	0.19/0.11	0.15/0.10
rs4252134	6:161153527	*PLG*	C<T	**1.23E-10**	1.28 [1.19-1.39]	0.04	49	0.33/0.29	0.36/0.28	0.36/0.33	0.32/0.30	0.32/0.28	0.34/0.33	0.38/0.23	0.38/0.28	0.47/0.28	0.41/0.30
rs128738	5:131540875	*P4HA2*	T<G	**4.60E-09**	1.32 [1.20-1.45]	0.49	0	0.19/0.14	0.22/0.17	0.16/0.14	0.17/0.16	0.20/0.17	0.18/0.16	0.22/0.18	0.23/0.15	0.18/0.16	0.16/0.18

GCA, giant cell arteritis; CTRL, controls; GRCh37, genome reference consortium human genome build 37; OR, per-allele odds ratio for the minor allele; CI, confidence interval; Q, Cochran's Q test.

[a]N (GCA/CTRL): Spain = 805/1,323, United Kingdom = 352/2,965, Italy = 271/960, North America = 176/1,181, Germany = 160/667, France = 114/488, Norway = 104/121, Netherlands = 69/638, Switzerland = 46/500, Ireland = 37/282, Combined = 2,134/9,125.

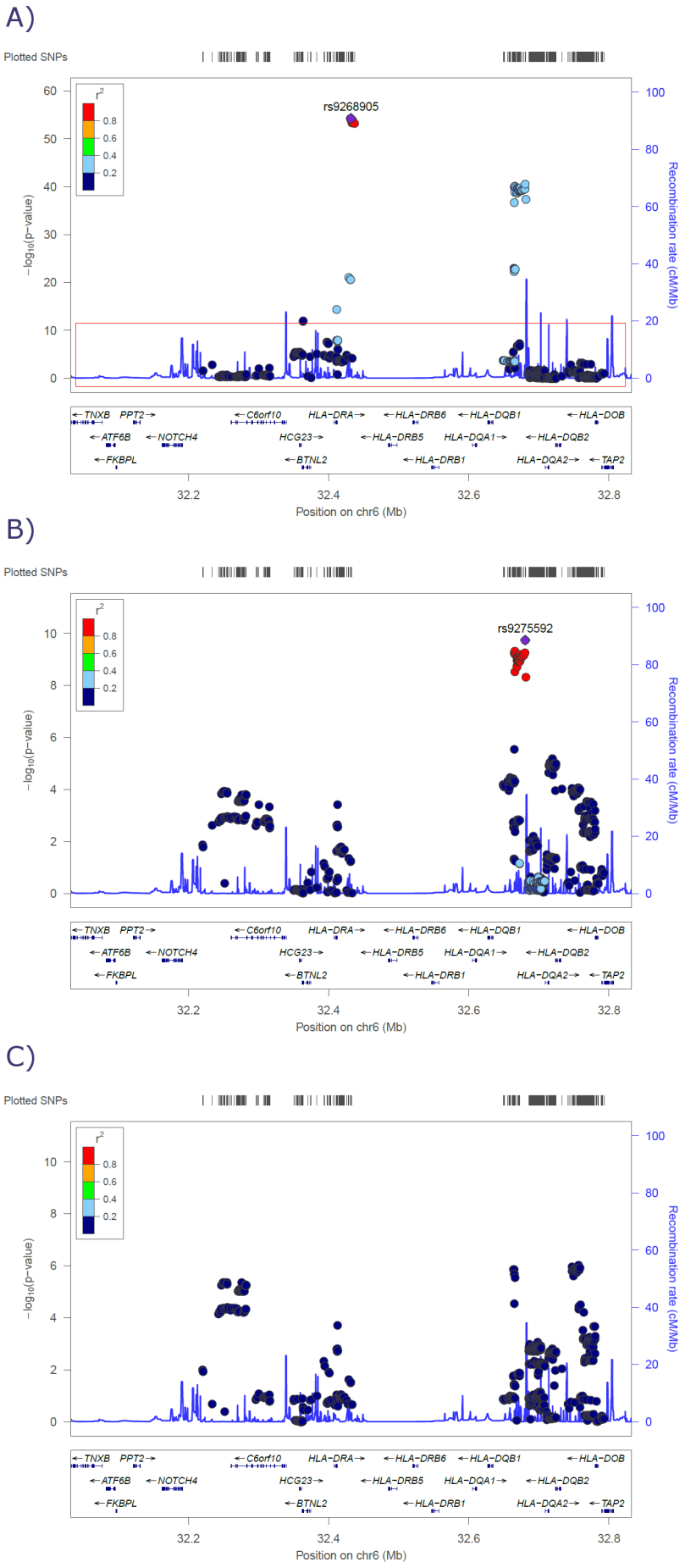
Table 2. Prioritised variants of the giant cell arteritis-associated non-HLA regions accordingly to the functional annotations of the lead signals and their proxies using ENCODE data. Chromatin marks in blood tissues are highlighted in bold.

						Regulatory chromatin marks			**Protein binding and regulatory motifs**		**GRASP QTL and eQTL annotations**			
SNP	Location (GRCh37)	*Locus*	Distance from hit (bp)	r2	RDB score	Promoter histone marks	Enhancer histone marks	DNase hypersensitivity	Proteins bound (ChIP-seq)	Regulatory motifs altered (PWM)	GRASP QTL traits	P-value	Ref (PMID)	eQTL hits
rs4252134	6:161153527	*PLG* (intronic)	0	-	6	-	-	-	-	-	Gene expression of DDR1 in peripheral blood monocytes.	3.31E-06	20502693	
rs4252126	6:161152294	*PLG* (intronic)	1,233	0.995	1f	YES	-	YES	CTCF	Pax-6				
rs4252135	6:161154232	*PLG* (intronic)	705	1.000	1f	YES	**YES**	**YES**	CTCF, RAD21, SMC3, ZNF143, NFKB, FOXA1, FOXA2, ZNF263	COMP1	Gene expression of DDR1 in peripheral blood monocytes.	2.92E-06	20502693	
rs34126283	6:161130002	*PLG* (intronic)	23,525	0.920	3a	YES	YES	YES	GATA2	Cdc5, E2F, Evi-1, HDAC2, PLZF, Pou2f2, TATA				
rs4252125	6:161152240	*PLG* (missense)	1,287	0.995	3a	YES	-	YES	CTCF, RAD21	AP-3	Gene expression of UNQ9391 in liver.	4.09E-14	22006096	
rs128738	5:131540875	*P4HA2* (intronic)	0	-	ND	YES	**YES**	-	-	BDP1				24
rs101194	5:131515413	*P4HA2* (intergenic)	25,462	0.901	3a	YES	**YES**	YES	GATA2, P300	CEBPB	Gene expression of *P4HA2* in lymphoblastoid cell lines. Gene expression of SLC22A5 in peripheral blood monocytes.	2.00E-07 1.16E-08	17873874 20502693	30
rs152054	5:131519540	*P4HA2* (intergenic)	21,335	0.967	3a	**YES**	**YES**	**YES**	CFOS	BAF155, BCL, Egr-1, Ets, FEV, GATA, NERF1a, Nrf-2, PU.1, Pax-5, SETDB1, SIX5, Tel2, Znf143				25
rs152051	5:131539025	*P4HA2* (intronic)	1,850	1.000	3a	YES	**YES**	**YES**	YY1, NFKB	AIRE, Foxa, GR, VDR	Gene expression of P4HA2 in CEU-CHB-JPT-YRI lymphoblastoid cell lines.	4.66E-09	17873874	28
rs156023	5:131545168	*P4HA2* (intronic)	4,293	0.967	3a	**YES**	**YES**	-	-	AIRE, LBP-1, LBP-9				22

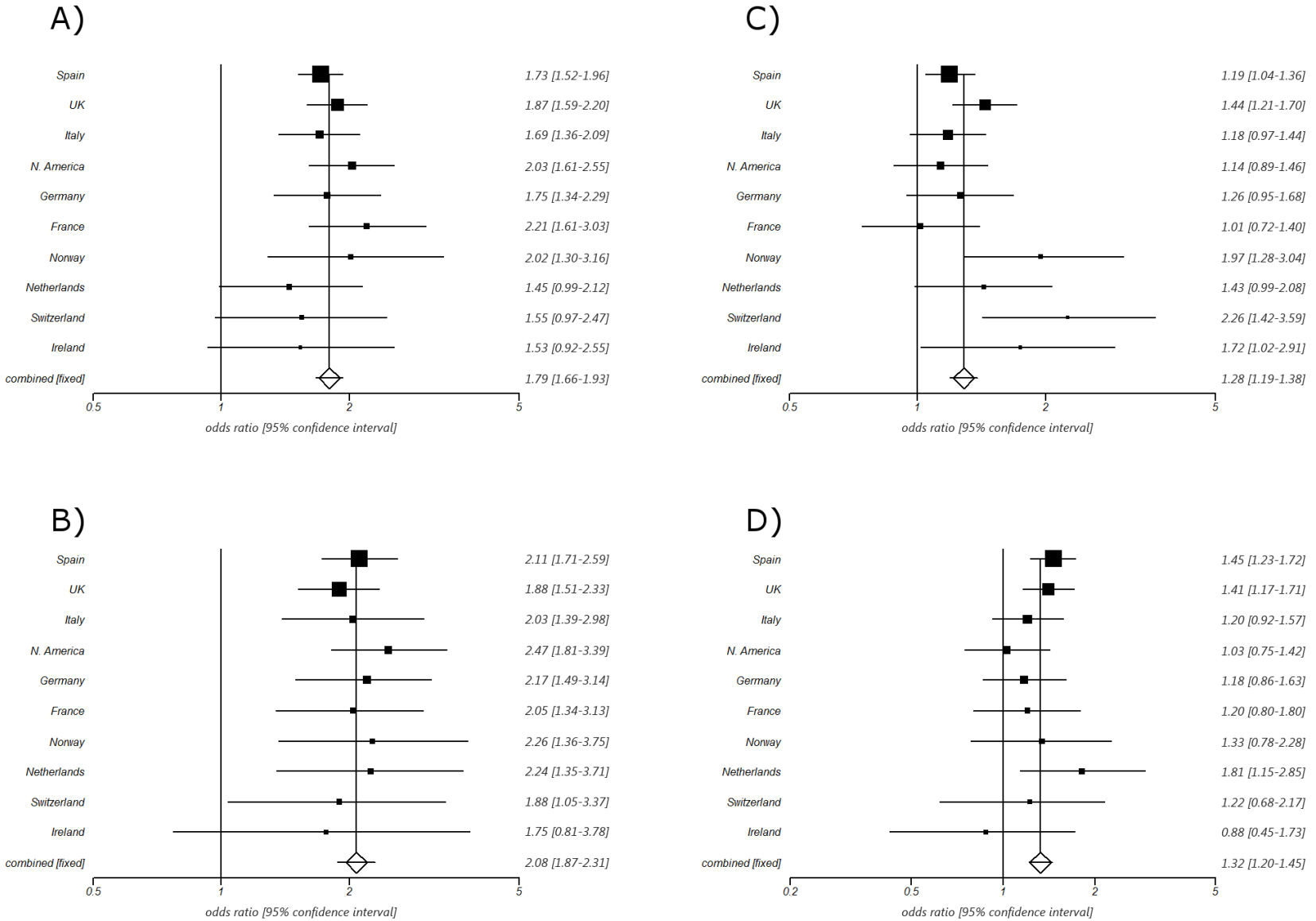
SNP, single nucleotide polymorphism; GRCh37, genome reference consortium human genome build 37; bp, base pair; RDB, Regulome database; PWM, position weight matrix ID; GRASP, genome-wide repository of associations between SNPs and phenotypes; PMID, PubMed identifier; eQTL, expression quantitative trait *loci.*

RDB scores: 1f, eQTL + TF binding / DNase peak; 3a, TF binding + any motif + DNase peak; 6, other. ND, no data.

Supplemental Figure 1. Manhattan plot representation of the step-wise conditional logistic regression of the HLA class II region. (A) Unadjusted test. (B) Results of the dependence analysis after adjusting for rs9268905. (C) Results of the dependence analysis after adjusting for rs9268905 and rs9275592. Lead variants are highlighted in violet. The red inset in (A) represents the scale in (B) and (C).



Supplemental Figure 2. Forest plot representations of the meta-analyses of the lead variants associated with giant cell arteritis in this study. (A) *HLA-DRA* / *HLA-DRB1* rs9268905. (B) *HLA-DQA1* / *HLA/DQA2* rs9275592. (C) *PLG* rs4252134. (D) *P4HA2* rs128738. Odds ratios with their 95% confidence intervals are shown for each individual population as well as for the overall meta-analysis.



Supplemental Figure 3. Manhattan plot representation of the step-wise logistic regression of the plasminogen region. (A) Unadjusted test. (B) Results of the dependence analysis after adjusting for rs4252134. Lead variants are highlighted in violet.