Concise report

HLA allele variation as a potential explanation for the geographical distribution of granulomatosis with polyangiitis

Richard A. Watts1,2, Alex J. MacGregor2 and Sarah L. Mackie3

Abstract

Objective. Granulomatosis with polyangiitis (GPA) is a rare autoimmune systemic vasculitis considered to result from the interaction of environmental factors with a genetically predisposed host. The HLA-DPB1*0401 allele, the PI*Z allele of the gene encoding α1-antitrypsin (SERPINA1) and the proteinase 3 (PRTN3) gene have been associated with GPA. The incidence of GPA is lower in non-Caucasian populations and has been associated with higher latitude. Our aim was to determine whether variation in population carrier frequency of the HLA-DPB1*0401 and PI*Z alleles could explain in part the variation in GPA incidence between countries.

Methods. We systematically identified published reports on the incidence of GPA and used previously published data on the frequency of HLA-DPB1*0401 and PI*Z alleles. The relationship between GPA incidence, latitude and population HLA-DPB1*0401 and PI*Z allele frequencies was assessed by linear regression.

Results. On multivariate analysis GPA incidence was associated with HLA-DPB1*0401 allele frequency (P = 0.001) but not with PI*Z allele frequency or latitude.

Conclusion. HLA-DPB1*0401 is a GPA susceptibility allele and HLA-DPB1*0401 population allele frequencies may help explain variations in GPA incidence described in the literature.

Key words: granulomatosis with polyangiitis, epidemiology, genetics, HLA, vasculitis.

Introduction

Granulomatosis with polyangiitis (GPA) is one of the ANCA-associated vasculitides (AAVs), GPA is a relatively uncommon condition characterized by the development of necrotizing vasculitis (in particular affecting the kidneys), granulomata formation and upper respiratory tract involvement. The presence of ANCA directed against proteinase 3 (PR3) in serum is associated with GPA. The aetiopathogenesis of GPA is unknown but is widely thought to involve the interaction of environmental factors with a genetically predisposed host.

Evidence of a genetic risk comes from descriptive epidemiology studies that suggest that the prevalence of GPA is greater in white Caucasian populations and significantly lower in non-Caucasian populations from South East Asia, sub-Saharan Africa and the Indian subcontinent [1–4]. Direct comparison between a UK white Caucasian population and a Japanese population indicated that while the overall incidence of AAV was the same, GPA was more common in the UK than in Japan [5]. Higher latitude (reflecting exposure to ultraviolet radiation) has been proposed as a possible explanation for the greater prevalence of GPA in Northern European populations [6, 7].

Candidate gene studies suggest associations between GPA and HLA-DPB1*0401 allele [odds ratio (OR) 3.91 and 3.01] [8], PTNP22 620W allele (OR 2.01) [9], CTL4A and the PI*Z allele of the α1-antitrypsin gene (SERPINA1) [10]. The serpin A1 gene is of particular interest, as it encodes α1-antitrypsin, a serine protease that has PR3 as one of its substrates. These studies have generally been conducted in white European (or European descent) populations.
In Han Chinese, a population in which GPA is relatively less common, PR3-ANCA-positive vasculitis has been associated with HLA-DRB1*1202 [11]. HLA-DRB1*1501 has been associated with PR3-ANCA-positive vasculitis in both African Americans (OR 73.3) and Caucasians (OR 2.2) [4].

A genome-wide association study (GWAS) of AAVs in European Caucasian patients reported that the strongest genetic associations were with the antigenic specificity of ANCA rather than with the clinical syndrome [12]. Anti-PR3-ANCA was associated with HLA-DP, SERPINA1 and PRTN3 (PRTN3 encodes PR3). The immune response against the autoantigen PR3 is a central aetiological feature of PR3-ANCA-associated vasculitis. Haplotype analysis suggested that the causal variant at SERPINA1/SERPINA11 is either the z allele of SERPINA1 or in close linkage disequilibrium with it. The precise variant in HLA-DP was not determined, but there was evidence of only a single genetic association in this region.

The global distribution of the Pi*Z allele of SERPINA1 have been well studied because of the well-known association between α1-antitrypsin deficiency and chronic pulmonary obstructive disease and hepatic disease (hepatitis, cirrhosis and hepatocellular carcinoma). De Serres and Blanco [13] collated all the available data on the global prevalence of the Pi*Z allele and showed that there are striking geographical differences in prevalence. The Pi*Z allele is most frequently found in Scandinavia (7–27 per 1000), Western and central Europe (7–30 per 1000) and those countries colonized by Europeans (North America, 10–12 per 1000; Australia and New Zealand, 12–26 per 1000). The Pi*Z alleles are rare in China, Japan and Indonesia (0–1 per 1000). A recent study has suggested that the incidence of AAVs is 10-fold greater in those with severe α1-antitrypsin deficiency [14].

The HLA-DRB1*0401 allele also shows variation worldwide [15]. There is a fairly consistent allele frequency in Europe of ~0.36–0.47. The allele is much less frequent in Japan (0.050), China (0.095, Han Cantonese) and US African Americans (0.104); these are three populations where PR3-ANCA vasculitis is very much less common than in Europe. In a small population of Han Chinese, PR3-ANCA vasculitis was associated with HLA-DRB1*1202, which is relatively more common in that population than in Europe [11]. In the USA, HLA-DRB1*15 was associated with PR3-ANCA vasculitis in both Caucasians and African Americans [4]. There was a difference in the genotype associations, as HLA-DRB1*1501 was associated with vasculitis in African Americans whereas HLA-DRB1*1503 was underrepresented. The HLA-DRB1*1501 allelic variant is of Caucasian descent, whereas HLA-DRB1*1503 is of African American descent. We hypothesized that the occurrence of GPA in different populations is determined by the relative occurrence of HLA-DRB1*0401 and Pi*Z alleles in each population.

Methods

A literature search in PubMed was conducted with combinations of the medical subject heading terms granulomatosis with polyangiitis, Wegener’s granulomatosis and epidemiology. Hand searching was also performed in the reference lists of retrieved articles, in review articles and in textbooks. We chose to use studies reporting incidence rather than prevalence of GPA, because there are more studies reporting incidence than prevalence. Studies were included if they were available in full text and included an estimate of the annual incidence, time period of the study, method of case definition, population studied and geographical location of the study. Studies that appeared to report duplicate or overlapping populations were excluded. Studies completing recruitment before 1990 were excluded in case of time trends in the incidence of GPA, and because the quality of the reporting was generally lower for the older studies. Where more than one report existed for a single country (unless in ethnically distinct populations), then the one with a later average period of recruitment was preferred. To identify ethnically matched HLA-DRB1*0401 population allele frequencies for each geographical region (city, province or country) identified, the Allele Frequency Net Database (http://www.allelefrequencies.net) was consulted. Where data for the relevant geographical region on the Allele Frequency Net Database were unavailable or based on one small study only, a search in PubMed was conducted to identify relevant articles (usually the control cohort of case–control genetic association studies for other diseases). Search terms were HLA-DRB1*0401 (limit humans and $(<country>$ or $<$nationality$>$). The best available match to the geographical region specified in each incidence paper was sought. Where possible, at least two studies were identified to check the consistency of the population allele frequencies. Where inconsistency was seen, further studies were sought and the recruitment method of the control population in each study was reviewed. Data on Pi*Z allele frequencies were obtained from previously published literature collating its distribution [13].

Geographical latitude was estimated for each incidence study by identifying the coordinates of the relevant geographical region, usually the reporting centre. Google Maps (https://maps.google.com) was used to identify latitude coordinates. Data were displayed graphically and linear regression was used to identify relationships between HLA-DRB1*0401 allele frequencies, geographical latitude and the incidence of GPA in different countries. Due to heterogeneity of the GPA incidence studies and the limited number of populations with available data, formal meta-regression was not performed. Statistical significance was taken as P < 0.05. Analyses were performed using SPSS (IBM, Armonk, NY, USA).

Results

Nineteen studies were included for which there was adequate information about classification method and denominator population: 11 were from Europe, 2 from the Americas and 6 from Australasia. For HLA, 15 data points were included in the dataset. Within this dataset HLA-DRB1*0401 was significantly more common at higher
latitudes ($r = 0.804, P = 0.001$). For PI*Z, 19 data points were included. For PI*Z there was no significant association with latitude ($r = 0.263, P = 0.139$).

Significant predictors of GPA incidence on univariable analysis were HLA-DPB1*0401 population allele frequency ($\beta = 0.804, P < 0.0001$, adjusted $R^2 = 0.619$) and latitude ($\beta = 0.553, P = 0.014$, adjusted $R^2 = 0.265$). There was no association between PI*Z and GPA incidence ($\beta = 0.263, P = 0.277$, adjusted $R^2 = -0.014$). In multivariable analysis, only HLA-DPB1*0401 was a significant predictor of incidence.

**Discussion**

In this study we demonstrated that there is an association between HLA-DPB1*0401 and GPA occurrence, which is independent of both latitude and PI*Z. HLA-DPB1*0401 population allele frequencies may help explain variations in GPA incidence between populations described in the literature. We were unable to demonstrate an effect of PI*Z allele frequency on GPA occurrence. The genetic background of GPA has been shown to be strongly associated with HLA-DP, SERPINA1 and PRTN3. In the GWAS, the strongest association was with HLA-DP ($P = 6.2 \times 10^{-59}$) compared with SERPINA1 and PRTN3 ($P = 5.6 \times 10^{-12}$ and $2.6 \times 10^{-07}$) [12]. The published data from the GWAS do not include the precise variant in HLA-DP, but the candidate gene data suggest that the causal variant may be HLA-DPB1*0401. This study provides further epidemiological evidence for this association. The failure to demonstrate an effect of PI*Z allele frequency on the occurrence of GPA perhaps reflects the weaker association between the PI*Z allele and GPA compared with HLA-DP.

In the univariable analysis we confirmed the previously observed association between high latitude and GPA occurrence [7], but multivariable analysis suggested that this effect was due to the distribution of HLA-DPB1*0401. Other genetic factors are likely to be important in determining risk for GPA. In Han Chinese, HLA-DRB1*1202 may be a more important HLA risk factor than HLA-DPB1*0401 [11], and the same may be true in African Americans, where HLA-DRB1*1501 may be the important risk factor [4]. This is the first study to explore the relationship between reported GPA incidence in relation to population HLA-DPB1*0401 allele frequency, PI*Z allele frequency and geographical latitude.

The main limitations of the present study are, first, the small number of studies on the epidemiology of GPA and, secondly a lack of data on the genes of interest in these populations. The study included 19 studies, but these were not evenly distributed across different ethnicities, with the majority reporting on white Caucasian populations from Europe, North American and Australasia and only two studies coming from countries with a different ethnic/genetic background (Japan and Taiwan). This might explain the failure to demonstrate association with PI*Z across populations, as only the Japanese and Taiwanese populations have a low frequency of the PI*Z allele. Insufficient data on the distribution of the PRTN3 single nucleotide polymorphism (rs62132295) was available to permit this type of analysis.

The 19 studies used different classification criteria. All but two used either the ACR (1990) criteria [16] or the European Medicines Agency classification algorithm [17]. One study used the Chapel Hill Consensus Conference (1994) definition [18] and one used the International Classification of Diseases, 10th revision (ICD-10) classification. Classification using ANCA specificity was not possible because the published epidemiological studies do not provide sufficient data on ANCA specificity. Further epidemiological studies from a broad range of ethnicities are required to explore the possible genetic associations of GPA in greater detail.

**Rheumatology key messages**

- HLA-DPB1*0401 is a granulomatosis with polyangiitis susceptibility allele.
- HLA-DPB1*0401 population allele frequencies may help explain variations in granulomatosis with polyangiitis incidence.

**Acknowledgements**

R.A.W. was supported by the Norfolk & Suffolk Comprehensive Local Research Network. S.L.M. is funded by a Clinician Scientist Fellowship from the National Institute for Health Research (NIHR). The views expressed are those of the authors and not necessarily those of the National Health Service, NIHR or the Department of Health.

**Disclosure statement:** The authors have declared no conflicts of interest.

**References**


