Large scale whole-genome sequencing reveals the genetic architecture of primary membranoproliferative glomerulonephritis and C3 glomerulopathy

Adam P. Levine, Melanie M. Y. Chan, Omid Sadeghi-Alavijeh, Edwin K. S. Wong, H. Terence Cook, Sofie Ashford, Keren Carss, Martin T. Christian, Matthew Hall, Claire Harris, Paul McAlinden, Kevin J. Marchbank, Stephen D. Marks, Heather Maxwell, Karyn Megy, Christopher J. Penkett, Monika Mozere, Kathleen E. Stirrups, Salih Tuna, Julie Wessels, Deborah Whitehorn, MPGN/DDD/C3 Glomerulopathy Rare Disease Group, NIHR BioResource, Sally A. Johnson, Daniel P. Gale

Significance Statement

Primary MPGN, including C3 glomerulopathy, is a rare untreatable kidney disease. A minority of cases are familial, caused by mutations in complement genes, and non-familial cases have also been reported to harbor such mutations. To characterize the genetic architecture of this disease we analyzed whole-genome data from 165 PMG cases and 10,250 non-PMG control individuals (146 and 6,442, respectively that were unrelated and of European ancestry). We observed no significant enrichment of rare variants – in complement genes or exome-wide, but did observe a strong and statistically significant common variant association at the HLA locus which was replicated in an independent cohort. These findings imply that in most cases the disease is driven by autoimmunity rather than an underlying monogenic disorder of complement regulation.

Large scale whole-genome sequencing reveals the genetic architecture of primary membranoproliferative glomerulonephritis and C3 glomerulopathy

Adam P. Levine[1], Melanie M. Y. Chan[1], Omid Sadeghi-Alavijeh[1], Edwin K. S. Wong[2,3,4], H. Terence Cook[5], Sofie Ashford[6], Keren Carss[6,7], Martin T. Christian[8], Matthew Hall[9], Claire Harris[3], Paul McAlinden[3], Kevin J. Marchbank[3,4], Stephen D. Marks[10], Heather Maxwell[11], Karyn Megy[6,7], Christopher J. Penkett[6,7], Monika Mozere[1], Kathleen E. Stirrups[6,7], Salih Tuna[6,7], Julie Wessels[12], Deborah Whitehorn[6,7], MPGN/DDD/C3 Glomerulopathy Rare Disease Group[13]†, NIHR BioResource[6]†, Sally A. Johnson[3,4,14], Daniel P. Gale[1]

1. Department of Renal Medicine, University College London, London, UK
2. Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK
3. Institute of Cellular Medicine, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK
4. The National Renal Complement Therapeutics Centre, Royal Victoria Infirmary, Newcastle upon Tyne, UK
5. Centre for Inflammation Research, Department of Medicine, Imperial College London, London, UK
6. NIHR BioResource, Cambridge University Hospitals, Cambridge Biomedical Campus, Cambridge, UK
7. Department of Haematology, University of Cambridge, Cambridge Biomedical Campus, Cambridge, UK
8. Children's Renal and Urology Unit, Nottingham Children's Hospital, QMC, Nottingham University Hospitals NHS Trust, Nottingham, UK
9. Nottingham University Hospitals NHS Trust, Nottingham, UK
10. Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK
11. Royal Hospital for Children, NHS Greater Glasgow and Clyde, Glasgow, UK
12. University Hospitals of North Midlands NHS Trust, Stoke-on-Trent, UK
13. RaDaR (http://www.rarerenal.org), UK Renal Registry, Bristol, UK
14. Department of Paediatric Nephrology, Great North Children's Hospital, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK

† The members of the MPGN/DDD/C3 Glomerulopathy Rare Disease Group and the NIHR BioResource may be found in the Supplementary Appendix.

Running title: Whole genome sequencing in MPGN

Corresponding author: Dr Daniel P. Gale, UCL Department of Renal Medicine, Royal Free Hospital, Rowland Hill Street, London NW3 2P, United Kingdom, d.gale [at] ucl.ac.uk

Significance Statement

Primary MPGN, including C3 glomerulopathy, is a rare untreatable kidney disease. A minority of cases are familial, caused by mutations in complement genes, and non-familial cases have also been reported to harbor such mutations. To characterize the genetic architecture of this disease we analyzed whole-genome data from 165 PMG cases and 10,250 non-PMG control individuals (146 and 6,442, respectively that were unrelated and of European ancestry). We observed no significant enrichment of rare variants – in complement genes or exome-wide, but did observe a strong and statistically significant common variant association at the HLA locus which was replicated in an independent cohort. These findings imply that in most cases the disease is driven by autoimmunity rather than an underlying monogenic disorder of complement regulation.

Abstract

Background: Primary membranoproliferative glomerulonephritis (PMG), including C3 glomerulopathy (C3G), is a rare, untreatable kidney disease characterized by glomerular complement deposition. Familial C3G can be caused by complement gene mutations and rare variants in such genes have been reported in non-familial PMG.

Methods: We analyzed whole-genome sequence data from 165 PMG cases and 10,250 non-PMG control individuals (146 and 6,442, respectively that were unrelated and of European ancestry) as part of the National Institute of Health Research BioResource-Rare Diseases Study. We examined copy number, rare and common variants.

Results: We observed no significant enrichment of rare variants in cases in the candidate genes (6.8% versus 5.9% in controls, p=0.37) or exome-wide. However, a significant common variant locus was identified at 6p21.32 (rs35406322, p=$3.29×10^{-8}$, OR 1.93) overlapping the HLA locus. Imputation of HLA types mapped this signal to a haplotype incorporating DQA1*05:01, DQB1*02:01 and DRB1*03:01 (p=$1.21×10^{-8}$, OR 2.89). This finding was replicated by analysis of HLA serotypes in 338 individuals with membranoproliferative glomerulonephritis and 15,614 individuals with non-immune renal failure (p=$1.4×10^{-4}$, OR 1.43).

Conclusions: These findings challenge the paradigm of complement gene mutations typically causing PMG and implicate an underlying autoimmune mechanism in most cases.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Introduction

Membranoproliferative glomerulonephritis (MPGN) refers to inflammatory kidney disease in which there is increased glomerular mesangial matrix and cellularity, thickening of the capillary walls and deposition of immunoglobulins and/or complement. Such appearances can be seen when the immune system is chronically activated; the term primary MPGN (PMG) refers to those cases in which an underlying infectious, neoplastic or autoimmune disorder is not identified. PMG is divided into immune complex PMG (IC-PMG), where there is positive immunostaining for immunoglobulins and complement, and C3 glomerulopathy (C3G), where complement C3 is the predominant immunoprotein deposited. C3G is subdivided by electron microscopic appearances into C3 glomerulonephritis (C3GN) and dense deposit disease (DDD), in which there is characteristic dense transformation of the glomerular basement membrane[1].

PMG is rare, with incidence estimated at 3-5 per million population[2–4]. In most cases the cause is not known but familial C3G has been linked to genomic rearrangements in the Complement Factor H Related genes (*CFHR1-5*)[5–8], bi-allelic loss of function variants of Complement Factor H (*CFH*)[9], and an activating mutation of *C3*[10]. In addition, studies of non-familial cases of PMG have identified rare variants in these and other complement genes (previously associated with atypical hemolytic uremic syndrome, aHUS) in up to 40% of patients[11–14]. These findings, together with the almost invariable presence of C3 in the glomerulus, have implicated complement alternative pathway activation as a key causal mechanism and testing for a complement gene mutations is currently recommended in C3G, especially where living related renal transplantation is considered[15].

However, the current paradigm, in which the disease is frequently assumed to result from a rare genetic defect of complement regulation, seems incompatible with the following observations: first, the disease is usually not familial; second, a C3 nephritic factor (C3NeF), an autoantibody that activates the complement alternative pathway in the blood, is detectable in a substantial proportion of patients, including those in whom a rare variant in a complement gene is identified[11]; and third, there is a recognized association of MPGN with other autoimmune diseases[16–18] including a very substantially increased rate of type 1 diabetes mellitus (DM1) in relatives of patients with DDD[19].

Here, we use whole genome sequencing (WGS) to investigate the role of genetic variation in the etiology of PMG in the United Kingdom (UK) population, and resolve all three of these anomalous observations: while rare genetic variation in the *a priori* candidate genes was not enriched in PMG (or the subset with C3G), there is a strong association with common variation at the HLA locus, explaining the phenotypic association with established autoimmune diseases and implicating autoimmunity as the key causal mechanism.

Methods

National Institute for Health Research (NIHR) BioResource Rare Diseases (BR-RD) Study

This study is a part of the NIHR BR-RD Study[20] in which WGS has been undertaken on 13,342 individuals: 12,525 across 16 rare disease domains and 817 apparently healthy subjects (see Supplementary Table 2). Given the potential for a shared genetic etiology with PMG, cohorts with diseases with a known immunological basis (pulmonary artery hypertension and primary immunodeficiencies) and steroid resistant nephrotic syndrome were excluded. Clinical phenotypic data for all participants was encoded using Human Phenotype Ontology[21], SNOMEDCT and ORPHANET codes. Amongst the non-PMG subjects, three participants with the phenotypes microangiopathic haemolytic anemia, thrombocytopenia acute kidney injury, or SNOMEDCT or ORPHANET codes compatible with hemolytic uremic syndrome, were identified and excluded from the control cohort, as were eight subjects with evidence of retinal drusen or macular degeneration.

A summary of the analytic workflow, number of samples analyzed and main findings is provided in Fig. S1.

PMG cohort

Recruitment of patients with PMG was undertaken from 10 British pediatric (64 patients) and 18 adult centers (120 patients, of whom 21 had pediatric onset of disease). Patients with histologically confirmed MPGN either with or without immune-complex deposition (IC-PMG or C3G, respectively) in the absence of a known or suspected underlying systemic cause[22] were considered eligible. No genetic pre-screening was applied. Clinical data were extracted from the UK Rare Renal Disease Registry (RaDaR, http://rarerenal.org/radar-registry). Where available, kidney biopsies were reviewed centrally to confirm the histological diagnosis and to classify as IC-PMG, C3GN or DDD. Serum C3 nephritic factor (C3NeF) and complement C3 and C4 levels were measured using standard clinically validated assays.

Whole-genome sequencing - data generation, variant calling, annotation, relatedness and ancestry

The methods employed for data generation and variant calling have been previously described[20] and are further detailed, along with information on quality control, variant annotation and the identification of a subset of unrelated individuals of European ancestry, in the Supplementary Methods.

Structural and copy number variants

The occurrence of previously described rare SVs and CNVs for PMG[5–7,23,24] was examined by manually inspecting all SVs and CNVs involving the genes of relevance in unrelated PMG individuals of all ethnicity. Subsequent analyses were restricted to the unrelated European cohort of cases and controls. A genome-wide comparison of the frequency of deletions per gene between PMG and controls was undertaken with p-values calculated by permutation testing (n = 100,000).

Comparison with previously described PMG and aHUS variants

The occurrence of common and rare variants in *C3, CD46, CFB, CFH, CFHR1, CFHR3, CFHR5, CFI, DGKE,* and *THBD* previously observed in patients with atypical hemolytic uremic syndrome (aHUS), age-related macular degeneration (AMD), C3G or thrombotic microangiopathy (TMA), as per the Database of Complement Gene Variants (DCGV), a compilation of rare variant data from 3,128 patients with aHUS and 443 with C3G tested in six national reference laboratories (http://www.complement-db.org)[14] and a further study[11], was examined.

Rare variant candidate gene and exome-wide coding variant burden analysis

Rare coding variants (gnomAD-NFE MAF < 0.0001) of moderate or high impact were extracted. Per-gene rare variant burden was enumerated as the proportion of individuals (cases versus controls) with at least one alternate allele in each gene with significance calculated using the exactCMC function in RVTESTS[25], which employs a Fisher's exact test. Analyses were also conducted filtering variants based on their predicted deleteriousness using CADD scores[26].

Common variant GWAS

Common, high quality variants (MAF ≥ 0.05 in gnomAD-NFE and BR-RD) were retained. Standard quality control procedures[27] were undertaken (see Supplementary Methods). The final dataset included 5,897,512 variants with the call rate across the samples exceeding 0.999. GWAS was undertaken with PLINK v1.9 assuming additive allele effects using logistic regression with the first five principal components as covariates.

HLA imputation

HLA genotyping was performed using BWAKIT/BWAMEM v0.7.15 (https://github.com/lh3/bwa/tree/master/bwakit) and using HLA-HD v1.2.0.1[28] (see Supplementary Methods). Alleles with a MAF < 0.05 in controls were excluded. Logistic regression with the first five

principal components as covariates was performed using PLINK v1.9. Haplotype association analysis was performed using PLINK v1.07.

Replication

HLA serotypes from the National Health Service Blood and Transplant (NHSBT) service were utilized as an independent replication cohort. The analyzed cohort was a subset of data from all Caucasian individuals listed for a kidney transplant in the UK within the last 25 years. HLA serotype data were available for HLA-A, HLA-B, HLA-C, HLA-DR and HLA-DQ. Only those serotypes observed at a frequency > 0.05 in controls were analyzed for association (n = 28). Chi-square allelic tests and logistic regression were performed using PLINK v1.9. Serotypes were converted to molecular subtypes using the HLA Dictionary[29].

Results

The initial PMG cohort comprised 184 participants. Following centralized biopsy review, 19 were excluded, most because the biopsy showed mesangial proliferative glomerulonephritis rather than MPGN, leaving 165 (47.9% male). Histological subtypes, complement abnormalities and clinical features are summarized in Table 1. Full clinical details for all individuals with C3G (C3GN and DDD) are provided in Table S1. A C3NeF was more likely to have been observed in subjects with DDD (Table S1), and almost all subjects with a C3NeF or DDD had exhibited a low C3 level on at least one occasion. Consistent with previous reports[30], transiently low serum C4 was documented in some patients with DDD and C3GN but this almost always normalized within weeks of initial presentation[30].

The total number of individuals in each BR-RD cohort is shown in Table S2. Excluding PID, PAH and SRNS, and control individuals with phenotypic codes compatible with aHUS or AMD (n=11) there were 10,250 individuals for use as controls. Of these, 6,491 were unrelated and of European ancestry. All the cases were genetically unrelated with the exception of one sibling pair and the majority (n = 146) were genetically classified as of European ancestry (Fig. S2). Following further quality control measures, the final dataset comprised 146 PMG cases and 6,442 non-PMG controls.

In the genes typically screened in patients with C3G and aHUS (namely *C3, CD46, CFB, CFH, CFHR5, CFI, DGKE* and *THBD*), there was no enrichment of rare (gnomAD-NFE MAF < 0.0001) variants of moderate or high predicted impact in the PMG cohort (Fig. 1). The number of individuals with at least one such variant was 10 (6.8%, 95% CI 3.5-12.6%) in the PMG cohort, compared with 381 (5.9%, 95% CI 5.4-6.5%) in the non-PMG controls (p=0.37, one-tailed Fisher's exact test), consistent across each

of the control cohorts (Fig. S3). Amongst the PMG cohort, there was no difference in the candidate gene rare variant burden between the histological subgroups C3GN, DDD, IC-PMG and PMG unclassified, between those with or without C3NeF or those with low C3 (Fig. S3). The details of the 11 and 318 variants identified in the PMG subjects and in non-PMG subjects, respectively, are provided in Tables S3 and S4. Analyses were also performed imposing a variable CADD threshold (none to $\geq$ 20) and control allele frequency (gnomAD-NFE MAF < 0.0001 to < 0.01); however in none of these permutations was there a significant difference between PMG and controls (Fig. S4). Furthermore, there was no enrichment of rare variants previously classified as pathogenic or likely pathogenic in the DCGV, with one and thirteen such variants identified in PMG and non-PMG individuals, respectively (Table S5).

Sixteen previously described common complement gene variants (gnomAD-NFE MAF ≥ 0.05) were identified (Table S6). Computing pairwise LD demonstrated that these variants represented 10 independent signals at $r^2 < 0.8$. Association analysis using logistic regression including principal components as covariates identified four variants in the genes *C3* and *CFH*, representing two independent signals, that were statistically significant after correction for multiple testing (Bonferroni threshold for 10 loci, $p<0.005$) (Table 2). The full association statistics for all 16 variants are provided in Table S6. There was no evidence of epistasis between the associated variants ($p > 0.05$).

Across the whole exome, there was no enrichment of rare variants with a moderate or high predicted impact per gene in PMG (Fig. S5). The minimum p-value across the exome was 1.9×10^{-4} as compared with the exome-wide significance threshold, correcting for 28,252 genes, of $p < 1.77\times10^{-6}$. The QQ plot showed no evidence of deviation from the null (Fig. S6). When filtering the data using a CADD threshold of ≥15, the minimum p-value was also 1.9×10^{-4}.

The only previously reported rare SV observed in PMG cases was the 6.3 kbp *CFHR5* tandem duplication (chr1:196950207-196956508) known to cause CFHR5 nephropathy[5], present in a single individual of Cypriot ancestry. The common *CFHR3-CFHR1* deletion was observed at a similar frequency in European PMG cases and controls at 0.164 and 0.201, respectively (Fisher's exact test p = 0.139), similar to that in the UK population[31]. Across all the candidate genes, a total of 65 SVs and CNVs were identified of which only one was seen in a PMG case: a 128.3 kbp heterozygous deletion involving exon 1 of *CFH* (chr1:196498350-196626665). Genome-wide, there was no enrichment of deletions in PMG cases either in total or per-gene after correcting for multiple testing by permutation analysis.

Common variant GWAS examining unrelated individuals of European ancestry, identified one locus achieving genome-wide significance ($p < 5\times10^{-8}$) at 6p21.32 (Fig. 2). The genomic inflation (lambda)

was 1.017 (QQ plot in Fig. S7). Association statistics for all variants achieving $p < 5 \times 10^{-8}$ are provided in Table S7. At the 6p21.32 locus, the lead variant (rs35406322) was associated at $p=3.29 \times 10^{-8}$ with OR 1.93 (95% CI: 1.53-2.44). Significance was maximal for variants within the gene *C6orf10* (Fig. 3). The control allele frequency of the lead variant (0.361) approximated to that in gnomAD-NFE (0.376) and was consistent across all of the BR-RD control cohorts (Fig. S8). There was no statistically significant difference in the frequency of the lead variants by PMG histological subtype, C3NeF status or in those with low C3 (Fig. S8). Conditioning on the lead variant abrogated the signal. The second lead variant is a known eQTL for multiple genes in multiple tissue types at genome-wide significance including *HLA-DRB5*, *CYP21A1P*, *C4A* and *NOTCH4* (Table S8). There was no evidence of epistasis (p > 0.05) between the 6p21.32 variants and the nominally associated common candidate complement gene variants. Testing dominant and recessive models for the lead variant showed weaker evidence of association, suggesting an additive genetic model best explains the association. A second locus at 12q14.1 was not statistically significant at the genome-wide level (lead variant rs61938185, $p=6.14 \times 10^{-8}$).

Fully imputed HLA genotypes at all six loci were available for all 146 of the PMG cases and 6,386 of the controls. A total of 39 HLA alleles were observed with a frequency > 0.05. The strongest association was with DQA1*05:01 at $p = 2.09 \times 10^{-8}$ with OR 1.94 (95% CI 1.54-2.45) followed by DRB1*03:01 and DQB1*02:01 (Table 3). Full association details for all alleles tested are provided in Table S9. The association with both DQB1*02:01 and DRB1*03:01 was abrogated by conditioning on DQA1*05:01. The DQA1*05:01|DQB1*02:01|DRB1*03:01 haplotype was observed in cases at a frequency of 0.233 compared with 0.122 in controls ($p = 1.21 \times 10^{-8}$, OR 2.19, 95% CI: 1.66-2.89). The control frequency of this haplotype approximates to that observed in 1,899 European American individuals (0.131)[32]. These analyses were repeated using HLA types imputed using HLA-HD, yielding similar results (Table S9).

HLA serotypes from the NHSBT were available from 338 individuals with MPGN (both primary and secondary) and 15,614 non-MPGN controls with renal failure of non-immune or unknown etiology, the largest groups of which were: unknown (n = 6,836), polycystic kidney disease (n = 4,442) and pyelonephritis/interstitial nephritis (n = 1,958). Employing a Bonferroni threshold of $p < 1.8 \times 10^{-3}$ (n = 28), three serotypes were statistically significantly associated with MPGN with OR of approximately 1.4: DR17 (corresponding to DRB1*03:01/04), B8 (B*08) and DQ2 (DQB1*02:01/02/03/04/05) (Table 4). The frequency of the most significant serotype, DR17, was approximately consistent across each of the control cohorts, particularly those with larger sample sizes (Fig. S9) and equal to the frequency in 1,043 UK blood donors[33]. The significance of B8 and DQ2 was abrogated by conditioning on DR17. The DR17|DQ2 haplotype was observed in cases at a frequency of 0.175 compared with 0.129 in controls ($p = 4.55 \times 10^{-4}$, OR 1.43, 95% CI: 1.17-1.75).

Discussion

In this study we have examined the genetics of PMG using whole-genome sequence data generated from a UK-wide collection of cases and a large number of genetic ancestry-matched non-PMG controls. The high prevalence of C3NeFs and of reduced serum C3 levels, especially in patients with DDD and/or a C3NeF, is consistent with previous literature[11,13] and suggests that the cohort under study was comparable with previously reported PMG cohorts. While we did observe rare, protein altering variants in the candidate genes (encoding components of the complement alternative pathway and other genes observed in the related disease aHUS), PMG cases were not enriched for such variants, which occurred at a frequency of approximately 7% across all cohorts. Our study of 146 European cases and 6,442 European controls had >92% power to detect a >15% burden of rare complement gene variants in PMG. We also observed association of PMG with common alleles of the candidate complement genes and, although not statistically significant at the genome-wide level, this is consistent with previous data and provides evidence that variation in genes encoding components of the complement alternative pathway affects susceptibility to PMG.

Power calculation, which we performed before recruitment to this study[34], indicated that, using whole exome analysis, 100 subjects would have provided >80% power to detect association with rare variants in a novel gene accounting for 20% of unexplained cases under a dominant model (power would be >95% under a recessive model). We recruited a greater number than this, which suggests that any currently unrecognized monogenic disorders are unlikely to account for a significant proportion of PMG in the UK population.

Analysis of the frequency of common genetic variants across the genome identified a single locus achieving genome-wide significance of $p < 5 \times 10^{-8}$. Numerous markers at the HLA locus were strongly associated with PMG, and imputation identified an associated haplotype containing DQA1*05:01, DQB1*02:01 and DRB1*03:01. This finding was replicated in an independent cohort that included both primary and secondary MPGN, in which we observed association with the corresponding HLA serotypes DQ2 (DQB1*02:01) and DR17 (DRB1*03:01) which are associated with a number of immune-mediated disorders. These genes encode components of the MHC class II molecule that are found on the surface of antigen presenting cells and are important in initiation of the adaptive immune response, including antibody production[35]. This suggests that a key step in the pathogenesis of these disorders is an aberrant adaptive immune response[36], which is consistent with the high frequency of autoantibodies (i.e. C3NeFs) in these patients. However, the possibility has not been excluded that the observed associations are mediated by one of the non-HLA genes spanned by this haplotype, including

those encoding complement components C2, Factor B and C4. Variation in dose of *C4A* and *C4B*, which encode isotypes of C4 that preferentially bind antibody-protein or antibody-cell surface complexes respectively[37], is known to affect serum C4 activity and has previously been implicated in systemic lupus erythematosus (SLE) and schizophrenia[38,39]. However, direct comparison of the copy number of each of these genes showed no significant differences between PMG cases and controls. Examining C4NeF levels and the relationship with the observed variants may be informative.

We observed shared genetic risk factors in all the subgroups of PMG (IC-PMG, C3GN and DDD), as well as those with and without C3NeF or low C3 (Fig. S5) in this study, implying shared underlying disease mechanisms. We did not observe any significant genetic differences between these subgroups and it is likely larger studies would be needed to identify such differences, if they exist. Tests for the presence of C4 nephritic factors or autoantibodies against complement regulators (which have been reported in patients with PMG[40]) were not available for the whole cohort, so we are unable to determine the proportion of patients in whom an autoantibody was present, however the HLA association we observed is consistent with other reports in which multiple autoantibodies are detectable in cohorts of patients with PMG[40] and previous observations showing an association between MPGN/C3G and autoimmune disorders [16–18,19]. The lead variant (e.g. rs3117137) and HLA alleles that we observed to be associated with PMG have previously been associated with a number of immune mediated diseases including membranous nephropathy[41], rheumatoid arthritis[42], myasthenia gravis[43], asthma[44], coeliac disease[45] and DM1[46,47], potentially explaining the observed phenotypic association between these different disorders. Together, these findings imply that, rather than resulting from a primary genetic disorder of complement alternative pathway regulation, in most cases PMG is actually an autoimmune disease.

Author contributions

DPG conceived and designed the study. APL, MMYC and OS analyzed the data with the assistance of KC, CP and ST. SJ directed the establishment of the PMG cohort. HTC undertook the centralized biopsy review. DPG and APL drafted the manuscript. All other authors were responsible for recruitment of subjects to the study and data acquisition.

Acknowledgments

Funding

Supplementary Appendix

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- Supplementary Tables
    - Table S1: Histological characterization and clinical details of C3G subjects
    - Table S2: BR-RD cohorts and sample sizes with filtering
    - Table S3: Prioritized rare variants in candidate genes in PMG subjects
    - Table S4: Prioritized rare variants in candidate genes in non-PMG subjects
    - Table S5: Rare variants in candidate genes in the Database of Complement Gene Variants
    - Table S6: Association statistics for common variants in candidate genes previously identified in aHUS/MPGN
    - Table S7: Association statistics for all variants achieving $p<5\times10^{-8}$ in the genome-wide association study
    - Table S8: eQTL results for rs3117135
    - Table S9: Association statistics for imputed HLA alleles using BWAKIT/BWAMEM and HLA-HD

Page **13** of **23**

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

References

1.   Sethi S, Nester CM, Smith RJH: Membranoproliferative glomerulonephritis and C3 glomerulopathy: resolving the confusion. *Kidney Int.* 81: 434–441, 2012

2.   Briganti EM, Dowling J, Finlay M, Hill PA, Jones CL, Kincaid-Smith PS, Sinclair R, McNeil JJ, Atkins RC: The incidence of biopsy-proven glomerulonephritis in Australia. *Nephrol. Dial. Transplant.* 16: 1364–1367, 2001

3.   Swaminathan S, Leung N, Lager DJ, Melton LJ, Bergstralh EJ, Rohlinger A, Fervenza FC: Changing incidence of glomerular disease in Olmsted County, Minnesota: a 30-year renal biopsy study. *Clin J Am Soc Nephrol* 1: 483–487, 2006

4.   Sethi S, Zand L, Leung N, Smith RJH, Jevremonic D, Herrmann SS, Fervenza FC: Membranoproliferative glomerulonephritis secondary to monoclonal gammopathy. *Clin J Am Soc Nephrol* 5: 770–782, 2010

5.   Gale DP, de Jorge EG, Cook HT, Martinez-Barricarte R, Hadjisavvas A, McLean AG, Pusey CD, Pierides A, Kyriacou K, Athanasiou Y, Voskarides K, Deltas C, Palmer A, Frémeaux-Bacchi V, de Cordoba SR, Maxwell PH, Pickering MC: Identification of a mutation in complement factor H-related protein 5 in patients of Cypriot origin with glomerulonephritis. *Lancet* 376: 794–801, 2010

6.   Malik TH, Lavin PJ, Goicoechea de Jorge E, Vernon KA, Rose KL, Patel MP, de Leeuw M, Neary JJ, Conlon PJ, Winn MP, Pickering MC: A hybrid CFHR3-1 gene causes familial C3 glomerulopathy. *J. Am. Soc. Nephrol.* 23: 1155–1160, 2012

7.   Chen Q, Wiesener M, Eberhardt HU, Hartmann A, Uzonyi B, Kirschfink M, Amann K, Buettner M, Goodship T, Hugo C, Skerka C, Zipfel PF: Complement factor H-related hybrid protein deregulates complement in dense deposit disease. *J. Clin. Invest.* 124: 145–155, 2014

8.   Tortajada A, Yébenes H, Abarrategui-Garrido C, Anter J, García-Fernández JM, Martínez-Barricarte R, Alba-Domínguez M, Malik TH, Bedoya R, Cabrera Pérez R, López Trascasa M, Pickering MC, Harris CL, Sánchez-Corral P, Llorca O, Rodríguez de Córdoba S: C3 glomerulopathy-associated CFHR1 mutation alters FHR oligomerization and complement regulation. *J. Clin. Invest.* 123: 2434–2446, 2013

9.   Ault BH, Schmidt BZ, Fowler NL, Kashtan CE, Ahmed AE, Vogt BA, Colten HR: Human factor H deficiency. Mutations in framework cysteine residues and block in H protein secretion and intracellular catabolism. *J. Biol. Chem.* 272: 25168–25175, 1997

10.   Martínez-Barricarte R, Heurich M, Valdes-Cañedo F, Vazquez-Martul E, Torreira E, Montes T, Tortajada A, Pinto S, Lopez-Trascasa M, Morgan BP, Llorca O, Harris CL, Rodríguez de Córdoba S: Human C3 mutation reveals a mechanism of dense deposit disease pathogenesis and provides insights into complement activation and regulation. *J. Clin. Invest.* 120: 3702–3712, 2010

11.   Servais A, Noël L-H, Roumenina LT, Le Quintrec M, Ngo S, Dragon-Durey M-A, Macher M-A, Zuber J, Karras A, Provot F, Moulin B, Grünfeld J-P, Niaudet P, Lesavre P, Frémeaux-Bacchi V: Acquired and genetic complement abnormalities play a critical role in dense deposit disease and other C3 glomerulopathies. *Kidney Int.* 82: 454–464, 2012

12.   Bu F, Borsa NG, Jones MB, Takanami E, Nishimura C, Hauer JJ, Azaiez H, Black-Ziegelbein EA, Meyer NC, Kolbe DL, Li Y, Frees K, Schnieders MJ, Thomas C, Nester C, Smith RJH: High-Throughput Genetic Testing for Thrombotic Microangiopathies and C3 Glomerulopathies. *J. Am. Soc. Nephrol.* 27: 1245–1253, 2016

13.   Iatropoulos P, Noris M, Mele C, Piras R, Valoti E, Bresin E, Curreri M, Mondo E, Zito A, Gamba S, Bettoni S, Murer L, Fremeaux-Bacchi V, Vivarelli M, Emma F, Daina E, Remuzzi G: Complement gene variants determine the risk of immunoglobulin-associated MPGN and C3 glomerulopathy and predict long-term renal outcome. *Mol. Immunol.* 71: 131–142, 2016

14. Osborne AJ, Breno M, Borsa NG, Bu F, Frémeaux-Bacchi V, Gale DP, van den Heuvel LP, Kavanagh D, Noris M, Pinto S, Rallapalli PM, Remuzzi G, Rodríguez de Cordoba S, Ruiz A, Smith RJH, Vieira-Martins P, Volokhina E, Wilson V, Goodship THJ, Perkins SJ: Statistical Validation of Rare Complement Variants Provides Insights into the Molecular Basis of Atypical Hemolytic Uremic Syndrome and C3 Glomerulopathy. *J. Immunol.* 200: 2464–2478, 2018

15. Goodship THJ, Cook HT, Fakhouri F, Fervenza FC, Frémeaux-Bacchi V, Kavanagh D, Nester CM, Noris M, Pickering MC, Rodríguez de Córdoba S, Roumenina LT, Sethi S, Smith RJH, Conference Participants: Atypical hemolytic uremic syndrome and C3 glomerulopathy: conclusions from a "Kidney Disease: Improving Global Outcomes" (KDIGO) Controversies Conference. *Kidney Int.* 91: 539–551, 2017

16. Dizdar O, Kahraman S, Gençtoy G, Ertoy D, Arici M, Altun B, Yasavul U, Turgan C: Membranoproliferative glomerulonephritis associated with type 1 diabetes mellitus and Hashimoto's thyroiditis. *Nephrol. Dial. Transplant.* 19: 988–989, 2004

17. Chihara J, Takebayashi S, Taguchi T, Yokoyama K, Harada T, Naito S: Glomerulonephritis in diabetic patients and its effect on the prognosis. *Nephron* 43: 45–49, 1986

18. Srikanta S, Malaviya AN, Rajagopalan P, Bhuyan UN, Ahuja MM: Association of type I (insulin-dependent) diabetes mellitus, autoimmunity, antinuclear antibody, and membranoproliferative glomerulonephritis. *Diabetes Care* 6: 71–74, 1983

19. Lu D-F, Moon M, Lanning LD, McCarthy AM, Smith RJH: Clinical features and outcomes of 98 children and adults with dense deposit disease. *Pediatr. Nephrol.* 27: 773–781, 2012

20. Ouwehand WH: Whole-genome sequencing of rare disease patients in a national healthcare system. *bioRxiv* [Internet] 2019 Available from: http://biorxiv.org/lookup/doi/10.1101/507244 [cited 2019 Feb 11]

21. Köhler S, Carmody L, Vasilevsky N, Jacobsen JOB, Danis D, Gourdine J-P, Gargano M, Harris NL, Matentzoglu N, McMurry JA, Osumi-Sutherland D, Cipriani V, Balhoff JP, Conlin T, Blau H, Baynam G, Palmer R, Gratian D, Dawkins H, Segal M, Jansen AC, Muaz A, Chang WH, Bergerson J, Laulederkind SJF, Yüksel Z, Beltran S, Freeman AF, Sergouniotis PI, Durkin D, Storm AL, Hanauer M, Brudno M, Bello SM, Sincan M, Rageth K, Wheeler MT, Oegema R, Lourghi H, Della Rocca MG, Thompson R, Castellanos F, Priest J, Cunningham-Rundles C, Hegde A, Lovering RC, Hajek C, Olry A, Notarangelo L, Similuk M, Zhang XA, Gómez-Andrés D, Lochmüller H, Dollfus H, Rosenzweig S, Marwaha S, Rath A, Sullivan K, Smith C, Milner JD, Leroux D, Boerkoel CF, Klion A, Carter MC, Groza T, Smedley D, Haendel MA, Mungall C, Robinson PN: Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Research* 47: D1018–D1027, 2019

22. Pickering MC, D'Agati VD, Nester CM, Smith RJ, Haas M, Appel GB, Alpers CE, Bajema IM, Bedrosian C, Braun M, Doyle M, Fakhouri F, Fervenza FC, Fogo AB, Frémeaux-Bacchi V, Gale DP, Goicoechea de Jorge E, Griffin G, Harris CL, Holers VM, Johnson S, Lavin PJ, Medjeral-Thomas N, Paul Morgan B, Nast CC, Noel L-H, Peters DK, Rodríguez de Córdoba S, Servais A, Sethi S, Song W-C, Tamburini P, Thurman JM, Zavros M, Cook HT: C3 glomerulopathy: consensus report. *Kidney Int.* 84: 1079–1089, 2013

23. Togarsimalemath SK, Sethi SK, Duggal R, Le Quintrec M, Jha P, Daniel R, Gonnet F, Bansal S, Roumenina LT, Fremeaux-Bacchi V, Kher V, Dragon-Durey M-A: A novel CFHR1-CFHR5 hybrid leads to a familial dominant C3 glomerulopathy. *Kidney Int.* 92: 876–887, 2017

24. Xiao X, Ghossein C, Tortajada A, Zhang Y, Meyer N, Jones M, Borsa NG, Nester CM, Thomas CP, de Córdoba SR, Smith RJH: Familial C3 glomerulonephritis caused by a novel CFHR5-CFHR2 fusion gene. *Mol. Immunol.* 77: 89–96, 2016

25. Zhan X, Hu Y, Li B, Abecasis GR, Liu DJ: RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics* 32: 1423–1426, 2016

26. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M: CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* 47: D886–D894, 2019

27. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT: Data quality control in genetic case-control association studies. *Nat Protoc* 5: 1564–1573, 2010

28. Kawaguchi S, Higasa K, Shimizu M, Yamada R, Matsuda F: HLA-HD: An accurate HLA typing algorithm for next-generation sequencing data. *Hum. Mutat.* 38: 788–797, 2017

29. Holdsworth R, Hurley CK, Marsh SGE, Lau M, Noreen HJ, Kempenich JH, Setterholm M, Maiers M: The HLA dictionary 2008: a summary of HLA-A, -B, -C, -DRB1/3/4/5, and -DQB1 alleles and their association with serologically defined HLA-A, -B, -C, -DR, and -DQ antigens. *Tissue Antigens* 73: 95–170, 2009

30. Vargas R, Thomson KJ, Wilson D, Cameron JS, Turner DR, Gill D, Chantler C, Ogg CS: Mesangiocapillary glomerulonephritis with dense "deposits" in the basement membranes of the kidney. *Clin. Nephrol.* 5: 73–82, 1976

31. Holmes LV, Strain L, Staniforth SJ, Moore I, Marchbank K, Kavanagh D, Goodship JA, Cordell HJ, Goodship THJ: Determining the population frequency of the CFHR3/CFHR1 deletion at 1q32. *PLoS ONE* 8: e60352, 2013

32. Klitz W, Maiers M, Spellman S, Baxter-Lowe LA, Schmeckpeper B, Williams TM, Fernandez-Viña M: New HLA haplotype frequency reference standards: high-resolution and large sample typing of HLA DR-DQ haplotypes in a sample of European Americans. *Tissue Antigens* 62: 296–307, 2003

33. Pingel J, Solloch UV, Hofmann JA, Lange V, Ehninger G, Schmidt AH: High-resolution HLA haplotype frequencies of stem cell donors in Germany with foreign parentage: how can they be used to improve unrelated donor searches? *Hum. Immunol.* 74: 330–340, 2013

34. Zhi D, Chen R: Statistical guidance for experimental design and data analysis of mutation detection in rare monogenic mendelian diseases by exome sequencing. *PLoS ONE* 7: e31358, 2012

35. Klein J, Sato A: The HLA system. First of two parts. *N. Engl. J. Med.* 343: 702–709, 2000

36. Dendrou CA, Petersen J, Rossjohn J, Fugger L: HLA variation and disease. *Nat. Rev. Immunol.* 18: 325–339, 2018

37. Law SK, Dodds AW, Porter RR: A comparison of the properties of two classes, C4A and C4B, of the human complement component C4. *EMBO J.* 3: 1819–1823, 1984

38. Jüptner M, Flachsbart F, Caliebe A, Lieb W, Schreiber S, Zeuner R, Franke A, Schröder JO: Low copy numbers of complement C4 and homozygous deficiency of C4A may predispose to severe disease and earlier disease onset in patients with systemic lupus erythematosus. *Lupus* 27: 600–609, 2018

39. Sekar A, Bialas AR, de Rivera H, Davis A, Hammond TR, Kamitaki N, Tooley K, Presumey J, Baum M, Van Doren V, Genovese G, Rose SA, Handsaker RE, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Daly MJ, Carroll MC, Stevens B, McCarroll SA: Schizophrenia risk from complex variation of complement component 4. *Nature* 530: 177–183, 2016

40. Noris M, Donadelli R, Remuzzi G: Autoimmune abnormalities of the alternative complement pathway in membranoproliferative glomerulonephritis and C3 glomerulopathy. *Pediatr. Nephrol.* 2018

41. Stanescu HC, Arcos-Burgos M, Medlar A, Bockenhauer D, Kottgen A, Dragomirescu L, Voinescu C, Patel N, Pearce K, Hubank M, Stephens HAF, Laundy V, Padmanabhan S, Zawadzka A, Hofstra JM, Coenen MJH, den Heijer M, Kiemeney LALM, Bacq-Daian D, Stengel B, Powis SH, Brenchley P, Feehally J, Rees AJ, Debiec H, Wetzels JFM, Ronco P, Mathieson PW, Kleta R: Risk HLA-DQA1 and PLA(2)R1 alleles in idiopathic membranous nephropathy. *N. Engl. J. Med.* 364: 616–626, 2011

42. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, Kochi Y, Ohmura K, Suzuki A, Yoshida S, Graham RR, Manoharan A, Ortmann W, Bhangale T, Denny JC, Carroll RJ, Eyler AE, Greenberg JD, Kremer JM, Pappas DA, Jiang L, Yin J, Ye L, Su D-F, Yang J, Xie G, Keystone E, Westra H-J, Esko T, Metspalu A, Zhou X, Gupta N, Mirel D, Stahl EA, Diogo D, Cui J, Liao K, Guo MH, Myouzen K, Kawaguchi T,

Coenen MJH, van Riel PLCM, van de Laar MAFJ, Guchelaar H-J, Huizinga TWJ, Dieudé P, Mariette X, Bridges SL, Zhernakova A, Toes REM, Tak PP, Miceli-Richard C, Bang S-Y, Lee H-S, Martin J, Gonzalez-Gay MA, Rodriguez-Rodriguez L, Rantapää-Dahlqvist S, Arlestig L, Choi HK, Kamatani Y, Galan P, Lathrop M, RACI consortium, GARNET consortium, Eyre S, Bowes J, Barton A, de Vries N, Moreland LW, Criswell LA, Karlson EW, Taniguchi A, Yamada R, Kubo M, Liu JS, Bae S-C, Worthington J, Padyukov L, Klareskog L, Gregersen PK, Raychaudhuri S, Stranger BE, De Jager PL, Franke L, Visscher PM, Brown MA, Yamanaka H, Mimori T, Takahashi A, Xu H, Behrens TW, Siminovitch KA, Momohara S, Matsuda F, Yamamoto K, Plenge RM: Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 506: 376–381, 2014

43. Gregersen PK, Kosoy R, Lee AT, Lamb J, Sussman J, McKee D, Simpfendorfer KR, Pirskanen-Matell R, Piehl F, Pan-Hammarstrom Q, Verschuuren JJGM, Titulaer MJ, Niks EH, Marx A, Ströbel P, Tackenberg B, Pütz M, Maniaol A, Elsais A, Tallaksen C, Harbo HF, Lie BA, Raychaudhuri S, de Bakker PIW, Melms A, Garchon H-J, Willcox N, Hammarstrom L, Seldin MF: Risk for myasthenia gravis maps to a (151) Pro→Ala change in TNIP1 and to human leukocyte antigen-B*08. *Ann. Neurol.* 72: 927–935, 2012

44. Hirota T, Takahashi A, Kubo M, Tsunoda T, Tomita K, Doi S, Fujita K, Miyatake A, Enomoto T, Miyagawa T, Adachi M, Tanaka H, Niimi A, Matsumoto H, Ito I, Masuko H, Sakamoto T, Hizawa N, Taniguchi M, Lima JJ, Irvin CG, Peters SP, Himes BE, Litonjua AA, Tantisira KG, Weiss ST, Kamatani N, Nakamura Y, Tamari M: Genome-wide association study identifies three new susceptibility loci for adult asthma in the Japanese population. *Nat. Genet.* 43: 893–896, 2011

45. van Heel DA, Franke L, Hunt KA, Gwilliam R, Zhernakova A, Inouye M, Wapenaar MC, Barnardo MCNM, Bethel G, Holmes GKT, Feighery C, Jewell D, Kelleher D, Kumar P, Travis S, Walters JRF, Sanders DS, Howdle P, Swift J, Playford RJ, McLaren WM, Mearin ML, Mulder CJ, McManus R, McGinnis R, Cardon LR, Deloukas P, Wijmenga C: A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat. Genet.* 39: 827–829, 2007

46. Hakonarson H, Grant SFA, Bradfield JP, Marchand L, Kim CE, Glessner JT, Grabs R, Casalunovo T, Taback SP, Frackelton EC, Lawson ML, Robinson LJ, Skraban R, Lu Y, Chiavacci RM, Stanley CA, Kirsch SE, Rappaport EF, Orange JS, Monos DS, Devoto M, Qu H-Q, Polychronakos C: A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. *Nature* 448: 591–594, 2007

47. Erlich H, Valdes AM, Noble J, Carlson JA, Varney M, Concannon P, Mychaleckyj JC, Todd JA, Bonella P, Fear AL, Lavant E, Louey A, Moonsamy P, Type 1 Diabetes Genetics Consortium: HLA DR-DQ haplotypes and genotypes and type 1 diabetes risk: analysis of the type 1 diabetes genetics consortium families. *Diabetes* 57: 1084–1092, 2008

Table 1 – Histological categorization and clinical details. Ultrastructural and immunostaining data to allow sub-classification into IC-PMG, C3GN or DDD were available for 114 subjects. Results of clinically accredited C3NeF assays were available for 67 and serum C3 and C4 levels were available for 71 subjects, but only a small number of those with end stage renal disease.

Category	Total (%)	C3NeF detected (% of 69)	Low C3 (<0.68 g/L) (%)	Low C4 (<0.18 g/L) (%)
All subjects	165 (100)	25 (36.2)	58 (69.0)	38 (45.2)
IC-PMG	53 (46.5)	6 (24.0)	19 (63.3)	19 (63.3)
C3GN	39 (34.2)	8 (33.3)	19 (67.9)	8 (28.6)
DDD	22 (19.3)	10 (58.8)	16 (84.2)	9 (47.4)
Pediatric onset	85 (51.5)	23 (40.4)	51 (77.2)	33 (50.0)
Immunosuppression	75 (56.0)	18 (38.3)	42 (77.8)	27 (50.0)
End stage renal disease	42 (25.4)	5 (41.7)	6 (42.9)	1 (7.1)
Renal transplant	30 (18.2)	4 (57.1)	4 (44.4)	0 (0.0)
C3NeF detected	25 (36.2)	-	23 (92.0)	11 (44.0)

Table 2 - Association statistics comparing PMG with controls for four common variants in complement genes previously described at altered frequency in individuals with aHUS/MPGN. gnomAD-NFE: allele frequency in non-Finish Europeans in the gnomAD database. Positions shown are for Build 37 of the human genome. The two chromosome 19 variants and two chromosome 1 variants are in linkage disequilibrium ($r^2 = 0.844$ and $r^2 = 0.999$, respectively).

Chr	Pos	Ref	Alt	rsID	Gene	HUGO Effect	Case	Control	gnomAD-NFE	OR (95% CI)	P
1	196654324	A	C	rs1061147	*CFH*	p.A307A	0.534	0.620	0.617	0.71 (0.56-0.89)	3.5×10^{-3}
1	196659237	C	T	rs1061170	*CFH*	p.H402Y	0.534	0.620	0.616	0.71 (0.56-0.89)	3.4×10^{-3}
19	6713262	G	A	rs1047286	*C3*	p.P314L	0.284	0.213	0.200	1.47 (1.13-1.90)	3.9×10^{-3}
19	6718387	G	C	rs2230199	*C3*	p.R102G	0.295	0.218	0.206	1.49 (1.16-1.93	2.1×10^{-3}

Table 3 - Association statistics comparing PMG and controls for the three most significant HLA alleles imputed using BWAKIT/BWAMEM.

Allele	Case	Control	OR (95% CI)	P
DQA1*05:01	0.291	0.154	1.94 (1.54-2.45)	2.09×10^{-8}
DRB1*03:01	0.264	0.146	1.94 (1.51-2.50)	2.46×10^{-7}
DQB1*02:01	0.264	0.144	1.81 (1.43-2.29)	7.69×10^{-7}

Table 4 - Association statistics for the HLA serotypes associated with MPGN in the NHSBT data after correcting for multiple testing ($p < 1.8 \times 10^{-3}$).

Serotype	Case	Control	OR (95% CI)	P
DR17	0.188	0.137	1.46 (1.20-1.78)	1.4×10^{-4}
B8	0.186	0.140	1.42 (1.16-1.72)	4.7×10^{-4}
DQ2	0.283	0.228	1.33 (1.12-1.58)	9.3×10^{-4}

Figures

Fig. 1
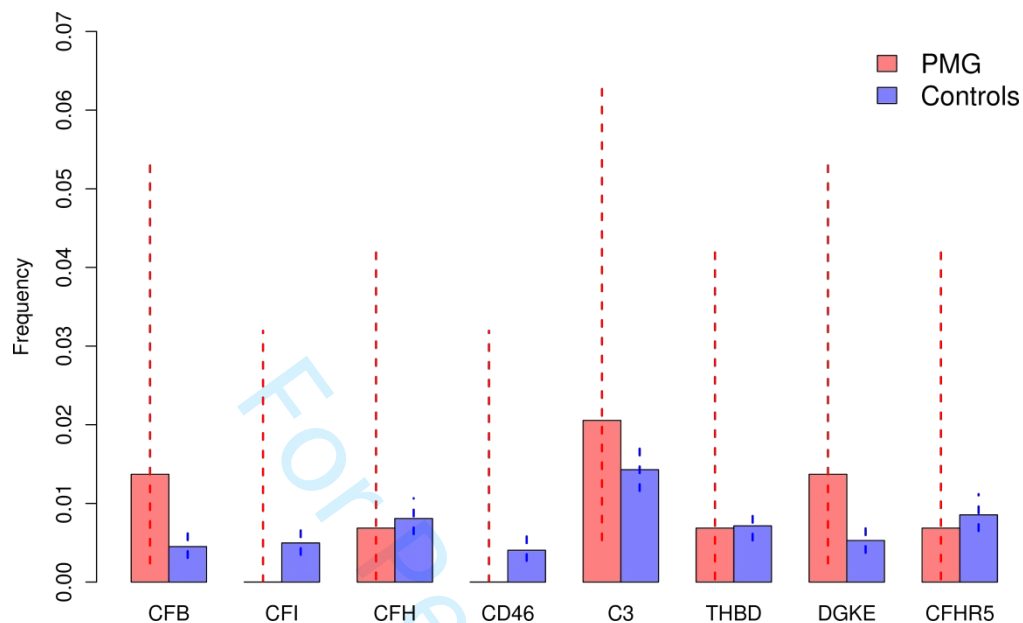
Burden of rare variants with moderate or high predicted impact in candidate genes comparing unrelated European PMG (n = 146) with controls (n = 6,442). Vertical dotted lines indicate the 95% confidence intervals.
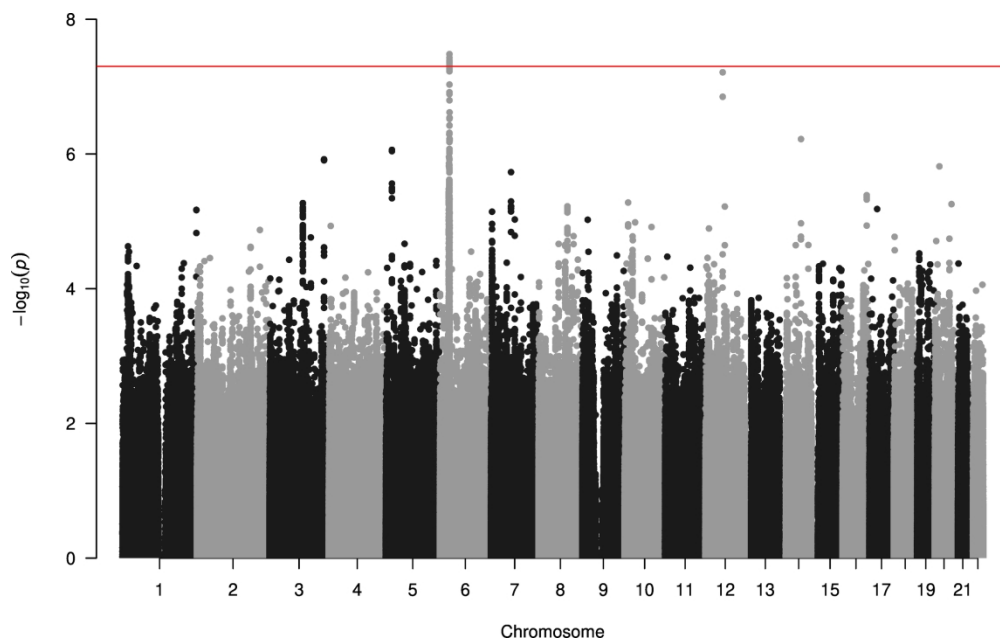
Page **21** of **23**

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Fig. 2
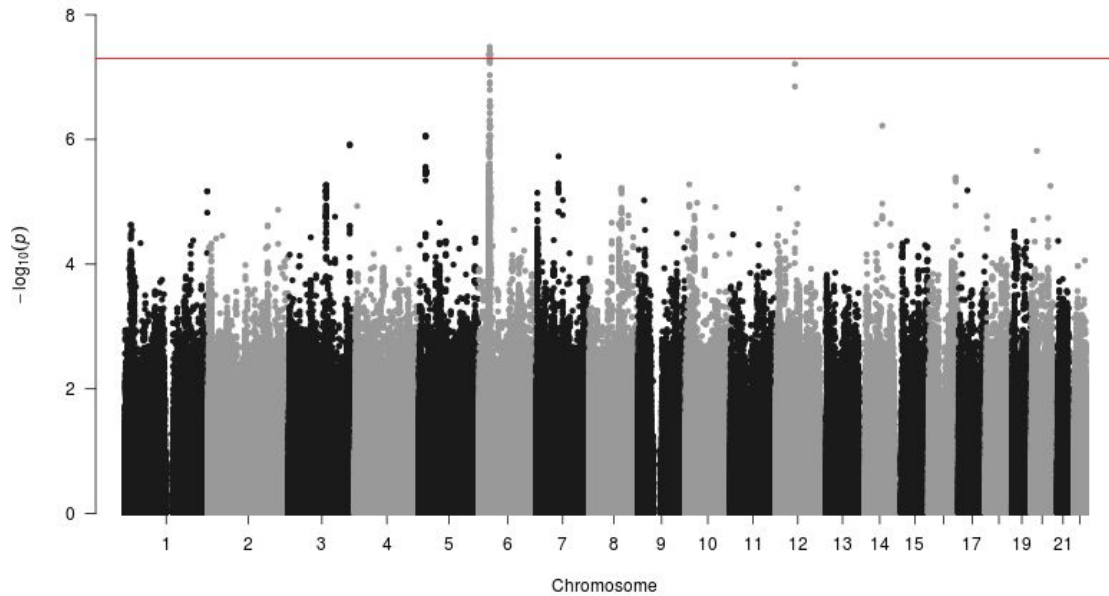
GWAS Manhattan plot comparing European unrelated PMG cases (n = 146) and controls (n = 6,442). One locus surpasses the genome-wide significance threshold (p < 5×10^{-8}, indicated by the red line) on chromosomes six.

Fig. 3
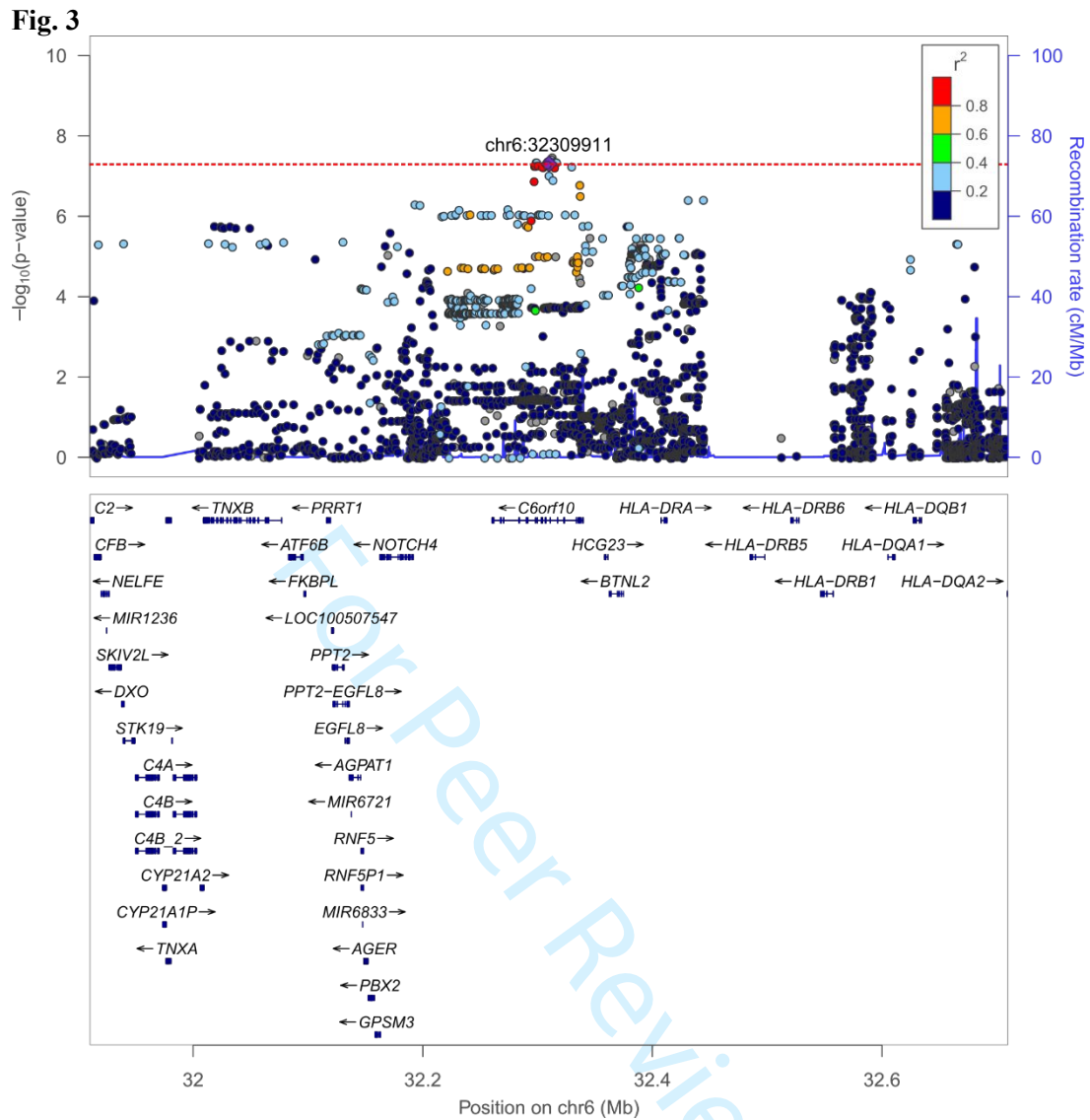
LocusZoom plot for the chromosome 6 locus associated with PMG at genome-wide significance with variants colored based on their linkage disequilibrium (LD) using 1000 Genomes (November 2014) European data. As there were no reference LD data available for the lead marker, the second most significant marker was used (chr6:32313531, rs3117135). A horizontal red dotted line indicates $p=5\times10^{-8}$.

243x151mm (300 x 300 DPI)

Large scale whole-genome sequencing reveals the genetic architecture of primary membranoproliferative glomerulonephritis and C3 glomerulopathy

Supplementary Appendix

Adam P. Levine, Melanie M. Y. Chan, Omid Sadeghi-Alavijeh, Edwin K. S. Wong, H. Terence Cook, Sofie Ashford, Keren Carss, Martin T. Christian, Matthew Hall, Claire Harris, Paul McAlinden, Kevin J. Marchbank, Stephen D. Marks, Heather Maxwell, Karyn Megy, Christopher J. Penkett, Monika Mozere, Kathleen E. Stirrups, Salih Tuna, Julie Wessels, Deborah Whitehorn, MPGN/DDD/C3 Glomerulopathy Rare Disease Group, NIHR BioResource, Sally A. Johnson, Daniel P. Gale

Correspondence: d.gale@ucl.ac.uk

1

Contents

1 Consortia

The MPGN/C3G Rare Disease Group

Su Brimble, Terence Cook, Daniel Gale, Julie Gibbs, Rodney Gilbert, Lorraine Harper, Claire Harris, Kim Jessup, Sally Johnson, Helen Jones, David Kavanagh, Adam Levine, Hannah Lomax-Browne, Andrew Longfellow, Roger Malcomson, Kevin Marchbank, Stephen Marks, Heather Maxwell, Paul McAlinden, David Milford, Matthew Pickering, Sandra Richardson, Stephen Richardson, Neil Sebire, Mark Taylor, Julie Wessels, Sarah Whittall, Edwin Wong.

Physicians also recruiting to the study

Saeed Ahmed, Vicki Ayub, Jonathan Barratt, Preetham Boddana, Gary Campbell, Maggie Fitzpatrick, Sian Griffiths, Shivaram Hegde, Praveen Jeevaratnam, Caroline Jones Nicholas Jones, Sarah Lawman, Andrew Lewington, Sinha Manish, Patrick Mark, Jason Moore, Albert Ong, Nick Webb.

The NIHR BioResource-Rare Diseases Consortium

Timothy J Aitman, David L Bennett, Mark Caulfield, Patrick F Chinnery, Peter H Dixon, Kathleen Freson, Daniel P Gale, Ania Koziell, Taco W Kuijpers, Michael A Laffan, Eamonn R Maher, Hugh S Markus, Nicholas Morrell, Willem H Ouwehand, F Lucy Raymond, Irene Roberts, Kenneth G C Smith, Adrian J Thrasher, Hugh Watkins, Catherine Williamson, Christopher Geoffrey Woods, Sofie Ashford, John R Bradley, Debra Fletcher, Roger James, Mary A Kasanicki, Nathalie Kingston, Christopher J Penkett, Hannah Stark, Kathleen E Stirrups, Marijcke W M Veltman, Timothy Young, Lindsey Berends, Matthew Brown, Naomi Clements Brod, John Davis, Eleanor F Dewhurst, Helen Dolling, Marie Erwood, Amy J Frary, Rachel Linger, Jennifer Martin, Sofia Papadia, Karola Rehnstrom, Crina Samarghitean, Emily Staples, Katherine Yates, Csaba Halmagyi, Vera Matser, Catherine Titterton, Julie von Ziegenweidt, Ping Yu, Anthony Attwood, Abigail Crisp-Hihn, Karen Edwards, James Fox, Jennifer Jolley, Rutendo Mapeta, Paula J Rayner-Matthews, Ilenia Simeoni, Simon Staines, Jonathan Stephens, Christopher Watt, Deborah Whitehorn, Yvette Wood, Sri V V Deevi, Fengyuan Hu, Stuart Meacham, Olga Shamardina, Salih Tuna, Louise C Daugherty, Kate Downes, Daniel Duarte, Karyn Megy, William J Astle, Marta Bleda, Keren Carss, Stefan Gräf, Daniel Greene, Matthias Haimel, Hana Lango Allen, Alba Sanchis-Juan, Ernest Turro, Eliska Zlamalova, Sylvia Richardson, Paul Calleja, Robert Klima, Ignacio Medina Castello, Stuart Rankin, Wojciech Turek, Jeff Almeida-King, Giselle Kerry, Aoife McMahon, Joannella Morales, Gerome Breen, John Chambers, Matthew Hurles, Mark I McCarthy, Nilesh Samani, Michael Simpson, Nicholas Wood, Julie Anderson, Christine Bryson, Tracey Hammerton, Coleen McJannet, Sophie Stock, Lindsey Walker, Jayanthi Alamelu, Raza Alikhan, David J Allsup, Arif Alvi, Steve Austin, Trevor Baglin, Tamam Bakchoul, Tadbir Bariana, Gareth Baynam, Neha Bhatnagar, Tina Biss, Sara Boyce, Elizabeth Chalmers, Melissa V Chan,

3

Janine Collins, Peter W Collins, Nicola S Curry, Tina Dutt, Wendy N Erber, Gillian Evans, Tamara Everington, Remi Favier, Bruce Furie, Michael Gattens, Keith Gomez, Andreas Greinacher, Paolo Gresele, Daniel Hart, Yvonne M C Henskens, Rashid Kazmi, David Keeling, Anne M Kelly, Michele Lambert, Claire Lentaigne, Ri Liesner, Bella Madan, Mike Makris, Sarah Mangles, Mary Mathias, Carolyn Millar, Andrew D Mumford, Paquita Nurden, Samya Obaji, K John Pasi, Jeanette H Payne, Kathelijne Peerlinck, Fernando Pinto, Shoshana Revel-Vilk, Mike Richards, Matthew T Rondina, Catherine Roughley, Sol Schulman, Harald Schulze, Marie Scully, Susan E Shapiro, Keith Sibson, Matthew C Sims, Suthesh Sivapalaratnam, Matthew J Stubbs, R Campbell Tait, Kate Talks, Gordon B Taylor, Jecko Thachil, Chantal Thys, Cheng-Hock Toh, Chris Van Geet, Minka J A Vries, Timothy Warner, Henry G Watson, Sarah K Westbury, John-Paul Westwood, Julian Barwell, Anand Dixit, Anna M Drazyk, Kirsty Harkness, Ahamad Hassan, Sarju Mehta, Keith W Muir, Bartlomiej Piechowski-Jozwiak, Julia Rankin, Emma E Richards, Peter Rothwell, Pankaj Sharma, Rhea Y Y Tan, Matthew Traylor, Sarah Trippier, Dean Waugh, Angela Welch, David Werring, Munaza Ahmed, Natalie Canham, Alex Henderson, Hanadi Kazkaz, Catherine L Mercer, David Parry, Miranda Splitt, Anthony M Vandersteen, John Baski, Rachel Bates, Paul Brennan, Rachel J Buchan, Teofila Bueser, Gerald Carr-White, Stuart Cook, Matthew J Daniels, Timothy Dent, Andrew R Harper, Anna C Michell, Elizabeth Ormondroyd, Antonis Pantazis, Georgina Parsons, Sanjay Prasad, Leema Robert, Linda Sneddon, Kate Thomson, Roddy Walsh, James S Ware, Irshad Ahmed, Rita Arya, Annette Briley, Jenny Chambers, Floria Cheng, Tony Elston, Maria C Estiu, Joanna C Girling, William M Hague, Sean Hughes, Mark Johnson, Bryony Jones, Mahantesh Karoshi, Alison Kirkpatrick, Ellen Knox, Malcolm Macdougall, Lucy H Mackillop, Laura Magee, Hanns-Ulrich Marschall, Muna Noori, Hassan Shehata, Marta Vazquez-Lopez, David J Williams, Stephen Abbs, Louise Allen, Gavin Arno, Courtney E French, Robert H Henderson, Dragana Josifova, Michel Michaelides, Anthony T Moore, Richard Scott, Andrew R Webster, Philip Griffiths, Rita Horvath, Gavin Hudson, Neringa Jurkute, Angela Pyle, Wei Wei, Patrick Yu-Wai-Man, Julian Adlard, Ruth Armstrong, Carole Brewer, Ruth Casey, Trevor R P Cole, Dafydd Gareth Evans, Lynn Greenhalgh, Helen L Hanson, Jonathan Hoffman, Louise Izatt, Ajith Kumar, Fiona Lalloo, Kai Ren Ong, Soo-Mi Park, Joan Paterson, Claire Searle, Lucy Side, Katie Snape, Marc Tischkowitz, James Whitworth, Emma Woodward, Gautam Ambegaonkar, Maria A K Bitner-Glindzicz, Angela Brady, Richard Brown, Jackie Buck, Jenny Carmichael, Manali Chitre, Emma M Clement, Virginia Clowes, Patrick B Deegan, Charu Deshpande, Rainer Doffinger, Helen Firth, Frances A Flinter, Alice Gardham, Neeti Ghali, Paul Gissen, Kathleen M Gorman, Detelina Grozeva, Anke Hensiek, Simon Holden, Muriel Holder, Henry Houlden, Jane A Hurst, Deepa Krishnakumar, Manju A Kurian, Melissa Lees, Robert MacLaren, Anna Maw, Amy McTague, Elaine Murphy, Joanne Ng, Adeline Ngoh, Apostolos Papandreou, Alasdair P J Parker, Chris Patch, David Perry, Sanjay Raina, Evan Reid, Elisabeth Rosser, Richard Sandford, Saikat Santra, Aman Sohal, Penelope E Stein, Emily Symington, Ellen Thomas, Dorothy A Thompson, Julie Vogt, Emma L Wakeling, Evangeline Wassmer, Sam Chong, Ingrid Emmerson, Lionel Ginsberg, David Gosal, Robert D M Hadden, Mohamed Mahdi-Rogers, Adnan Manzur, Andrew Marshall, Emma Matthews, Mary M Reilly, Tara Renton, Andrew S C Rice, Andreas C Themistocleous, Tom Vale, Natalie van Zuydam, Suellen M Walker, Harm Bogaard, Colin Church, Gerry Coghlan, Robin Condliffe, Victoria Cookson, Paul Corris, Amanda Creaser-Myers,

4

Rosa Da Costa, Cesare Danesino, Natalie Dormand, Mlanie Eyries, Henning Gall, Stefano Ghio, Hossein-Ardeschir Ghofrani, J Simon R Gibbs, Barbara Girerd, Alan Greenhalgh, Charaka Hadinnapola, Arjan Houweling, Luke S Howard, Anna E Huis in 't Veld, Marc Humbert, Val Irvine, Fiona Kennedy, David G Kiely, Gabor Kovacs, Allan Lawrie, Rajiv D Machado, Robert V MacKenzie Ross, Larahmie Masati, Sharon Meehan, Shahin Moledina, David Montani, Michael Newnham, Andrea Olschewski, Horst Olschewski, Shokri Othman, Andrew Peacock, Joanna Pepke-Zaba, Gary J Polwarth, Lavanya Ranganathan, Laura Scelsi, Gwen Schotte, Werner Seeger, Deborah Shipley, Florent Soubrier, Laura Southgate, Jay Suntharalingam, Emilia Swietlik, Mark Toshner, Carmen Treacy, Richard Trembath, Anton Vonk Noordegraaf, Quinten Waisfisz, Ivy Wanjiku, John Wharton, Martin R Wilkins, Stephen J Wort, Zoe Adhya, Hana Alachkar, Ariharan Anantharachagan, Richard Antrobus, Gururaj Arumugakani, Chiara Bacchelli, Helen E Baxendale, Claire Bethune, Shahnaz Bibi, Claire Booth, Michael J Browning, Matthew S Buckland, Siobhan O Burns, Oliver S Burren, Anita Chandra, Nichola Cooper, Sophie Davies, Sarah Deacock, John Dempster, Lisa A Devlin, Elizabeth Drewe, J David M Edgar, William Egner, Tariq El-Shanawany, Henry Farmery, H Bobby Gaspar, Rohit Ghurye, Kimberly C Gilmour, Sarah Goddard, Pavels Gordins, Sofia Grigoriadou, Scott Hackett, Rosie Hague, Lorraine Harper, Grant Hayman, Archana Herwadkar, Stephen Hughes, Aarnoud P Huissoon, Stephen Jolles, Yousuf Karim, Peter Kelleher, Sorena Kiani, Nigel Klein, Dinakantha Kumararatne, James Laffan, Sara E Lear, Hilary Longhurst, Lorena E Lorenzo, Paul A Lyons, Jesmeen Maimaris, Ania Manson, Elizabeth M McDermott, Anoop Mistry, Valerie Morrisson, Sai H K Murng, Sadia Noorani, Eric Oksenhendler, Mark J Ponsford, Waseem Qasim, Ellen Quinn, Isabella Quinti, Alex Richter, Ravishankar B Sargur, Sinisa Savic, Suranjith L Seneviratne, W A Carrock Sewell, Fiona Shackley, Hans Stauss, Cathal L Steele, James E Thaventhiran, David Thomas, Moira J Thomas, Rafal Urniaz, Steven B Welch, Lisa Willcocks, Sarita Workman, Austen Worth, Nigel Yeatman, Patrick F K Yong, Saeed Ahmed, Waqar Ayub, Jonathan Barratt, Preetham Boddana, Gary Campbell, Melanie M Y Chan, Martin T Christian, H Terence Cook, Rodney D Gilbert, Sian Griffiths, Matthew Hall, Claire Harris, Shivaram Hegde, Praveen Jeevaratnam, Sally A Johnson, Caroline Jones, Sarah H A Lawman, Adam P Levine, Andrew Lewington, Kevin J Marchbank, Patrick B Mark, Stephen Marks, Heather Maxwell, Paul McAlinden, David Milford, Jason Moore, Monika Mozere, Albert C M Ong, Omid Sadeghi-Alavijeh, Manish D Sinha, Nick Webb, Julie Wessels, Edwin K S Wong, Philip Ancliff, Yesim Aydinok, Christian Babbs, Claire Harrison, Sally Kinsey, D Mark Layton, Eleni Louka, Simon J McGowan, Adam J Mead, Steven Okoli, Jennifer O'Sullivan, Beth Psaila, Anupama Rao, Noemi B Roy, Ajay Vora, Maryam Afzal, Agnieszka Bierzynska, Carmen Bugarin Diz, Elizabeth Colby, Melanie N Ekani, Moin Saleem, Simon Satchell.

2   Supplementary Methods

National Institute for Health Research (NIHR) BioResource Rare Diseases (BR-RD) Study

In addition to primary membranoproliferative glomerulonephritis (PMG), the phenotypes examined include: bleeding/thrombotic/platelet disorders (BPD)[1], cerebral small vessel disease (CSVD), Ehlers-Danlos syndrome (EDS), hypertrophic cardiomyopathy (HCM), intrahepatic cholestasis of pregnancy (ICP), Leber Hereditary Optic Neuropathy (LHON), multiple primary malignant tumors (MPMT)[2], pulmonary arterial hypertension (PAH)[3], [4], primary immune disorders (PID)[5], inherited retinal disorders (IRD)[6], [7], neurological and developmental disorders (NDD), neuropathic pain disorders (NPD), stem cell and myeloid disorders (SMD) and steroid resistant nephrotic syndrome. Data were also generated from process controls (CNTRL), individuals with one of 162 rare diseases with no known causal mutation by standard of care genetic testing (GEL) and from healthy individuals from the UK Biobank[8] with extreme red blood cell traits (UKBio). A breakdown of the number of individuals per cohort can be found in Table S2.

Ethics

Written informed consent was provided by all participants. The study was approved by the East of England Cambridge South National Research Ethics Committee (Reference 13/EE/0325) and the South West Central Bristol Research Ethics Committee (Reference 10/H0106/8).

Whole-genome sequencing - data generation, variant calling and annotation

In brief, DNA was extracted from whole blood, underwent initial quality control assessment and was prepared using the Illumina TruSeq DNA PCR-Free sample preparation kit (Illumina, Inc.). Subsequently 100-150 base pair paired-end sequencing was undertaken using an Illumina HiSeq 2500 or HiSeq X. The minimum coverage required per sample was at least 95% of the autosomal genome at 15 times read depth. Reads were aligned against the human genome (GRCh37) using Isaac (Illumina)[9].

Single nucleotide variants (SNVs) and indels were called using the Illumina Starling software. Sample duplicates (n = 136) and those with poor data quality (n = 14) were excluded. SNVs and indels were normalized and combined into gVCFs. For each variant, the overall pass rate (OPR) was enumerated as the product of the pass rate (the proportion of alternate genotype passing the original variant filtering) and the call rate (proportion of non-missing genotypes). A genotype quality (GQ) threshold of 20 and depth (DP) threshold of 10 were imposed per genotype per individual; calls failing to

6

meet either of these criteria were set to missing. Only variants with OPR ≥ 0.8 and frequency of missingness ≤ 0.01 were retained.

Variants were annotated using the Ensembl Variant Effect Predictor (v89)[10], their predicted deleteriousness based on CADD score[11] and their frequency in gnomAD (`http://gnomad.broadinstitute.org/variant`)[12]. Variants were filtered using bcftools (v1.8)[13] and further filtered and analyzed using custom scripts written in Python and R.

Relatedness and ancestry

A subset of high quality common variants was extracted for ancestry and relatedness estimation in the full BR-RD dataset. These variants were selected as they were present on three Illumina genotyping arrays (HumanOmni2.58v1.1, HumanCoreExome-12v1.1 and HumanCoreExome-24v1.0), were biallelic, were genotyped in all BR-RD individuals, had a minor allele frequency (MAF) ≥ 0.3, were not in linkage disequilibrium (LD) (pruned using PLINK v1.9[14] with $r^2 < 0.2$) and had OPR ≥ 0.99. An initial kinship matrix was computed using KING[15]. Subsequently, PC-AiR[16] and PC-Relate[17] in the R package GENESIS were utilized to correct the kinship matrix for population structure. The resulting kinship matrix was used as input in PRIMUS[18] to identify the maximal set of unrelated individuals. The ancestry of all BR-RD samples was ascertained by calculating principal components (PC) using unrelated 1000 Genomes individuals[19] and projecting the BR-RD genotypes onto this vector space. A multivariate model was then used to classify each subject as being either of non-Finnish European, Finnish, African, South Asian and East Asian based on the 1000 Genomes data.

Further quality control

As described below in the section entitled 'Common variant GWAS', per individual heterozygosity and missingness were computed using common variants and individuals with values greater than three standard deviations from the mean for each parameter were excluded (n = 49). This resulted in a final dataset comprising 146 PMG cases and 6,442 non-PMG controls that was used for all subsequent analyses (see Table S2).

Structural variants

Structural variants (SVs) and copy number variants (CNVs) were called using Manta[20] and Canvas[21], respectively. Variants were categorized as CNV gain/loss, translocations, deletions, tandem duplications and insertions. Only those variants within at least 10 base pairs of an exon were included. Known common benign variants[22] and those failing Illumina quality filters were excluded. Variants were filtered based on their allele frequency in all BR-RD samples excluding those at > 0.001. Deletions identified by both Canvas and Manta with a 20% minimum overlap were identified. The genome-wide comparison of the frequency of deletions per gene between cases and controls was

7

undertaken using those deletions identified by both Canvas and Manta using PLINK v1.07[23].

Comparison with previously described PMG and aHUS variants

Details on 843 common and rare variants within the genes *C3*, *CD46*, *CFB*, *CFH*, *CFHR1*, *CFHR3*, *CFHR5*, *CFI*, *DGKE*, and *THBD* previously observed in patients with atypical hemolytic uremic syndrome (aHUS), age-related macular degeneration (AMD), C3G or thrombotic microangiopathy (TMA) were extracted from the Database of Complement Gene Variants (DCGV) (`http://www.complement-db.org`)[24]. The variants were manually curated to remove duplicates leaving 830. Three additional common variants previously studied in MPGN but missing from the database were added[25]. Observed sequence variants were matched to these previously described variants based on either overlapping (i) genomic position, reference and alternate alleles, (ii) Human Genome Variation Society (HGVS)[26] protein effect or (iii) HGVS cDNA effect. For common variants (gnomAD non-Finnish Europeans (gnomAD-NFE) MAF $\geq$ 0.05)), logistic regression association analysis with five principal components as covariates, and epistasis analyses were performed using PLINK v1.9.

Rare variant candidate gene and genome-wide coding variant burden analysis

To extract coding variants, exon positions as defined by both Ensembl[27] and RefSeq[28] were utilized. Variants residing within these loci passing quality control were retained. Only data from European unrelated PMG and controls were examined. Variants equal to or exceeding an allele frequency of 0.0001 in gnomAD-NFE were excluded. Only variants of moderate (inframe indels and missense variants) or high (splice acceptor/donor, stop gain/loss, start loss, frameshift) impact were analyzed. Per-gene rare variant burden was enumerated as the proportion of individuals (cases versus controls) with at least one alternate allele in each gene with significance calculated using the exactCMC function in RVTESTS[29], which employs a Fisher's exact test. A Manhattan plot was produced using the R package qqman[30]. A QQ plot including 95% confidence intervals was produced using the R packge snpStats. As per the method employed by qqman, only p-values < 1 have been included in the QQ plot.

Common variant GWAS

Variants passing quality control filters that had a MAF > 0.05 in gnomAD-NFE and across all samples in BR-RD were retained (n = 5,939,292). Standard quality control procedures[31] were subsequently employed to remove samples and variants of poor quality. Initially, heterozygosity and per-sample missingness were computed using PLINK v1.9. Individuals with heterozygosity or missingness greater than three standard deviations from the mean were excluded. This resulted in the exclusion of 66 individuals (none

8

with PMG). Next, the data were filtered to remove variants demonstrating deviation from Hardy-Weinberg equilibrium at $p < 0.001$ and those at a MAF < 0.05 in controls. GWAS was undertaken with PLINK v1.9 assuming additive allele effects using logistic regression with the first five principal components as covariates. Following completion of the GWAS, a Manhattan plot was produced using the R package qqman[30]. A QQ plot including 95% confidence intervals was produced using the R packge snpStats. High resolution plots showing the LD between markers (as per 1000 Genomes November 2014 European data) were generated using LocusZoom[32]. eQTL data were extracted from the Genotype-Tissue Expression (GTEx) project (`https://www.gtexportal.org`)[33].

HLA imputation

HLA genotyping was performed by realignment of the raw sequence data to HLA contigs using BWAKIT / BWAMEM v0.7.15 (`https://github.com/lh3/bwa/tree/master/bwakit`). HLA alleles A, B, C, DQA1, DQB1 and DRB1 were examined. For each allele, only the most likely genotypes were carried forward.

HLA genotyping was also performed using HLA-HD v1.2.0.1[34] which maps the raw sequence data to an extensive dictionary of HLA alleles. This was run using default parameters. HLA alleles A, B, C, DQA1, DQB1, DRB1-9, DPA1, DMA, DMB, DOA, DOB, DRA, E, F, G, H, J, K, L and V were imputed.

The imputed results generated by the two methods were processed separately. The results were converted to PED/MAP format at four-digit resolution and filtered on missingness per individual (99%) and per variant (99%).

3 Supplementary Results

HLA replication

Identical serotypes at all five HLA types tested were observed for seven (2.1%) of the cases and 1106 (7.1%) of the controls. The coverage of the observed serotypes for each of the five HLA types were as follows: A 99.5%, B 99.5%, C 87.6%, DR 99.5%, DQ 92.8%.

Of the European PMG individuals in the discovery analysis, 22 had undergone renal transplantation and might have been included within the MPGN NHSBT data. Amongst these cases, four and two individuals were heterozygous and homozygous, respectively, for DRB1*03:01 (corresponding to DR17). Removing four heterozygous, two homozygous and sixteen wild type cases from the MPGN DR17 analysis, this serotype remained significantly associated with disease ($p = 2.0 \times 10^{-4}$), confirming independent replication.

9

4   Supplementary References

1. S. K. Westbury et al., Expanded repertoire of RASGRP2 variants responsible for platelet dysfunction and severe bleeding, Blood, vol. 130, no. 8, pp. 10261030, 24 2017.

2. J. Whitworth et al., Comprehensive Cancer-Predisposition Gene Testing in an Adult Multiple Primary Tumor Series Shows a Broad Range of Deleterious Variants and Atypical Tumor Phenotypes, Am. J. Hum. Genet., vol. 103, no. 1, pp. 318, Jul. 2018.

3. S. Gräf et al., Identification of rare sequence variation underlying heritable pulmonary arterial hypertension, Nat Commun, vol. 9, no. 1, p. 1416, Apr. 2018.

4. C. Rhodes et al., Genetic determinants of risk and survival in pulmonary arterial hypertension, May 2018.

5. P. Tuijnenburg et al., Loss-of-function nuclear factor B subunit 1 (NFKB1) variants are the most common monogenic cause of common variable immunodeficiency in Europeans, J. Allergy Clin. Immunol., vol. 142, no. 4, pp. 12851296, Oct. 2018.

6. K. J. Carss et al., Comprehensive Rare Variant Analysis via Whole-Genome Sequencing to Determine the Molecular Pathology of Inherited Retinal Disease, Am. J. Hum. Genet., vol. 100, no. 1, pp. 7590, 05 2017.

7. G. Arno et al., Biallelic Mutation of ARHGEF18, Involved in the Determination of Epithelial Apicobasal Polarity, Causes Adult-Onset Retinal Degeneration, Am. J. Hum. Genet., vol. 100, no. 2, pp. 334342, Feb. 2017.

8. C. Sudlow et al., UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age, PLoS Med., vol. 12, no. 3, p. e1001779, Mar. 2015.

9. C. Raczy et al., Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms, Bioinformatics, vol. 29, no. 16, pp. 20412043, Aug. 2013.

10. W. McLaren et al., The Ensembl Variant Effect Predictor, Genome Biol., vol. 17, no. 1, p. 122, 06 2016.

11. M. Kircher, D. M. Witten, P. Jain, B. J. ORoak, G. M. Cooper, and J. Shendure, A general framework for estimating the relative pathogenicity of human genetic variants, Nat. Genet., vol. 46, no. 3, pp. 310315, Mar. 2014.

12. M. Lek et al., Analysis of protein-coding genetic variation in 60,706 humans, Nature, vol. 536, no. 7616, pp. 285291, 18 2016.

13. H. Li et al., The Sequence Alignment/Map format and SAMtools, Bioinformatics, vol. 25, no. 16, pp. 20782079, Aug. 2009.

14. C. C. Chang, C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee, Second-generation PLINK: rising to the challenge of larger and richer datasets, Gigascience, vol. 4, p. 7, 2015.

10

15. A. Manichaikul, J. C. Mychaleckyj, S. S. Rich, K. Daly, M. Sale, and W.-M. Chen, Robust relationship inference in genome-wide association studies, Bioinformatics, vol. 26, no. 22, pp. 28672873, Nov. 2010.

16. M. P. Conomos, M. B. Miller, and T. A. Thornton, Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness, Genet. Epidemiol., vol. 39, no. 4, pp. 276293, May 2015.

17. M. P. Conomos, A. P. Reiner, B. S. Weir, and T. A. Thornton, Model-free Estimation of Recent Genetic Relatedness, Am. J. Hum. Genet., vol. 98, no. 1, pp. 127148, Jan. 2016.

18. J. Staples, D. A. Nickerson, and J. E. Below, Utilizing graph theory to select the largest set of unrelated individuals for genetic analysis, Genet. Epidemiol., vol. 37, no. 2, pp. 136141, Feb. 2013.

19. 1000 Genomes Project Consortium et al., A global reference for human genetic variation, Nature, vol. 526, no. 7571, pp. 6874, Oct. 2015.

20. X. Chen et al., Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications, Bioinformatics, vol. 32, no. 8, pp. 12201222, 15 2016.

21. E. Roller, S. Ivakhno, S. Lee, T. Royce, and S. Tanner, Canvas: versatile and scalable detection of copy number variants, Bioinformatics, vol. 32, no. 15, pp. 23752377, 01 2016.

22. M. Zarrei, J. R. MacDonald, D. Merico, and S. W. Scherer, A copy number variation map of the human genome, Nat. Rev. Genet., vol. 16, no. 3, pp. 172183, 2015.

23. S. Purcell et al., PLINK: a tool set for whole-genome association and population-based linkage analyses, Am. J. Hum. Genet., vol. 81, no. 3, pp. 559575, Sep. 2007.

24. A. J. Osborne et al., Statistical Validation of Rare Complement Variants Provides Insights into the Molecular Basis of Atypical Hemolytic Uremic Syndrome and C3 Glomerulopathy, J. Immunol., vol. 200, no. 7, pp. 24642478, Apr. 2018.

25. A. Servais et al., Acquired and genetic complement abnormalities play a critical role in dense deposit disease and other C3 glomerulopathies, Kidney Int., vol. 82, no. 4, pp. 454464, Aug. 2012.

26. J. T. den Dunnen et al., HGVS Recommendations for the Description of Sequence Variants: 2016 Update, Hum. Mutat., vol. 37, no. 6, pp. 564569, 2016.

27. D. R. Zerbino et al., Ensembl 2018, Nucleic Acids Res., vol. 46, no. D1, pp. D754D761, Jan. 2018.

28. N. A. OLeary et al., Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation, Nucleic Acids Res., vol. 44, no. D1, pp. D733-745, Jan. 2016.

11

29. X. Zhan, Y. Hu, B. Li, G. R. Abecasis, and D. J. Liu, RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data, Bioinformatics, vol. 32, no. 9, pp. 14231426, 01 2016.

30. S. D. Turner, qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots, May 2014.

31. C. A. Anderson, F. H. Pettersson, G. M. Clarke, L. R. Cardon, A. P. Morris, and K. T. Zondervan, Data quality control in genetic case-control association studies, Nat Protoc, vol. 5, no. 9, pp. 15641573, Sep. 2010.

32. R. J. Pruim et al., LocusZoom: regional visualization of genome-wide association scan results, Bioinformatics, vol. 26, no. 18, pp. 23362337, Sep. 2010.

33. L. J. Carithers et al., A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project, Biopreserv Biobank, vol. 13, no. 5, pp. 311319, Oct. 2015.

34. S. Kawaguchi, K. Higasa, M. Shimizu, R. Yamada, and F. Matsuda, HLA-HD: An accurate HLA typing algorithm for next-generation sequencing data, Hum. Mutat., vol. 38, no. 7, pp. 788797, 2017.
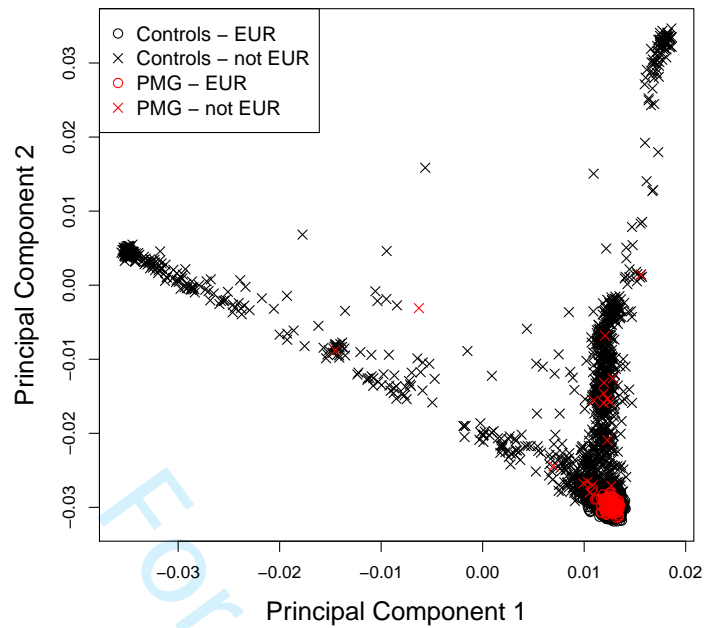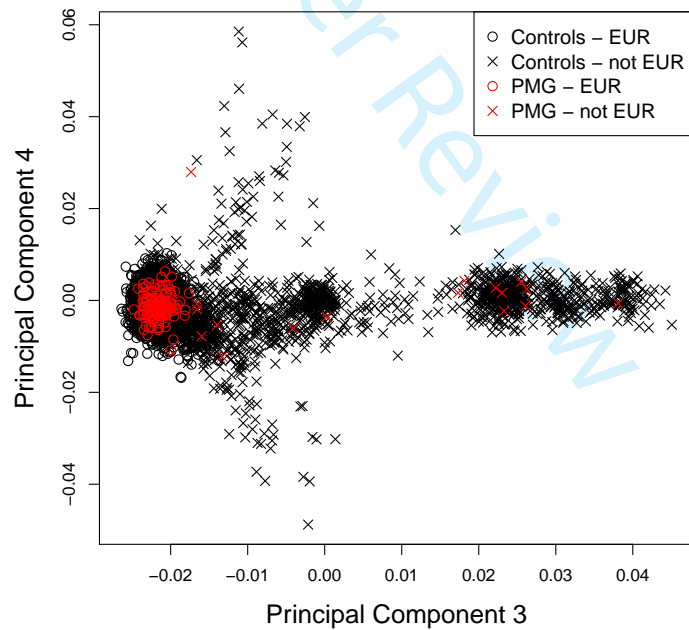
5 Supplementary Figures
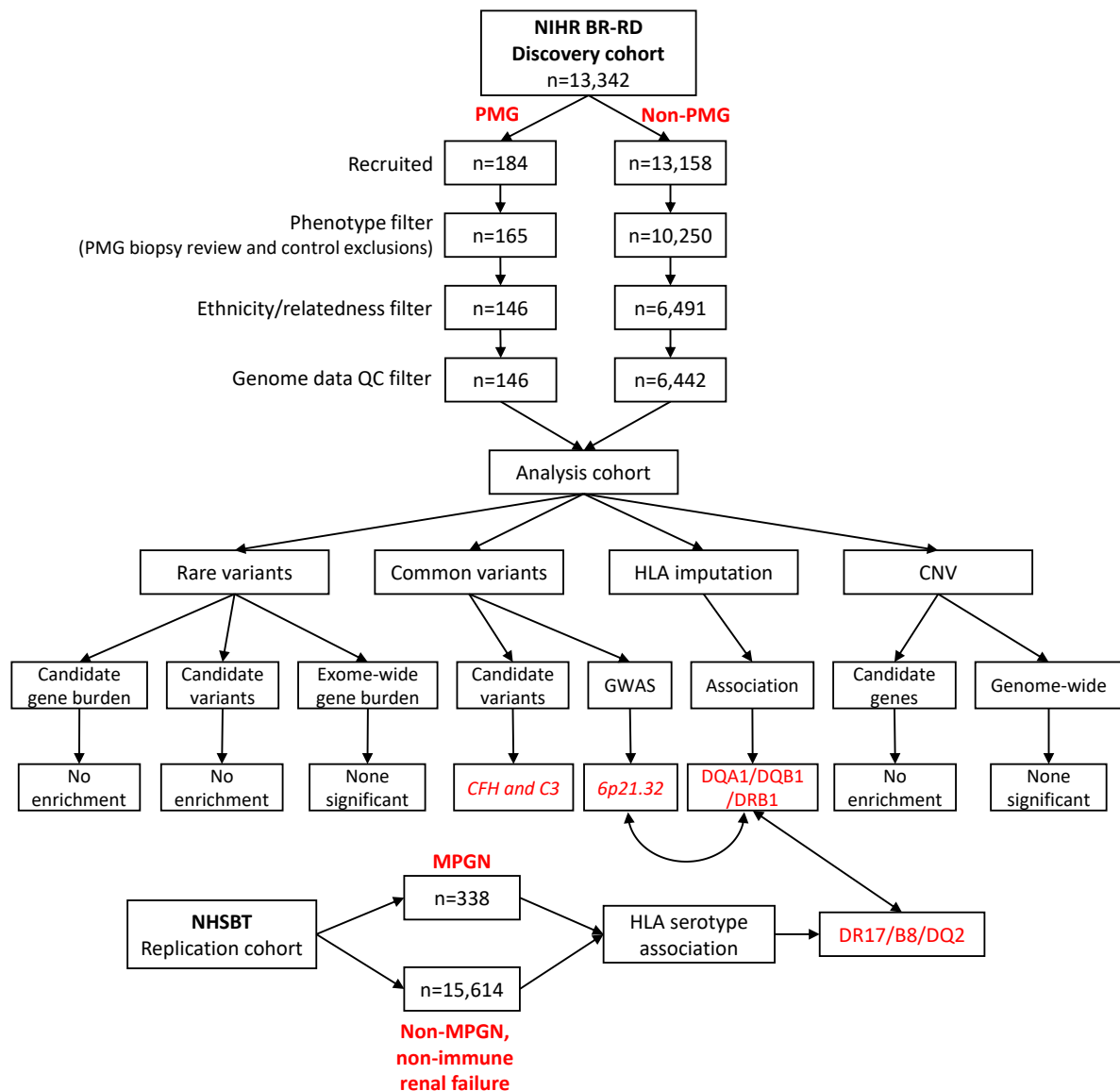
Figure S1: Sample and analytic workflow for both the discovery and replication components of the study. The flowchart shows the number of samples included, the analytical strategies employed and the main findings.

13

(a) Principal components 1 and 2

(b) Principal components 3 and 4

Figure S2: Principal component analysis showing the first four principal components highlighting PMG cases (red) and controls (black) by ethnicity (European circle, not European cross). All unrelated individuals post-exclusions, pre-common variant quality control have been plotted.

14

Candidate gene rare variant burden
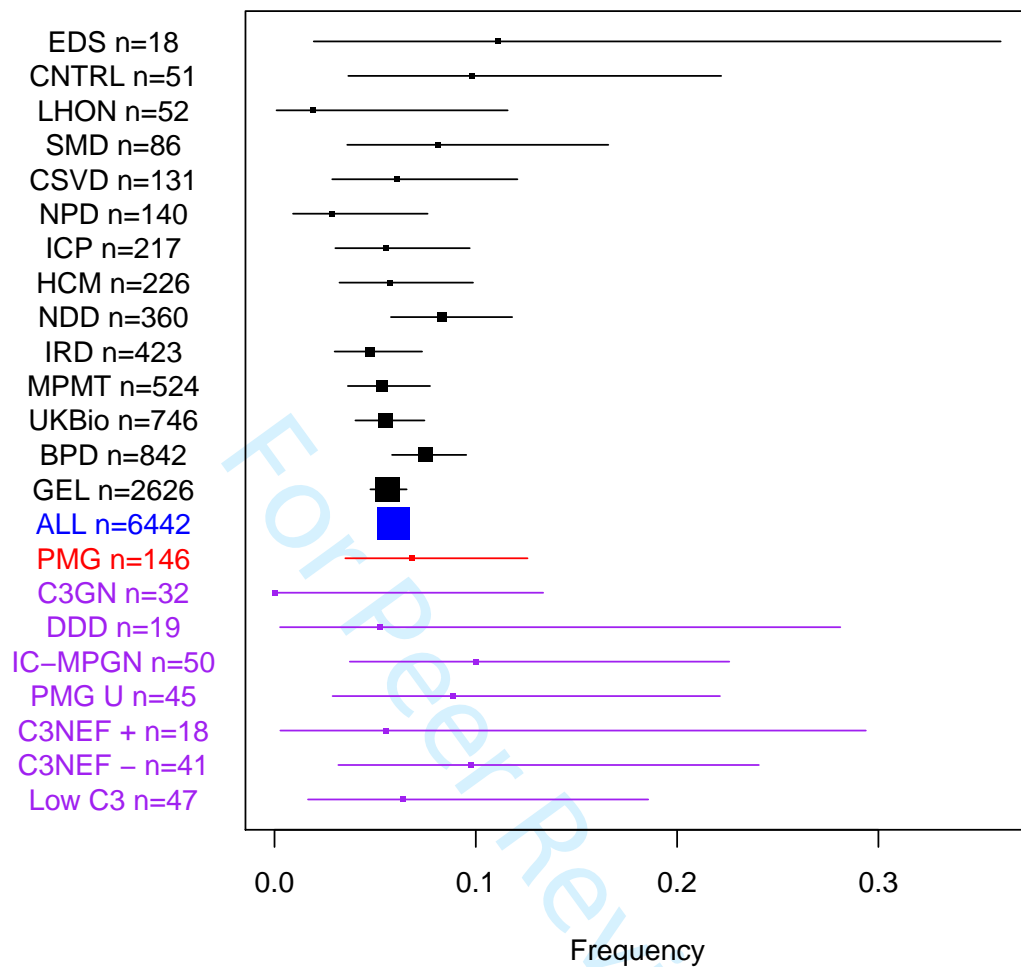
Figure S3: Cumulative burden of rare variants with moderate or high predicted impact in the PMG candidate genes in each of the control cohorts of BR-RD separately (black), together (blue), in PMG (red) and the subphenotypes of PMG based on histopathology (C3GN, DDD, IC-MPGN, PMG unclassified), C3NeF status (positive/negative) and those with low C3 (purple). Horizontal lines indicate the 95% confidence intervals.
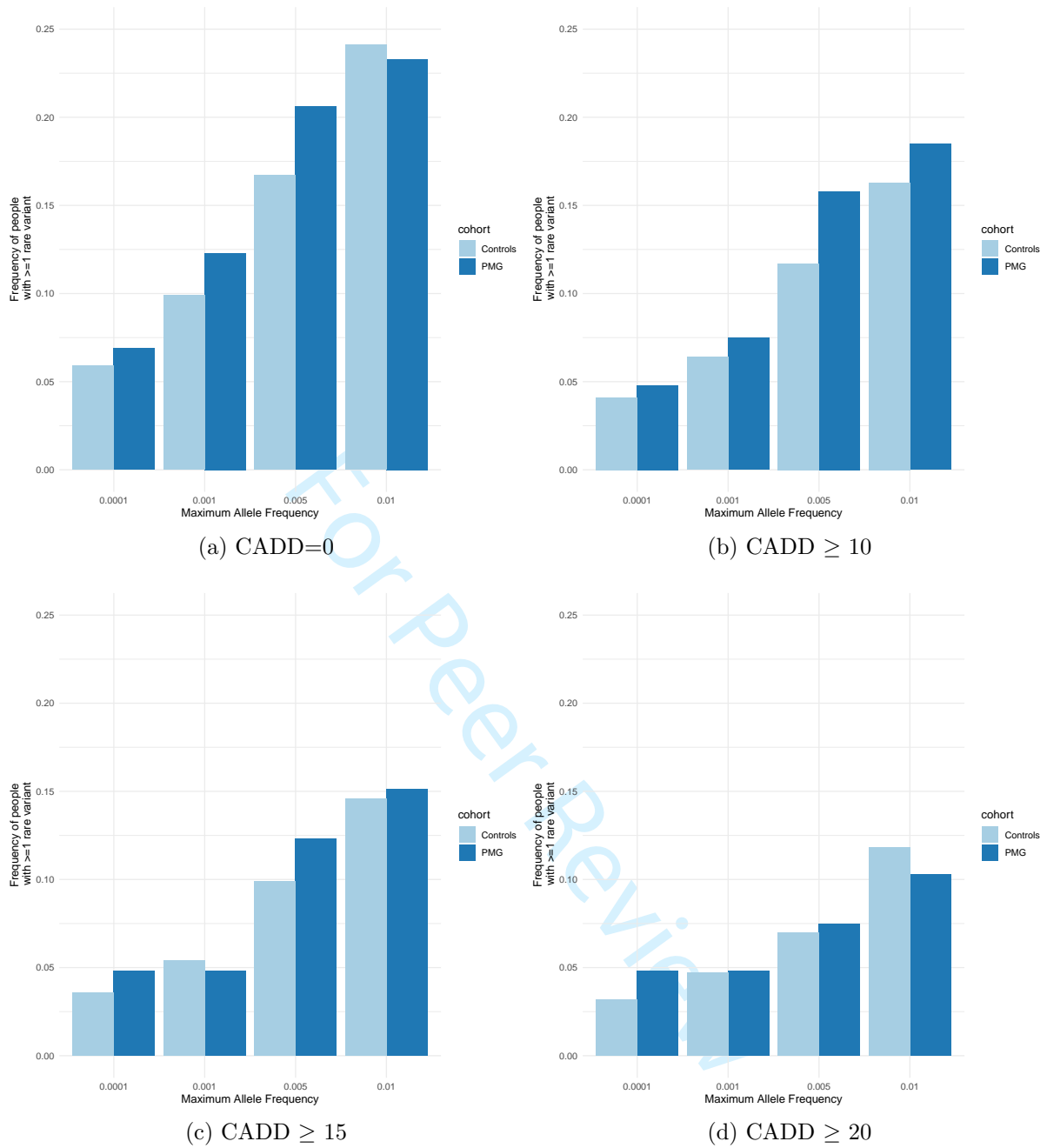
Figure S4: Cumulative burden of rare variants with moderate or high predicted impact in the PMG candidate genes with variable CADD threshold (none to ≥ 20) and control allele frequency (gnomAD-NFE MAF < 0.0001 to < 0.01), in PMG and control subjects.
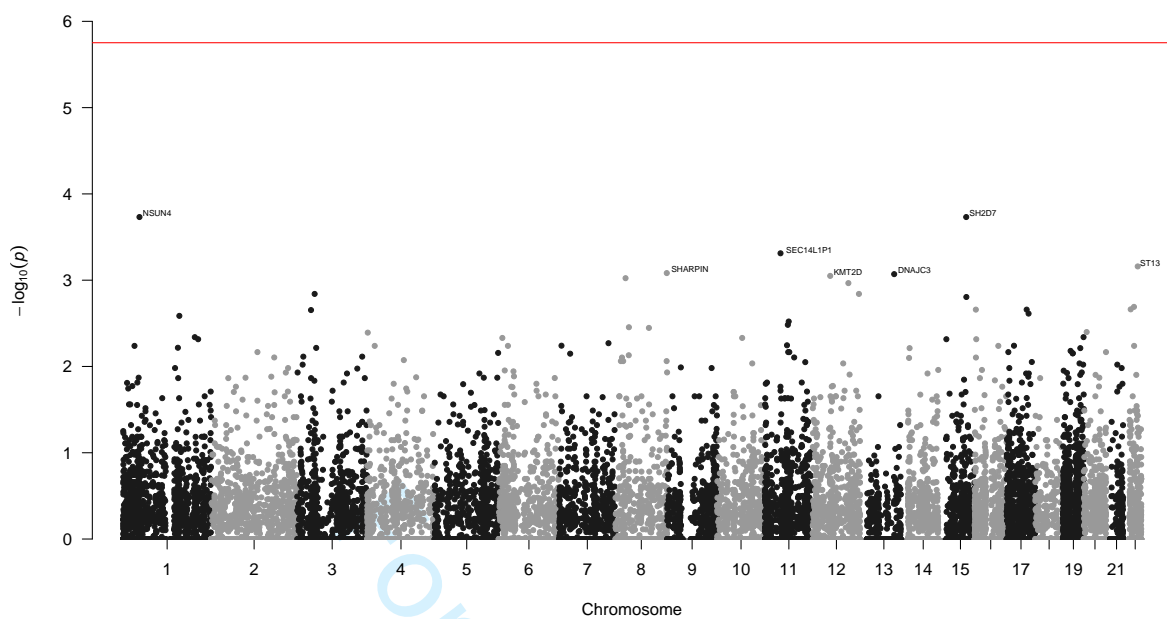
16

Figure S5: Exome-wide rare variant burden analysis Manhattan plot comparing European unrelated PMG cases and controls. No gene surpasses the exome-wide Bonferroni significance threshold indicated by a horizontal red line ($p < 1.77 \times 10^{-6}$). Genes achieving p $< 1 \times 10^{-3}$ have been annotated.
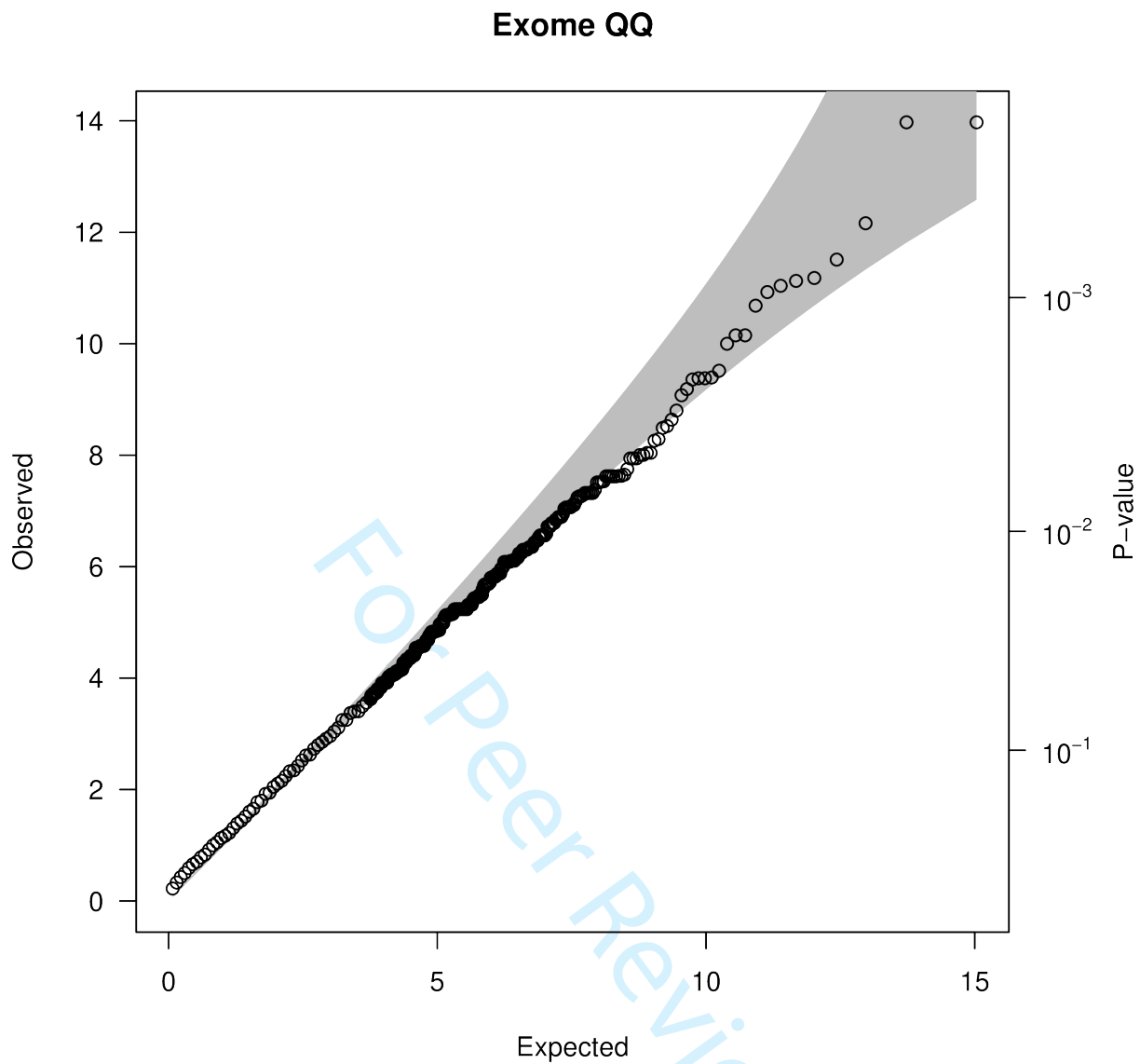
17

Exome QQ

Figure S6: QQ plot for exome-wide rare variant gene burden analysis. The observed/expected chi-square values and corresponding p-values are shown. The grey shaded area indicates the 95% confidence interval of the null.
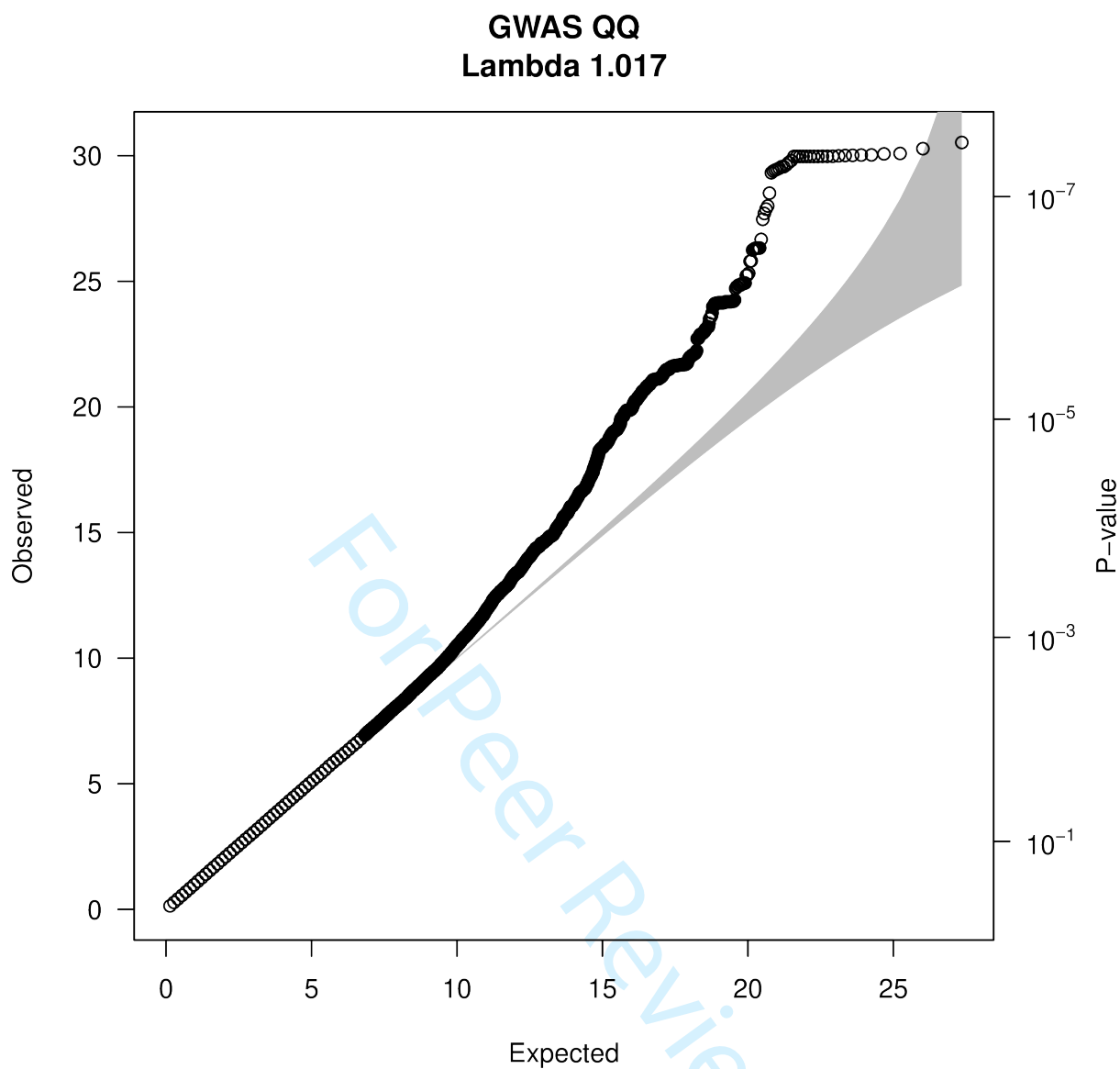
Figure S7: QQ plot for the common variant genome-wide association analysis. The observed/expected chi-square values and corresponding p-values are shown. The grey shaded area indicates the 95% confidence interval of the null. The genomic inflation (lambda) is 1.017.
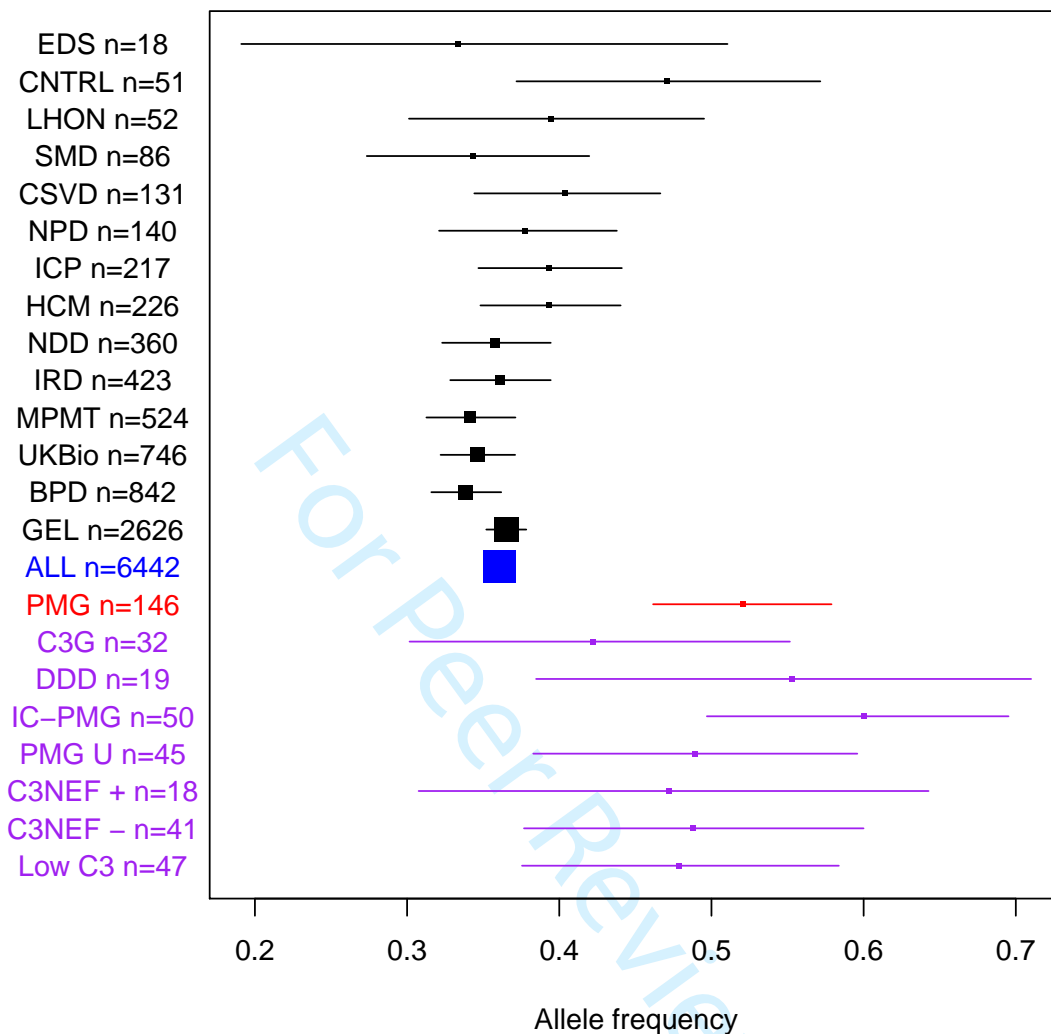
Figure S8: Allele frequency of the lead variant from the chromosome 6 locus in each of the control cohorts of BR-RD separately (black), together (blue), in PMG (red) and the subphenotypes of PMG based on histopathology (C3GN, DDD, IC-MPGN, PMG unclassified), C3NeF status (positive/negative) and those with low C3 (purple). Horizontal lines indicate the 95% confidence intervals.
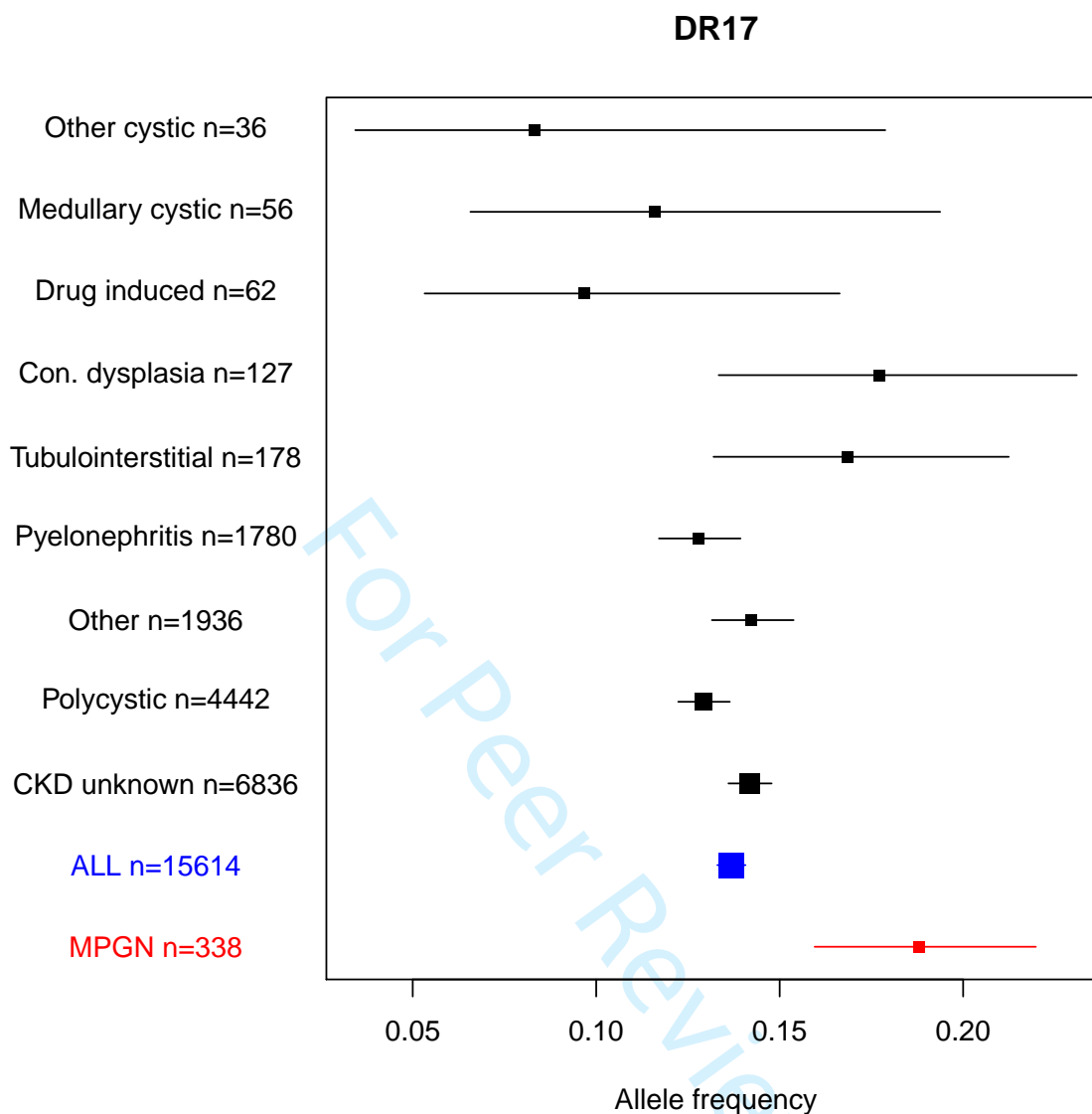
DR17

Figure S9: Frequency of HLA serotype DR17 in the NHSBT data showing the controls utilized separately, together (blue) and MPGN (red). Horizontal lines indicate the 95% confidence intervals.

21

6 Supplementary Tables

See separate Excel Spreadsheet.

Table S1

Full histological categorization and clinical details for all subjects with C3G (C3GN and DDD). In the **Source of histology data** column, the abbreviations used are as follows: LM light microscopy, EM electron microscopy and IS immunostain. **ESRD** is end-stage renal disease.

Table S2

Number of individuals in each of the BR-RD cohorts at each stage of filtering.

Table S3

Prioritized rare moderate/high impact variants identified in candidate genes in PMG subjects. **Chr** (chromosome), **Pos** (position), **Ref** (reference allele) and **Alt** (alternate alele) are given with reference to Build 37 of the human genome. **HGVSc** and **HGVSp** effects are given for the Ensembl transcript that is canonical or otherwise the transcript with the highest impact variant effect. **gnomAD_AF_NFE** is the frequency of the variant in non-Finish European individuals in gnomAD. **gnomAD_AF** is the frequency of the variant in all gnomAD cohorts. **Phenotype** is the histological subphenotype (PMG is PMG unclassified). **AC_Controls** gives the number of non-PMG individuals with each variant. **Controls_Cohort_AC** shows the non-PMG cohorts in which subjects with each variant are identified.

Table S4

Prioritized rare moderate/high impact variants identified in candidate genes in non-PMG subjects. The column definitions (where they overlap) are as per Table S3.

Table S5

Rare variants in candidate genes previously classified as pathogenic or likely pathogenic in the Database of Complement Gene Variants (DCGV). **DCGV_cDNA** and **DCGV_Protein** are the cDNA and protein effects of the variant as reported in DCGV. **DCGV_Conditions** gives the diseases each variant has been previously identified in. **DCGV_Path** is the pathogenicity classification of the variant as per DCGV (using American College of Medical Genetics and Genomics and the Association for Molecular Pathology criteria, see DCGV publication by Osborne et al.). **AC** gives the number of individuals with each

22

variant. Further column definitions (where they overlap) are as per Table S3. A single variant seen in a PMG case is shown in red.

Table S6

Association statistics for sixteen common variants in candidate genes previously identified in association with aHUS or MPGN comparing PMG with controls. Linkage disequilibrium, as calculated from the data itself, enable identification of independent signals, as shown with r^2. **A1** is the minor allele. **HUGO Effect** gives the effect of the variant as described in the Database of Complement Gene Variants (DCGV). **F_A** and **F_U** give the frequency of the minor allele (A1) in PMG cases and controls, respectively. **OR**, **L95** and **U95** give the odds ratio and 95% confidence intervals. **P** is the p-value as calculated using logistic regression with five principal components as covariates. Further column definitions (where they overlap) are as per Table S3. P-values achieving significance after correcting for multiple testing are shown in red.

Table S7

Association statistics for all variants achieving $p < 5 \times 10^{-8}$ in the genome-wide association study. The column definitions (where they overlap) are as per Table S6.

Table S8

GTEx eQTL results for rs3117135 (`https://gtexportal.org/home/snp/rs3117135`). **NES** is the normalized effect size.

Table S9

Association statistics for imputed HLA alleles using BWAKIT/BWAMEM and HLA-HD. For each HLA allele, for each imputation method (BWAKIT/BWAMEM (**BWA-**) and HLA-HD (**HLAHD-**)) the frequency in PMG (**_A**), controls (**_U**), odds ratio (**OR**), confidence intervals (lower **L95** and upper **U95**) and p-value (**P**) are shown.

23