Large-Scale Whole-Genome Sequencing Reveals the Genetic Architecture of Primary Membranoproliferative GN 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

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,

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,

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 S1.

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

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 genome-wide association study', 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 S1).

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

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 ≥ 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[31]. As per the method employed by qqman, only p-values < 1 have been included in the QQ plot.

Common variant genome-wide association study

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[32] 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 49 individuals

(none 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. Genome-wide association study 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 genome-wide association study, 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[33]. eQTL data were extracted from the Genotype-Tissue Expression (GTEx) project (`https://www.gtexportal.org`)[34].

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[35] 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%) using PLINK.

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.

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. 1026-1030, Aug. 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. 3-18, Jul. 2018.
3. S. Gräf et al., Identification of rare sequence variation underlying heritable pulmonary arterial hypertension, Nat Commun, vol. 9, no. 1, article 1416, Apr. 2018.
4. C. Rhodes et al., Genetic determinants of risk in pulmonary arterial hypertension: international genome-wide association studies and meta-analysis, Lancet Respir Med, vol. 7, no. 3, pp. 227-238, Mar. 2019.
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. 1285-1296, 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. 75-90, Jan. 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. 334-342, 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, article 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. 2041-2043, Aug. 2013.
10. W. McLaren et al., The Ensembl Variant Effect Predictor, Genome Biol., vol. 17, no. 1, article 122, Jun. 2016.
11. M. Kircher, D. M. Witten, P. Jain, B. J. O'Roak, 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. 310-315, Mar. 2014.
12. M. Lek et al., Analysis of protein-coding genetic variation in 60,706 humans, Nature, vol. 536, no. 7616, pp. 285-291, Aug. 2016.
13. H. Li et al., The Sequence Alignment/Map format and SAMtools, Bioinformatics, vol. 25, no. 16, pp. 2078-2079, 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, article 7, Feb. 2015.

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. 2867-2873, 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. 276-293, 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. 127-148, 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. 136-141, Feb. 2013.

19. 1000 Genomes Project Consortium et al., A global reference for human genetic variation, Nature, vol. 526, no. 7571, pp. 68-74, 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. 1220-1222, Apr. 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. 2375-2377, Aug. 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. 172-183, Mar. 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. 559-575, 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. 2464-2478, 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. 454-464, 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. 564-569, Jun 2016.

27. D. R. Zerbino et al., Ensembl 2018, Nucleic Acids Res., vol. 46, no. D1, pp. D754-D761, 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.

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. 1423-1426, May. 2016.

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

31. D. Clayton D. snpStats: SnpMatrix and XSnpMatrix classes and methods. R package version 1.36.0. doi:10.18129/B9.bioc.snpStats. 2019.

32. 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. 1564-1573, Sep. 2010.

33. R. J. Pruim et al., LocusZoom: regional visualization of genome-wide association scan results, Bioinformatics, vol. 26, no. 18, pp. 2336-2337, Sep. 2010.

34. L. J. Carithers et al., A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project, Biopreserv Biobank, vol. 13, no. 5, pp. 311-319, Oct. 2015.

35. 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. 788-797, Jul. 2017.

5 Supplementary Figures

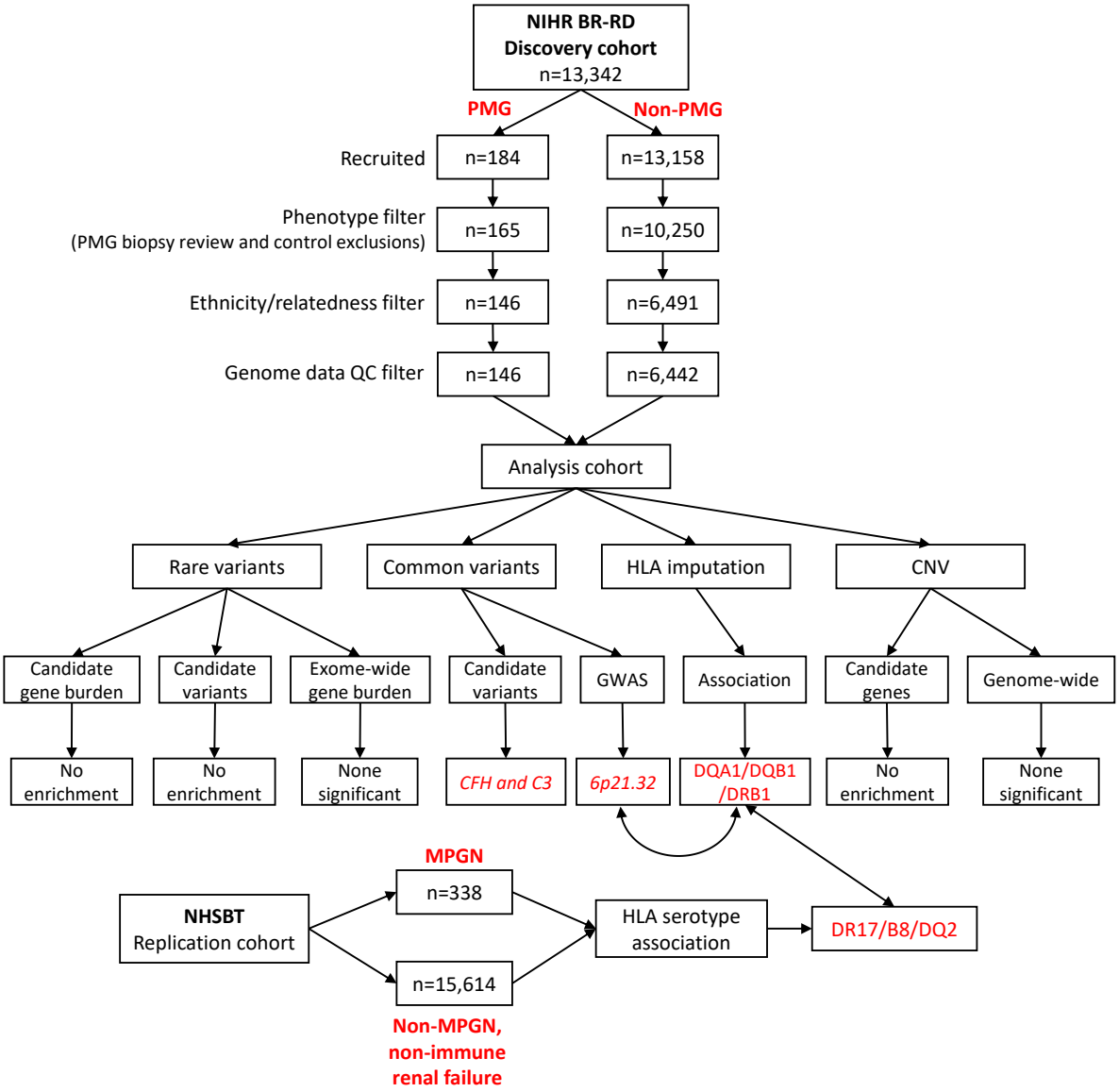
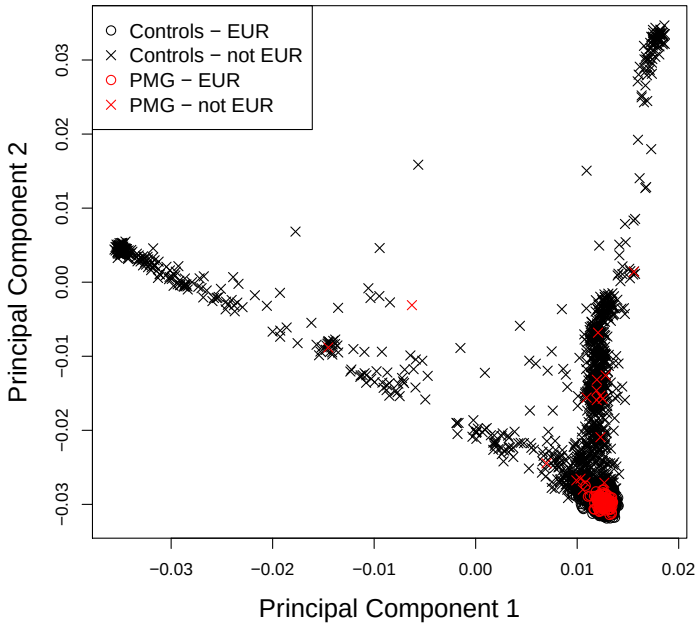
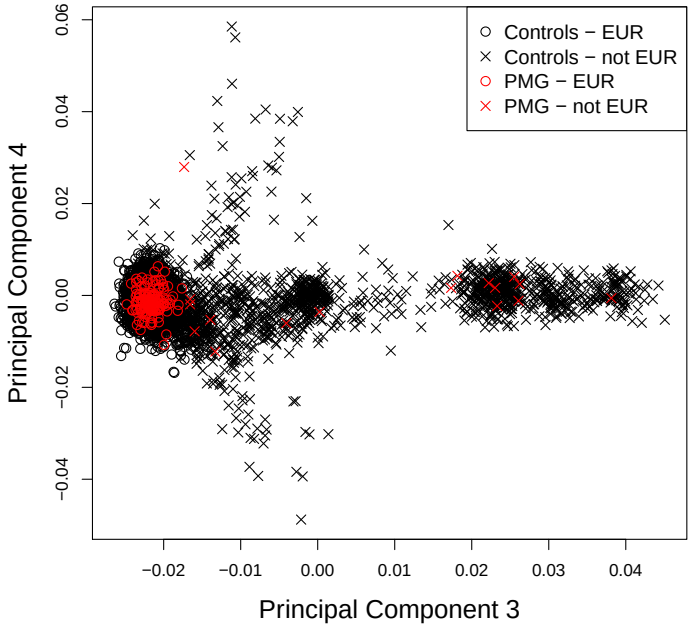


Figure S1: Sample and analytic workflow for both the discovery and replication components of the study. The flowchart shows the number of samples included, the analytical strategies employed and the main findings.



(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.

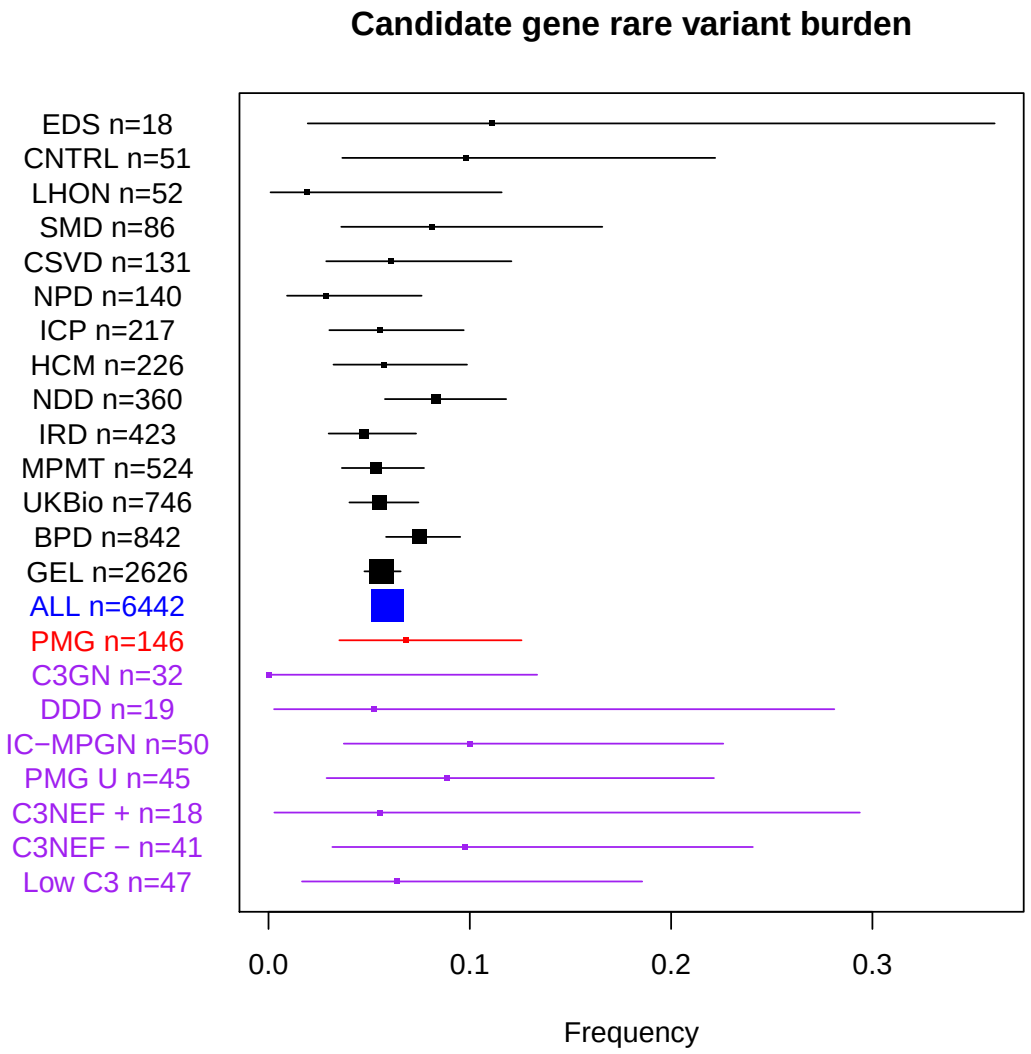
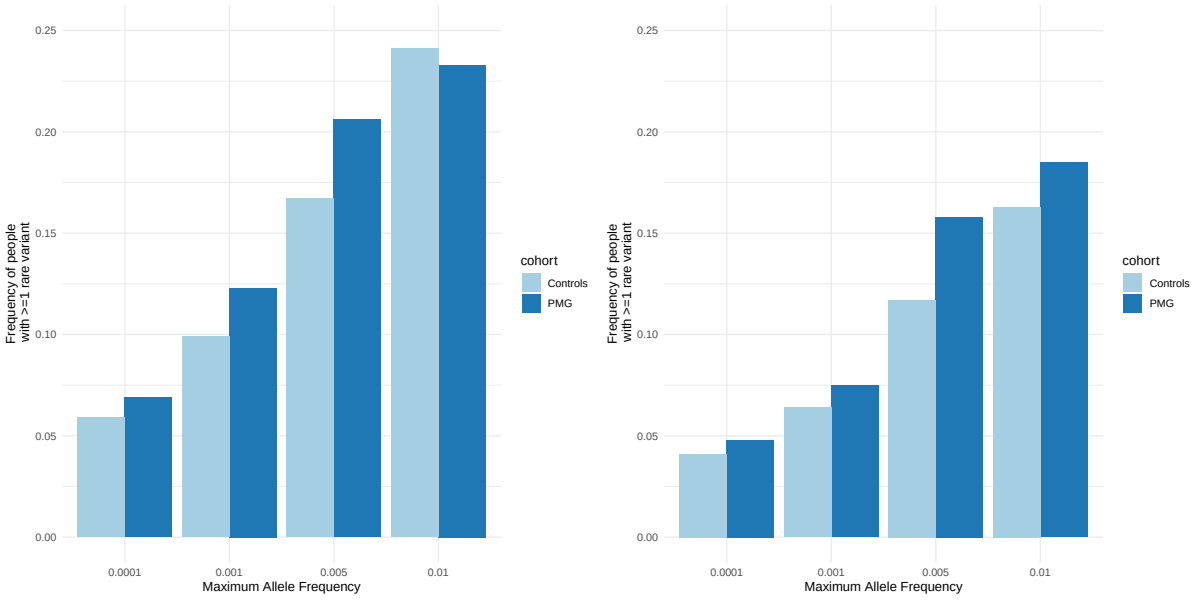
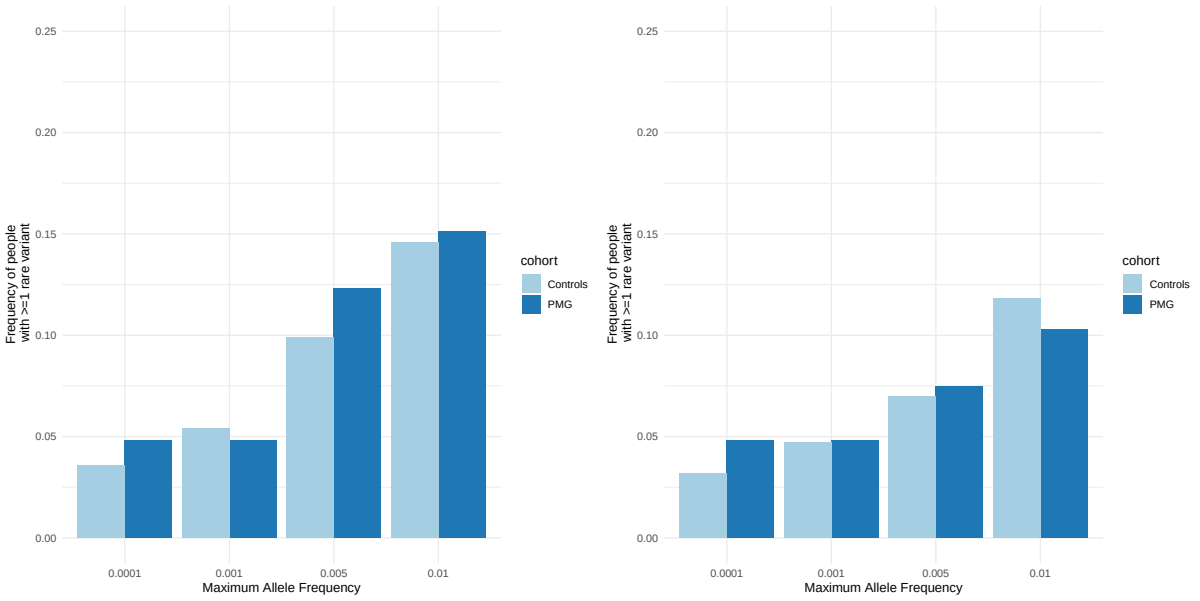


Figure S3: Cumulative burden of rare variants with moderate or high predicted impact in the PMG candidate genes in each of the control cohorts of BR-RD separately (black), together (blue), in PMG (red) and the subphenotypes of PMG based on histopathology (C3GN, DDD, IC-MPGN, PMG unclassified (PMG U)), C3NeF status (positive/negative) and those with low C3 (purple). Horizontal lines indicate the 95% confidence intervals.



(a) CADD=0

(b) CADD $\geq$ 10



(c) CADD $\geq$ 15

(d) CADD $\geq$ 20

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.

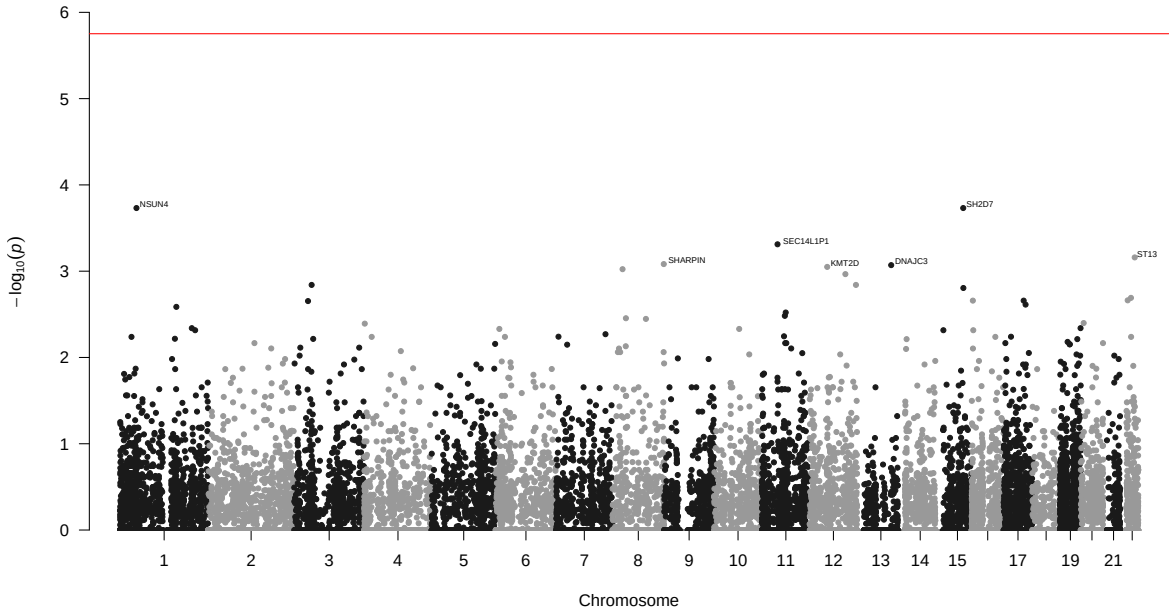


Figure S5: Exome-wide rare variant burden analysis Manhattan plot comparing European unrelated PMG cases and controls. No gene surpasses the exome-wide Bonferroni significance threshold indicated by a horizontal red line ($p < 1.77 \times 10^{-6}$). Genes achieving $p < 1 \times 10^{-3}$ have been annotated.

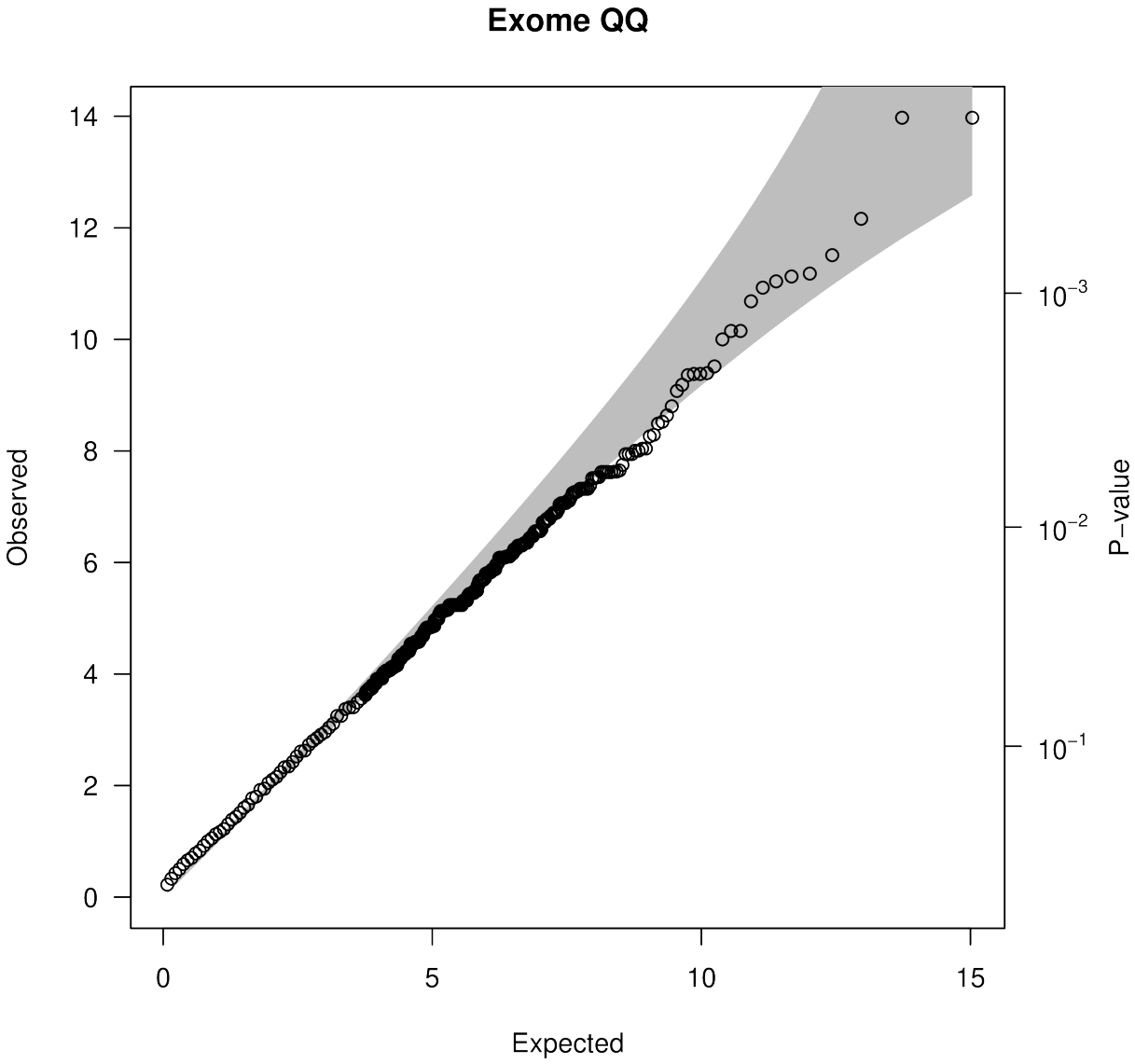


Figure S6: QQ plot for exome-wide rare variant gene burden analysis. The observed/expected chi-square values and corresponding p-values are shown. The grey shaded area indicates the 95% confidence interval of the null.

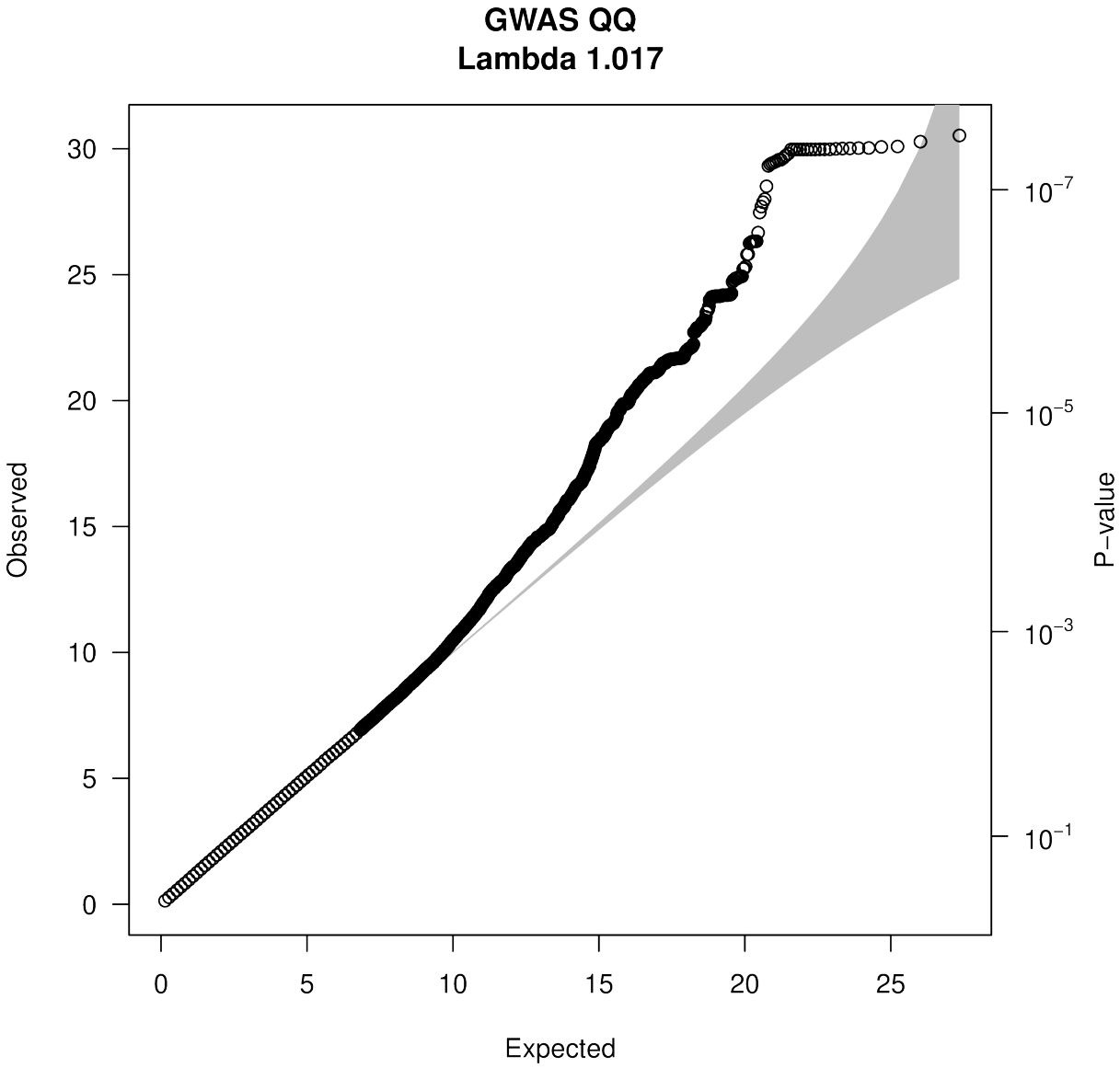


Figure S7: QQ plot for the common variant genome-wide association analysis. The observed/expected chi-square values and corresponding p-values are shown. The grey shaded area indicates the 95% confidence interval of the null. The genomic inflation (lambda) is 1.017.

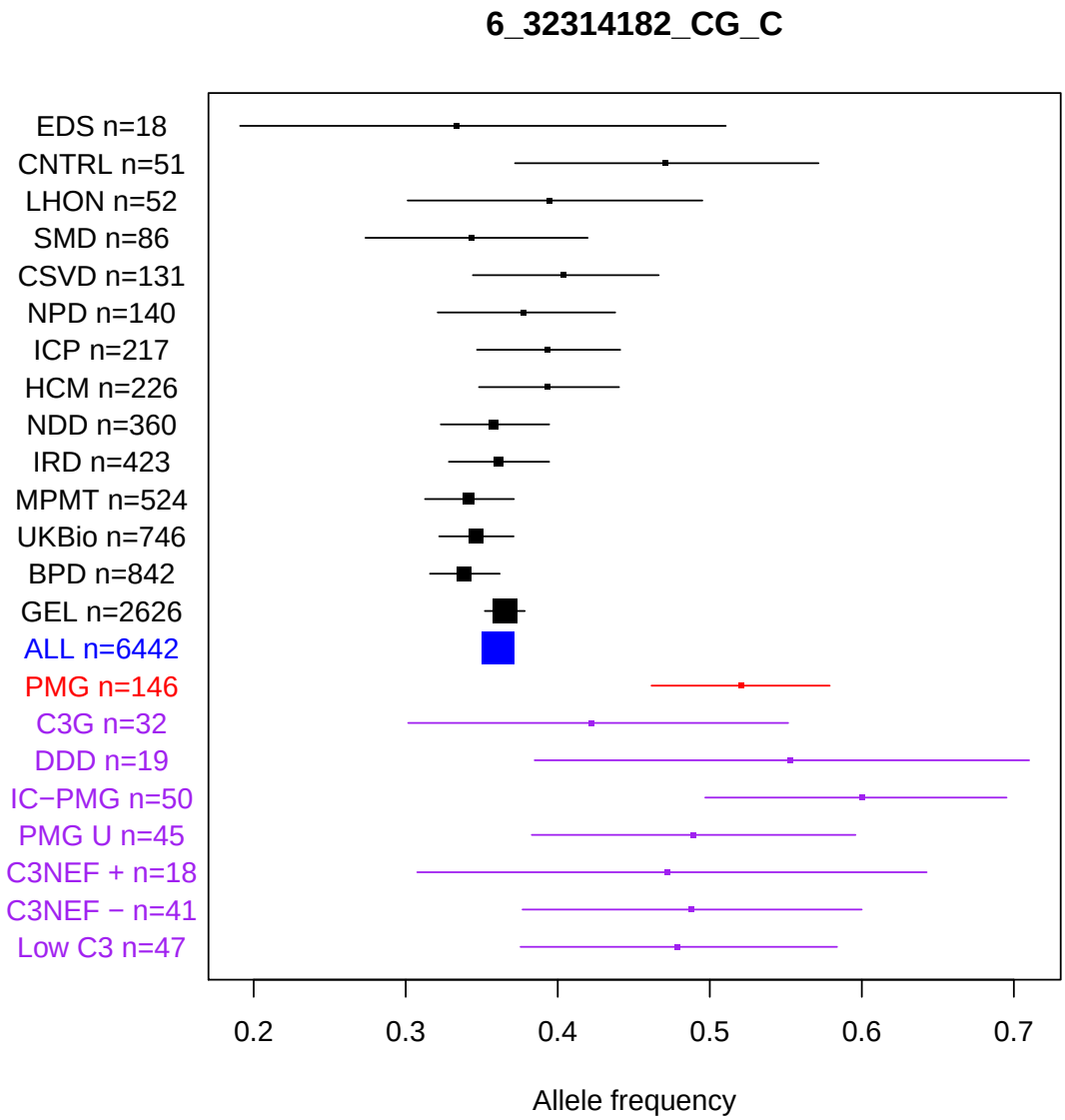


Figure S8: Allele frequency of the lead variant from the chromosome 6 locus in each of the control cohorts of BR-RD separately (black), together (blue), in PMG (red) and the subphenotypes of PMG based on histopathology (C3GN, DDD, IC-MPGN, PMG unclassified (PMG U)), C3NeF status (positive/negative) and those with low C3 (purple). Horizontal lines indicate the 95% confidence intervals.

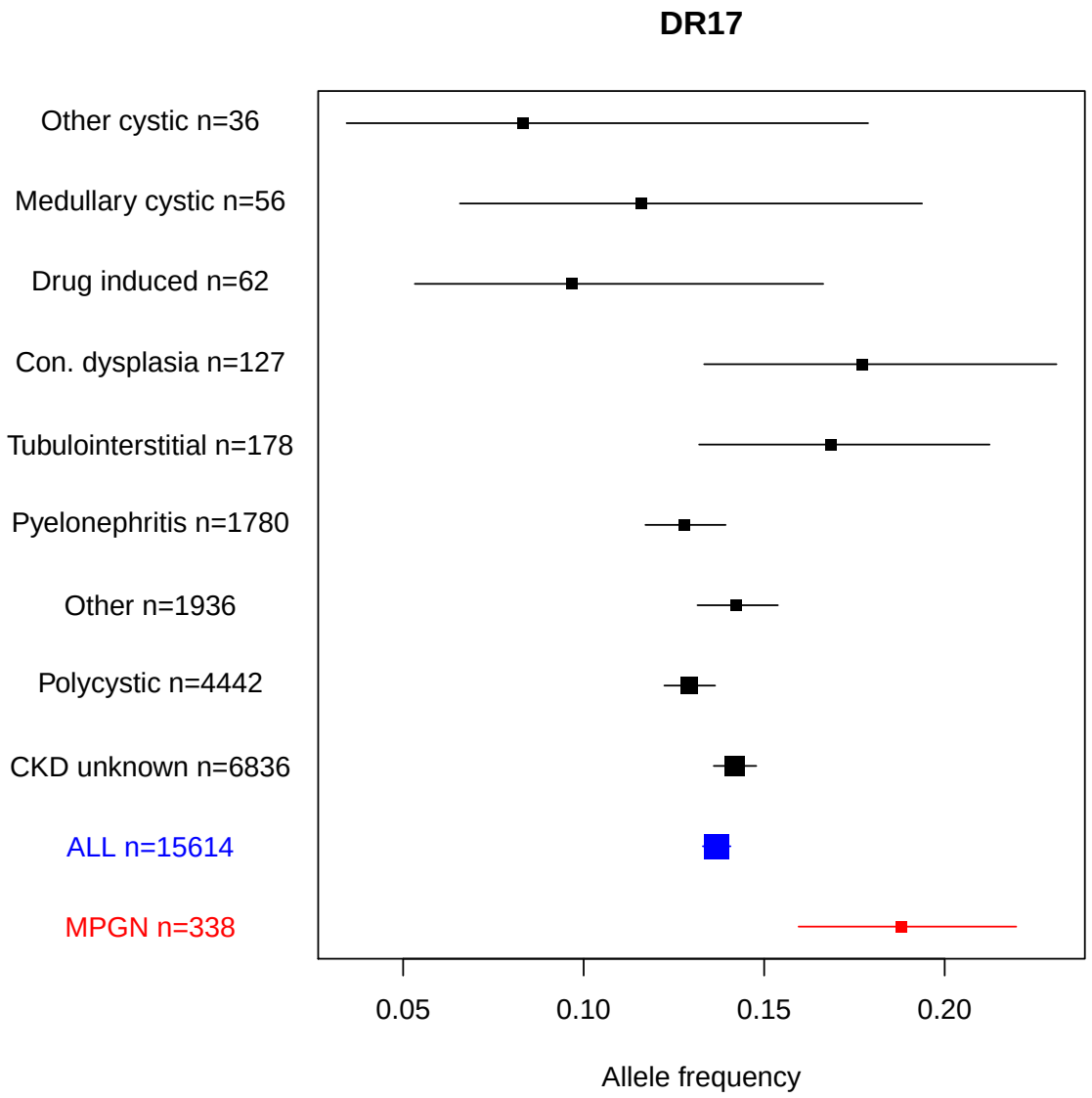


Figure S9: Frequency of HLA serotype DR17 in the NHSBT data showing the controls utilized separately, together (blue) and MPGN (red). Horizontal lines indicate the 95% confidence intervals.

6 Supplementary Tables

See separate Excel file.

Table S1

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

Table S2

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 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

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. **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.