The Journal of Immunology

Statistical Validation of Rare Complement Variants Provides Insights into the Molecular Basis of Atypical Hemolytic Uremic Syndrome and C3 Glomerulopathy

Amy J. Osborne, Matteo Breno, Nicolo Ghiringhelli Borsa, Fengxiao Bu, Véronique Frémeaux-Bacchi, Daniel P. Gale, Lambertus P. van den Heuvel, David Kavanagh, Marina Noris, Sheila Pinto, Pavithra M. Rallapalli, Giuseppe Remuzzi, Santiago Rodríguez de Cordoba, Angela Ruiz, Richard J. H. Smith, Paula Vieira-Martins, Elena Volokhina, Valerie Wilson, Timothy H. J. Goodship and Stephen J. Perkins

J Immunol 2018; 200:2464-2478; Prepublished online 2 March 2018;
doi: 10.4049/jimmunol.1701695
http://www.jimmunol.org/content/200/7/2464

The Journal of Immunology is published twice each month by
The American Association of Immunologists, Inc.,
1451 Rockville Pike, Suite 650, Rockville, MD 20852

Print ISSN: 0022-1767 Online ISSN: 1550-6606.

Statistical Validation of Rare Complement Variants Provides Insights into the Molecular Basis of Atypical Hemolytic Uremic Syndrome and C3 Glomerulopathy

Amy J. Osborne,* Matteo Breno,† Nicolo Ghiringhelli Borsa,‡ Fengxiao Bu,‡,§ Véronique Frémeaux-Bacchi,¶ Daniel P. Gale,‖ Lambertus P. van den Heuvel,#, David Kavanagh,††,‡‡ Marina Noris,† Sheila Pinto,§§ Pavithra M. Rallapalli,* Giuseppe Remuzzi,†,¶¶ Santiago Rodríguez de Cordoba,§§ Angela Ruiz,§§ Richard J. H. Smith,‡ Paula Vieira-Martins,¶ Elena Volokhina,# Valerie Wilson,‖‖ Timothy H. J. Goodship,‡‡ and Stephen J. Perkins***

Atypical hemolytic uremic syndrome (aHUS) and C3 glomerulopathy (C3G) are associated with dysregulation and overactivation of the complement alternative pathway. Typically, gene analysis for aHUS and C3G is undertaken in small patient numbers, yet it is unclear which genes most frequently predispose to aHUS or C3G. Accordingly, we performed a six-center analysis of 610 rare genetic variants in 13 mostly complement genes (*CFH*, *CFI*, *CD46*, *C3*, *CFB*, *CFHR1*, *CFHR3*, *CFHR4*, *CFHR5*, *CFP*, *PLG*, *DGKE*, and *THBD*) from >3500 patients with aHUS and C3G. We report 371 novel rare variants (RVs) for aHUS and 82 for C3G. Our new interactive Database of Complement Gene Variants was used to extract allele frequency data for these 13 genes using the Exome Aggregation Consortium server as the reference genome. For aHUS, significantly more protein-altering rare variation was found in five genes *CFH*, *CFI*, *CD46*, *C3*, and *DGKE* than in the Exome Aggregation Consortium (allele frequency < 0.01%), thus correlating these with aHUS. For C3G, an association was only found for RVs in *C3* and the N-terminal C3b-binding or C-terminal nonsurface-associated regions of *CFH*. In conclusion, the RV analyses showed nonrandom distributions over the affected proteins, and different distributions were observed between aHUS and C3G that clarify their phenotypes. ***The Journal of Immunology*****, 2018, 200: 2464–2478.**

Atypical hemolytic uremic syndrome (aHUS) and C3 glomerulopathy (C3G) are two severe ultrarare renal diseases that involve dysregulation of the alternative pathway (AP) in the complement system of innate immunity. In healthy individuals, the AP eliminates unwanted pathogens without compromising host cells due to the balance between AP activator and regulatory proteins. aHUS and C3G feature host cell attack by the AP, leading to end-stage renal failure. However, aHUS is a thrombotic microangiopathy with an acute presentation, whereas C3G is not a thrombotic microangiopathy but is characterized by an

*Department of Structural and Molecular Biology, University College London, London WC1E 6BT, United Kingdom; †Centro di Ricerche Cliniche per le Malattie Rare "Aldo e Cele Daccò," IRCCS-Istituto di Ricerche Farmacologiche "Mario Negri," 24020 Ranica Bergamo, Italy; ‡Molecular Otolaryngology and Renal Research Laboratories, Carver College of Medicine, University of Iowa, Iowa City, IA 52242; §Medical Genetics Center, Southwest Hospital, Chongqing 400038, China; ¶Assistance Publique-Hôpitaux de Paris, Hôpital Européen Georges Pompidou, Service d'Immunologie Biologique, 75015 Paris, France; ‖Centre for Nephrology, Royal Free Hospital, University College London, London NW3 2QG, United Kingdom; #Department of Pediatric Nephrology, Radboud University Medical Center, 6525 GA Nijmegen, the Netherlands; **Department of Pediatric Nephrology, Department of Growth and Regeneration, University Hospital Leuven, 3000 Leuven, Belgium; ††The National Renal Complement Therapeutics Centre, Newcastle upon Tyne NE1 4LP, United Kingdom; ‡‡Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne NE1 3BZ, United Kingdom; §§Department of Cellular and Molecular Medicine, Center for Biological Research and Center for Biomedical Network Research on Rare Diseases, 28040 Madrid, Spain; ¶¶Department of Biomedical and Clinical Sciences, University of Milan, 20122 Milan, Italy; and ‖‖Northern Molecular Genetics Service, Newcastle upon Tyne Hospitals National Health Service Foundation Trust, Newcastle upon Tyne NE1 3BZ, United Kingdom

ORCIDs: 0000-0001-6869-1176 (A.J.O.); 0000-0002-8397-637X (N.G.B.); 0000-0002-4865-8528 (V.F.-B.); 0000-0002-9170-1579 (D.P.G.); 0000-0003-4718-0072 (D.K.); 0000-0001-7651-5033 (M.N.); 0000-0002-8261-6209 (P.M.R.); 0000-0002-6194-3446 (G.R.); 0000-0001-6401-1874 (S.R.d.C.); 0000-0003-3063-9414 (P.V.-M.); 0000-0001-9218-9805 (S.J.P.).

Received for publication December 6, 2017. Accepted for publication January 31, 2018.

This work was supported by the Spanish Ministerio de Economía y Competitividad/Fondo Europeo de Desarrollo Regional (Grant SAF2015-66287-R to S.R.d.C.), the Seventh Framework Programme European Union Project (FP7/2007–2013) (EURenOmics) (Grant 305608 to S.R.d.C.), the Autonomous Region of Madrid (Grant S2017/BMD-3673 COMPLEMENTO II-CM to S.R.d.C.), the National Institutes of Health (Grant R01 DK110023 to R.J.H.S.), the Dutch Kidney Foundation (CP 14.27 COMBAT Consortium, Grants 13OI116, KFB 11.007, IP 10.22, and 160KKO1 to L.P.v.d.H. and E.V.), the European Renal Association – European Dialysis and Transplantation Association (Grants ERA STF 138-2013 and ERA LTF 203-2014 to L.P.v.d.H. and E.V.), and the European Society for Pediatric Nephrology (Grant 2014.03 to L.P.v.d.H. and E.V.). A.J.O. and S.J.P. received a Ph.D. studentship grant from Alexion to Complement U.K.

Initial results from this work were presented as an abstract and a poster at the 26th International Complement Workshop, September 5–8, 2016, Kanazawa, Japan.

Address correspondence and reprint requests to Prof. Stephen J. Perkins, Department of Structural and Molecular Biology, Darwin Building, University College London, Gower Street, London WC1E 6BT, U.K. E-mail address: s.perkins@ucl.ac.uk

The online version of this article contains supplemental material.

Abbreviations used in this article: AF, allele frequency; aHUS, atypical hemolytic uremic syndrome; AP, alternative pathway; *CD46*, cluster of differentiation 46; *CFB*, complement factor B; *CFH*, complement factor H; *CFHR*, complement factor H–related; *CFI*, complement factor I; *CFP*, complement factor properdin; C3G, C3 glomerulopathy; CNV, copy number variation; *DGKE*/DGKE, diacylglycerol kinase epsilon; EVS, Exome Variant Server; ExAC, Exome Aggregation Consortium; FB, factor B; FH, complement factor H; FHR, complement factor H related; FI, complement factor I; 1000GP, 1000 Genomes Project; LGR, large genomic rearrangement; MAF, minor AF; MCP, membrane cofactor protein; MG, macroglobulin; PDB, Protein Data Bank; *PLG*/PLG, plasminogen; RV, rare variant; SCR, short complement repeat; *THBD*/THBD, thrombomodulin.

abundance of C3 deposition in the renal glomeruli with a mostly chronic presentation, with some exceptions (1).

Rare genetic abnormalities in AP and thrombosis-related genes are present in ~60% of aHUS cases (2–4). These mostly drive AP dysregulation at the endothelial cell surface (1). The penetrance of predisposing rare variants (RVs) in aHUS is ~50% and is determined by the *complement factor H* (*CFH*) and membrane cofactor protein (MCP; cluster of differentiation 46 [*CD46*]) haplotype, additional RVs, and a trigger (5–8). As reported in our 2014 aHUS database (9), most genetic aHUS cases are heterozygous (10) and are attributed to the genes *CFH* (25–30% of cases), followed by *CD46* (8–10%), *C3* and *complement factor I* (*CFI*) (4–8% each), and *complement factor B* (*CFB*) (1–4%) (6, 11). The C-terminal short complement repeat (SCR)-19/20 domains of FH are a well-known RV hotspot for aHUS, this being attributed to its functional binding sites for C3b, C3d, and heparin (12). (We define SCR thus, rather than "short consensus repeat" as a better descriptor of the most abundant domain type in complement). As opposed to other genes, most RVs in *CD46* lead to a quantitative decrease in the protein product, and approximately one quarter are homozygous (10). Rare copy number variation (CNV) leading to large genomic rearrangements (LGRs) in the *complement factor H-related* (*CFHR*) region, such as *CFH/CFHR1* and *CFH/CFHR3* hybrid genes, are risk factors for aHUS (13–16). RVs in the noncomplement gene *thrombomodulin* (*THBD*) account for 3–4% of genetic aHUS (6), although no *THBD* variants were detected in the French aHUS cohort (17). RVs in the noncomplement gene *diacylglycerol kinase epsilon* (*DGKE*) account for ~27% of cases of aHUS presenting under the age of 1 y and for <4% of cases presenting under the age of 2 y (18, 19).

Complement gene RVs have been identified in ~20% of sporadic C3G cases (6, 20–22). Familial C3G is most often linked to highly penetrant heterozygous CNV in *CFHR1–5* genes, such as *CFHR5* nephropathy, as well as homozygous *CFH* deficiency and heterozygous gain-of-function mutation in *C3* (23–30). These frequently affect AP regulation in the fluid phase, with some exceptions (12). As in aHUS, unaffected carriers of these genetic abnormalities are seen, indicating that the genetic variant only predisposes for the manifestation of C3G (31). aHUS and C3G also involve anti-CFH autoantibodies, known as acquired factors (5–13% of cases) (6, 32).

Genetic sequencing and multiple ligation-dependent probe assessment screening panels for aHUS and C3G typically include up to 10 complement genes (*CFH*, *CFHR1*, *CFHR3*, *CFHR4*, *CFHR5*, *CFI*, *C3*, *CD46*, *CFB*, and *complement factor properdin* [*CFP*]) (33)), two coagulation genes (*THBD* and *plasminogen* [*PLG*]) (34), and one noncomplement gene (*DGKE*) (1). RVs in six genes (*CFH*, *CFI*, *C3*, *CD46*, *CFB*, and *DGKE*) are associated with aHUS, whereas RVs in *CFH* and *C3* associate with C3G. These associations result from studies of many aHUS patients and rather fewer C3G patients by linkage and familial segregation, case-control cohorts, and functional studies. However, the associations of RVs in the four genes *CFHR5*, *CFP*, *PLG*, and *THBD* with aHUS, and in all of the 13 genes with C3G, are less well defined (1). In addition, an increasing number of variants of unknown significance for aHUS and C3G are being identified among these 13 genes. These require rapid pathogenicity evaluations for clinical interpretation (e.g., almost one third of *CFH* variants have limited functional characterization) (35).

To clarify differences in the genetic and molecular basis of aHUS and C3G, we have analyzed the statistical allele frequencies (AFs) of RVs in multiple patient cohorts for comparison with reference datasets, and we have developed a new Database of Complement Gene Variants (http://www.complement-db.org). The AF provides the frequency of the variant in a given population. Variant pathogenicity can be identified by comparisons with genomic reference datasets, such as the Exome Aggregation Consortium (ExAC) (36). The database also provides structural biology and evolutionary tools to predict the effect of RVs on these proteins and present functional data from the literature and ClinVar. The database was used to analyze 610 rare genetic variants from >3500 patients in six renal centers; this is the largest dataset known to date for aHUS and C3G, with 371 aHUS and 82 C3G novel RVs. Following comparisons with 60,706 genomic reference sequences from ExAC (37) and 6,500 from the Exome Variant Server (EVS), we confirmed the associations of six genes (above) with aHUS and that three genes (*CFHR5*, *PLG*, and *THBD*) were not associated with aHUS. The statistical comparisons also confirmed that *CFH* and *C3* were associated with C3G and suggested the involvement of *CFB* and *THBD* with C3G. Our results explain how changes in the same proteins result in the different pathologies observed with aHUS and C3G. Through the use of AFs and burden testing in our database, the new database will inform patient management by enabling clinical immunologists to interpret new variants in terms of their associations with aHUS and C3G.

Materials and Methods

Data collection

The aHUS and C3G phenotype and variant data were sourced from six centers in this study as follows, including from literature searches (8, 9, 17). aHUS was diagnosed by the presence of one or more episodes of microangiopathic hemolytic anemia and thrombocytopenia defined on the basis of hematocrit < 30%, hemoglobin < 10 mg/dl, serum lactate dehydrogenase > 460 U/l, undetectable haptoglobin, fragmented erythrocytes in the peripheral blood smear, and platelet count < 150,000/μl, associated with acute renal failure, together with a negative Coombs test, ADAMTS13 activity > 10%, and negative Shiga toxin (1). C3G was diagnosed by the presence of C3 deposits by immunofluorescence in the absence or comparatively reduced presence of Igs. The identification of dense deposits within the glomerular basement membrane by electron microscopy led to the further classification of the C3G as dense deposit disease (1). The database included variant data for 13 genes (*C3*, *CD46*, *CFB*, *CFH*, *CFHR1*, *CFHR3*, *CFHR4*, *CFHR5*, *CFI*, *CFP*, *DGKE*, *PLG*, and *THBD*) (Table I). The estimated variant AFs in the aHUS and C3G datasets were based on data from the six centers only. Data from the three reference genome projects, namely ExAC (Version 0.3) (37), EVS (NHLBI GO Exome Sequencing Project, Seattle, WA (http://evs.gs.washington.edu/EVS), and the 1000 Genomes Project (1000GP) (38), were downloaded and used as surrogate control reference datasets for all genes with the exception of *CFHR1*, *CFHR3*, and *CFHR4*, which were involved in CNVs only. These latter three genes were excluded because, at the time of writing, no CNV data were available from ExAC, EVS, or 1000GP. These three reference datasets do not contain genomes from patients with rare renal disease phenotypes that include aHUS or C3G; however, ExAC and EVS contain genomes from myocardial infarction studies. Although this may affect the analyses of thrombosis-related genes, such as *THBD*, a recent study found that ExAC was not enriched in pathogenic RVs for these diseases (39). In ExAC, only variants with "PASS" filter status were included in our analyses. The corresponding AF of each variant in ExAC, EVS, and 1000GP was queried using Human Genome Variation Society nucleotide level nomenclature (c.) and the National Center for Biotechnology Information gene accession number. Published experimental data on each variant were sourced from literature searches using PubMed. Variants were described using DNA and protein level terms for Human Genome Variation Society and legacy nomenclature.

Data cleansing, duplications, and maintenance

Patient duplicate tests were carried out within and between datasets from the six renal centers, first by identifying potential duplicates using the patient's variant profile, disease, gender, and year of birth. Potential duplicates were investigated by requesting the full date of birth details from each renal center and deleting as necessary. For maintenance of the database in the long-term, each collaborating group will provide an annual data update as a spreadsheet that will be automatically uploaded to the database with automated duplication and error checks that are programmed into the

Table I. Summary of mRNA and protein identifiers for the 13 genes

Gene	Protein	RefSeq (National Center for Biotechnology Information)		Ensembl	
		mRNA	Protein	Transcript ID	Protein ID
CFH	FH	NM_000186.3	NP_000177.2	ENST00000367429	ENSP00000356399
CFI	FI	NM_000204.3	NP_000195.2	ENST00000394634	ENSP00000378130
C3	C3	NM_000064.3	NP_000055.2	ENST00000245907	ENSP00000245907
CD46	MCP	NM_002389.4	NP_002380.3	ENST00000358170	ENSP00000350893
CFB	FB	NM_001710.5	NP_001701.2	ENST00000425368	ENSP00000416561
CFHR1	FHR1	NM_002113.2	NP_002104.2	ENST00000320493	ENSP00000314299
CFHR3	FHR3	NM_021023.5	NP_066303.2	ENST00000367425	ENSP00000356395
CFHR4	FHR4	NM_001201550.2	NP_001188479.1	ENST00000367416	ENSP00000356386
CFHR5	FHR5	NM_030787.3	NP_110414.1	ENST00000256785	ENSP00000256785
THBD	THBD	NM_000361.2	NP_000352.1	ENST00000377103	ENSP00000366307
CFP	FP	NM_001145252.1	NP_001138724.1	ENST00000396992	ENSP00000380189
DGKE	DGKE	NM_003647.2	NP_003638.1	ENST00000284061	ENSP00000284061
PLG	Plasminogen	NM_000301	NP_000292	ENST00000308192	ENSP00000308938

The databases of annotated genomic, transcript, and protein reference sequences can be found at https://www.ncbi.nlm.nih.gov/refseq/ (RefSeq) and http://www.ensembl.org/index.html (Ensembl).

MySQL backend. As an open resource, other clinical centers are invited to upload formatted data updates through the curator. To prevent against SQL injection attacks and related software vulnerabilities, the database inputs were sanitized and tested using the penetration testing tool SQLMAP.

Web database development

The Database of Complement Gene Variants (http://www.complement-db.org) was constructed using a MySQL platform, and its user interface was developed using a combination of PHP, JavaScript, jQuery, CSS, and HTML. The design is based on our 2014 aHUS database and the European Association for Haemophilia and Allied Disorders Coagulation Factor IX Web database (http://www.factorix.org/) (9, 40). The database is mounted on a Linux server within the Information Services Division at University College London.

Data retrieval

Variant data were retrieved from the Database of Complement Gene Variants using simple or advanced search tools. The user first defined the data source and gene and then the other search options, including protein domain, location, and variant type and effect. The advanced search tool featured a customizable default ExAC AF cutoff value of 1%, which was used to ensure that any variant with an ExAC AF >1% was not retrieved. Variants were mapped to their genomic location using the reference human genomes GRCh37 and GRCh38. The protein and transcript locations were identified using RefSeq and Ensembl (Table I). The database provided protein structural views of each missense variant using the JSmol Java applet (https://sourceforge.net/projects/jsmol/). These views facilitated a structural understanding of the variant using the Web page "structure and function" tab. For each disease dataset, the statistics page showed the distribution of variants in genes by protein domain, exon/intron location, and variant type and effect; the RV burden per gene; and the number of RVs per case. The AF Web page summarized the number of variants in each gene using their reference genome AF, with links to their variant database entries, for each disease dataset. The world map Web page showed the number of laboratory-sourced cases that have citizenship in each country, for which data were available. The new variants Web page predicted the effects of any theoretical missense variants using AF, structural, and functional analyses. The structures Web page listed all known protein structural models, with links to their Protein Data Bank (PDB) code. The amino acid alignments Web page showed the multiple sequence alignment of up to 10 vertebrate species for each protein, depending on sequence availability; this was calculated using the UniProt database and the Probabilistic Alignment Kit program (41).

RV burden

The RV burden per gene was computed for the aHUS, C3G, ExAC, and EVS datasets. The RV burden was defined as the proportion of screened cases with an identified RV per gene. For aHUS and C3G, the burden was calculated by dividing the number of cases with an identified RV by the total number of cases screened per gene. For ExAC and EVS, this was calculated for each gene by dividing the total mean adjusted allele count, which was determined from the allele count minus the number of homozygous cases, by the total mean adjusted number of subjects with RVs (42). An AF cutoff of 0.01% was used for all RV burden calculations, which is advised for a Mendelian disease (43). The RV burden was calculated for all protein-altering RV only; truncating (non-sense, frameshift, and splice) and non-truncating (missense and in-frame). RVs classified as "benign" and "likely benign" (see next section) were not filtered out, because these data were unavailable for the ExAC and EVS datasets.

RV assessment

Each non-LGR RV was classified as "pathogenic," "likely pathogenic," "uncertain significance," "benign," or "likely benign" using categorization guidelines (1) that followed the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (44). The following categories were also added to cover all RVs in our datasets:

1) Minor AF (MAF) $< 0.1\%$; likely benign by clinical testing (Illumina; ClinVar); in silico analyses predict: tolerated, neutral, and benign. Categorized as "likely benign" for all genes with the exception of *C3* and *CFB*, which were "uncertain significance" instead.
2) MAF $< 0.1\%$; likely benign by clinical testing (Illumina; ClinVar); in silico analyses predict: uncertain deleterious effects. Categorized as "uncertain significance."
3) MAF $< 0.1\%$, predicted as loss-of-function (includes non-sense, frameshift, and splice acceptor/donor variants) in *C3* or *CFB*. Categorized as "uncertain significance." *C3* or *CFB* loss-of-function is unlikely to overactivate complement and lead to aHUS or C3G. Because their effects on complement have not been experimentally proven, it was safer to categorize these as "uncertain significance."
4) MAF $< 0.1\%$; synonymous change; in silico analyses predict: tolerated, neutral, and benign. Categorized as "likely benign."
5) $0.1\% <$ MAF $< 1\%$; no functional data; in *C3* or *CFB*. Categorized as "likely benign."

Loss-of-function RVs in all genes, with the exception of those involved only in LGR variants or in the complement activators *C3* and *CFB*, were classified as "likely pathogenic" for aHUS and C3G, unless functional data reported otherwise (1). The Ensembl Variant Effect Predictor tool (45) was used for PolyPhen-2 (46) and SIFT (47) analyses, and the results were used for combinatorial variant analyses. These classifications were made available for each variant identified in the aHUS and C3G datasets.

Spatial distribution of missense RVs in the proteins

For each protein domain, the AFs of missense RVs in the aHUS and C3G datasets were summed and then divided by the proportion of protein residues in the corresponding domain, to identify mutational hotspots. Missense RVs that were categorized as "benign" or "likely benign" were excluded from these analyses. For FH, complement factor I (FI), C3, MCP, and complement factor B (FB), where there were enough missense RVs to identify mutational hotspots, these were represented graphically as bar charts. For FH, FI, C3, MCP, and FB, each unique missense RV was mapped onto the structural model for visualization, noting that these did not take the aHUS or C3G AF into account. The structural models for FH [PDB code 3N0J (48)], C3 [PDB code 2A73 (49)], FI [PDB code 2XRC (50)], MCP [PDB code 3O8E (51)], and FB [PDB code 2OK5 (52)] were

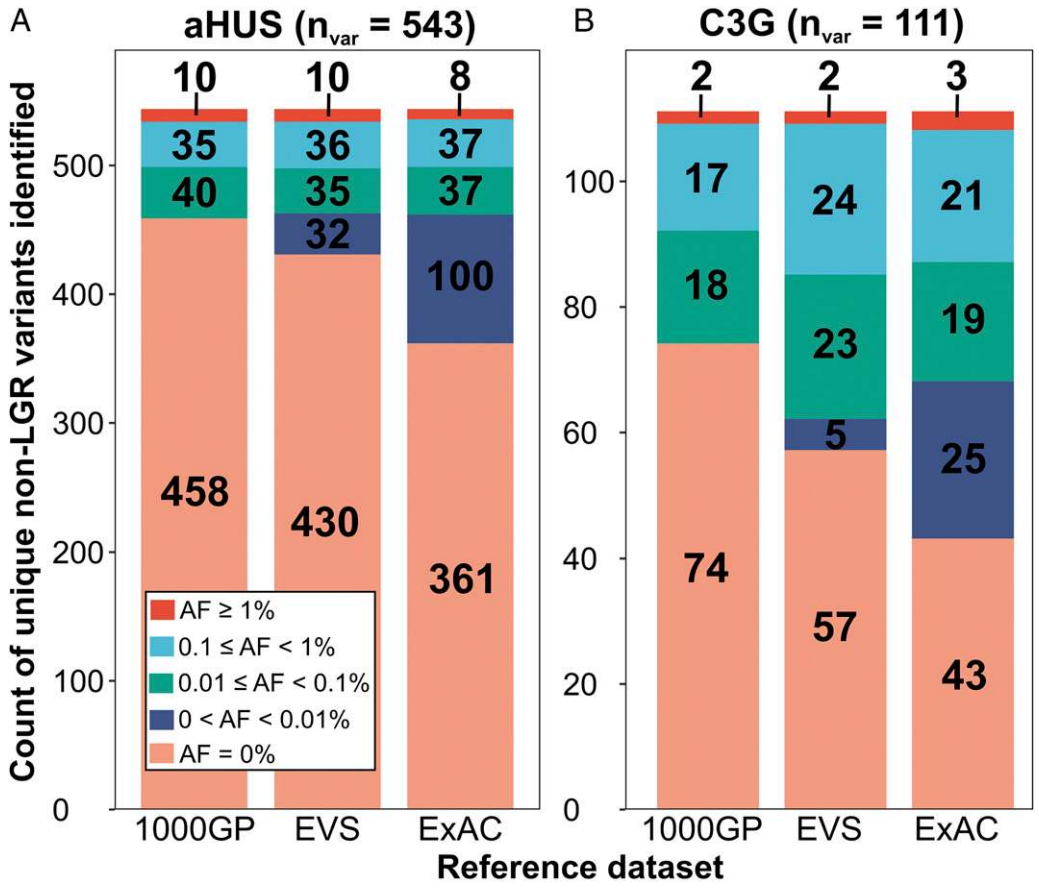

FIGURE 1. Stacked bar analyses showing the reference AF of the variants identified in the aHUS (**A**) and C3G (**B**) datasets. The total numbers of unique variants (n_{var}) are shown within the bars. A unique variant is defined as, when a variant is seen in more than one patient, it is only counted once. Unique variants (excluding the 13 aHUS and 3 C3G CNVs) are categorized by their AF in each of the three reference datasets: 1000 GP, with an allele number (AN; total number of alleles screened for each gene) of 3,775–5,008, EVS, with an AN of 8,202–13,005, and ExAC, with an AN of 14,708–121,412. The aHUS and C3G datasets had ANs of 634–6,256 and 208–886, respectively.

sourced from the PDB. An alternative model for FH, in which the C-terminal domains were extended and not the N-terminal domains, was also used (data not shown; PDB code 3GAV) (9). The R statistical package was used for statistical analyses and artwork (http://www.R-project.org/). PyMol was used for protein structural visualization and artwork (53).

Statistical analyses

All categorical variables were examined using the two-tailed Fisher exact test or two-tailed χ^2 test (most with the Yates correction) with a 0.05 significance level that was Bonferroni corrected where applicable. No statistics were used for Table I and Supplemental Table I. For the analyses on genetic variants in aHUS and C3G in Fig. 1, the two-tailed Fisher exact test was used with a significance level of 0.05. For the common variants analyses in Table II, the χ^2 test with the Yates correction was used and the 0.05 significance level was Bonferroni corrected by dividing by the 14 variants, to give 0.0036. For the RV frequencies in aHUS and C3G cases analyses in Fig. 2, the two-tailed Fisher exact test was used with a significance level of 0.05. For the RV profiles in aHUS and C3G cases analyses in Fig. 2 and Table III, the two-tailed Fisher exact test was used with a significance level of 0.05. For the RV abundance in genes analyses in Fig. 3, the two-tailed Fisher exact test was used with a significance of 0.05. No statistics were used for Table IV. For the patient gender analyses in Table V, the χ^2 test was used with a significance level of 0.05. For Table VI, the χ^2 test with Yates correction was used and the 0.05 significance level was Bonferroni corrected by dividing by 12 genes to give 0.0042 for aHUS and by 11 genes to give 0.0045 for C3G. The "ALL" genes category for aHUS and C3G were each subjected to a 0.05 significance level (no Bonferroni adjustment needed). For the RV burden analyses presented in Fig. 4 and Supplemental Tables II and III, the χ^2 test with Yates correction was used and the 0.05 significance level was Bonferroni corrected by dividing by the nine genes to give 0.0056. For the statistical analyses of *CFH* in Supplemental Table IV and Fig. 5, the χ^2 test was used with a 0.05 significance level.

For each row of one independent *t* test (protein-altering only) in Supplemental Tables II and III, we undertook a power calculation using the program PS Power and Sample Size Calculations (Version 3.0) (54). Our calculations used expected differences ≥ 5% for *CFH*, *C3*, and *CD46*, ≥2.6% for *DGKE* in the aHUS and EVS groups, ≥2.5% for *DGKE* in the other three groups, and ≥2% for all other genes and took into account the unequal sizes of the experimental (aHUS or C3G) and control (ExAC or EVS) groups. If the true difference between the experimental and control groups was as expected, we were able to reject the null hypothesis that their frequencies were equal with a certain probability (power). The type I error probability associated with all of these tests was 0.0056. The power was ≥80% in all tests, with the exception of *DGKE* (74%) for aHUS, and *DGKE* (55%) and *PLG* (32%) for C3G, when the ExAC reference dataset was used (Supplemental Table II), as well as *DGKE* (67%) for aHUS, and *DGKE* (68%), *PLG* (10%), *THBD* (38%), and *CFB* (41%) for C3G when using EVS as the reference dataset (Supplemental Table III).

AF analyses

A two-tailed χ^2 test with the Yates correction was used to assess the difference in protein-altering RV AF (ExAC AF < 0.1%) between the aHUS/C3G and reference datasets. Protein-altering variants included non-truncating and truncating variants only (42). For each RV, the 0.05 significance level was Bonferroni adjusted by dividing by the number of RVs identified in its gene for aHUS and ExAC (*C3*: 485, *CD46*: 191, *CFB*: 400, *CFH*: 469, *CFHR5*: 262, *CFI*: 215, *DGKE*: 193, *PLG*: 263, *THBD*: 145) and for C3G and ExAC (*C3*: 452, *CD46*: 130, *CFB*: 391, *CFH*: 334, *CFHR5*: 261, *CFI*: 173, *DGKE*: 176, *PLG*: 263, *THBD*: 143) and rounding to four decimal places to give significance levels of 0.0001, 0.0002, 0.0003, or 0.0004 accordingly. AF was not available in the aHUS dataset for the two *CFP* RVs, so *CFP* was not analyzed. *CFHR1*, *CFHR3*, and *CFHR4* were not analyzed because LGRs were not included in ExAC at the time of writing. Related individuals were taken out of the analyses by including familial alleles only once (twice for homozygous cases).

Table II. The 14 common genetic variants identified in at least one of the three reference datasets (1000GP, EVS, and ExAC) at AF ≥ 1%

Gene	Genetic Variant	Protein Variant	Dataset AF (%)				
			1000GP	EVS	ExAC	aHUS	C3G
C3	c.1407G>C	p.Glu469Asp	1.62[a]	1.41[a]	0.40	0.16	0
CFB	c.754G>A	p.Gly252Ser	1.02[a]	2.82[a]	2.22[a]	0.43	0
CFB	c.1598A>G	p.Lys533Arg	1.92[a,b]	0.44	1.05[a]	0.20	0.26
CFB	c.1697A>C	p.Glu566Ala	1.02	0.73	1.12[a]	0.53	0.13
CFB	c.1953T>G	p.Asp651Glu	1.04[a]	0.94[a]	0.22	0.04	0
CFH	c.1652T>C	p.Ile551Thr	1.91[a]	1.68[a]	0.50[a]	0.16	0
CFH	c.2669G>T	p.Ser890Ile	6.23[a]	6.57[a]	1.99[a]	0.34	0
CFH	c.2808G>T	p.Glu936Asp	20.33[a]	13.80[a]	19.55[a]	0.02	0
CFHR5	c.136C>T	p.Pro46Ser	0.90	1.13	0.70	0.63	0
CFHR5	c.1067G>A	p.Arg356His	1.04	2.09[a]	1.78[a]	1.42	0.12
CFI	c.884-7T>C	—	2.56[a]	0	2.31[a]	0.02	0
CFI	c.1534+5G>T	—	0.30	1.13	0.90	0.33	0.12
CD46	c.1058C>T	p.Ala353Val	0.40	1.25[a]	1.53[a]	0.60	0.24
PLG	c.1567C>T	p.Arg523Trp	0.24	1.01	0.68	0.35	0

[a]Significantly more common in ExAC than in aHUS using a two-tailed χ^2 test with Yates correction and a Bonferroni corrected significance level of 0.0036.
[b]Significantly more common in ExAC than in C3G using a two-tailed χ^2 test with Yates correction and a Bonferroni corrected significance level of 0.0036.

This procedure refers to the AF assessments that were displayed on the Web database and were also used to verify that none of the RVs was significantly more common in ExAC than in aHUS or C3G.

Results

Genetic variants in aHUS and C3G

To perform comparisons between aHUS and C3G using AF analyses, six reference centers from the United Kingdom, France, Italy, Spain, Holland, and the United States provided phenotype and variant data sets for 3128 and 443 patients from national and international registries. The total of disease variants was 543 for aHUS and 111 for C3G in 13 genes (Fig. 1). The aHUS and C3G variants were analyzed in terms of their reference AFs for these variants found in three reference datasets (1000GP, EVS, and ExAC). First, 10, 10, and 8 common variants with AF > 1% were filtered out (red, Fig. 1). The RVs that were retained were those with AF < 1% in the reference datasets (light blue, green, dark blue, and pink, Fig. 1), or were rare LGRs not covered by the reference datasets. By this filtering, 610 retained RVs (96–97% of the total) were identified in 13 genes (Table I) for aHUS (542 variants) and C3G (110 variants), with 42 shared by aHUS and C3G, including the LGRs. The remaining 3–4% of common variants with AF > 1% in at least one of the reference datasets consisted of 14 and 4 variants in aHUS and C3G, respectively (Table II). These 14 and 4 variants were just slightly over the 1% AF threshold, with the exception of three (*CFH* p.Ser890Ile, *CFH* p.Glu936Asp, and *CFI* c.884-7T>C). In the aHUS dataset, 14 (75), 19 (103), and 32% (174) of variants were found to have an AF between 0 and 1% (light blue/green/dark blue, Fig. 1A) in the 1000GP, EVS, and ExAC reference datasets, respectively. In the C3G dataset, these proportions were significantly higher: 32 (35), 47 (52), and 59% (65), respectively (light blue/green/dark blue, Fig. 1B). In confirmation of this outcome, all three analyses gave $p < 0.0001$ using the two-tailed Fisher exact test. This outcome was still true when the benign and likely benign RVs were excluded from our aHUS and C3G datasets ($p < 0.0001$, Fisher exact test) (see below). Our aHUS and C3G RV datasets were compared with the literature (9) to determine how many RVs were novel. Of the 542 aHUS and 110 C3G RVs (purple and yellow, Fig. 2A), 68 (371) and 75% (82), respectively, were not found in the literature and, therefore, were novel (purple, Fig. 2A). The RVs that were reported in the literature but not in the six reference centers were not part of our current analyses, because we could not calculate their aHUS and C3G AF values (green, Fig. 2A); however, they are included in the updated database from this study.

RV frequencies in cases

The frequencies of the RVs in the aHUS and C3G datasets were surveyed. First, we investigated differences in the screened cases with RVs for aHUS and C3G. The aHUS and C3G datasets contained 3128 and 443 patients, respectively (Fig. 2B; Table III). Of these totals, 1231 (39%) aHUS and 116 (26%) C3G patients harbored 542 and 110 unique RVs, respectively, with an overlap of 42 unique variants between them. A significantly greater proportion of screened aHUS patients had at least one RV compared with C3G patients ($p < 0.0001$, two-tailed Fisher exact test). There were 0.44 unique RVs per case for aHUS compared with 0.95 for C3G. This suggested that almost every C3G case had a different unique RV, whereas aHUS cases were more likely to share the same RV. The proportion of aHUS patients with RVs (1231/3128 = 39.3%) (Fig. 2B) was low compared with literature reports (~60%). This difference most likely arose from the omission in the AF analyses of the 119 literature-sourced aHUS RVs in the Web database, for which no aHUS patient AF data from the six centers were available (green, Fig. 2A).

RV profiles of cases

Next, we compared the genetic RV profiles of aHUS and C3G patients. Among the 1231 aHUS patients and 116 C3G patients with RVs, 1024 (83%) and 82 (68%), respectively, harbored a single RV in 1 of the 13 genes (Fig. 2B). Two RVs were identified in 182 (15%) aHUS cases and 31 (27%) C3G cases (light blue bars, Fig. 2B), three RVs were identified in 24 (2%) aHUS cases and 3 (3%) C3G cases (red bars, Fig. 2B), and four RVs were identified in 1 (<1%) aHUS case but no C3G cases. However, the number of RVs per aHUS case in the dataset from Reference Center 6 was unknown, with the exception of three homozygotes. All other RV occurrences in this dataset were analyzed as one heterozygous aHUS case. When the 322 aHUS cases from Reference Center 6 were excluded, 702 of 909 (77%) harbored a single RV, 182 of 909 (20%) had two RVs, 24 of 909 (3%) had three RVs, and 1 (<1%) had four RVs. Excluding Reference Center 6, no significant difference was seen between aHUS and C3G in the cases with two RVs ($p = 0.113$, two-tailed Fisher exact test). For aHUS, cases with two RVs in *CD46* (23 cases, including 19 homozygotes; 11% of all aHUS cases with *CD46* RVs), were the most frequent, followed by *CFH* (37 cases, including 18

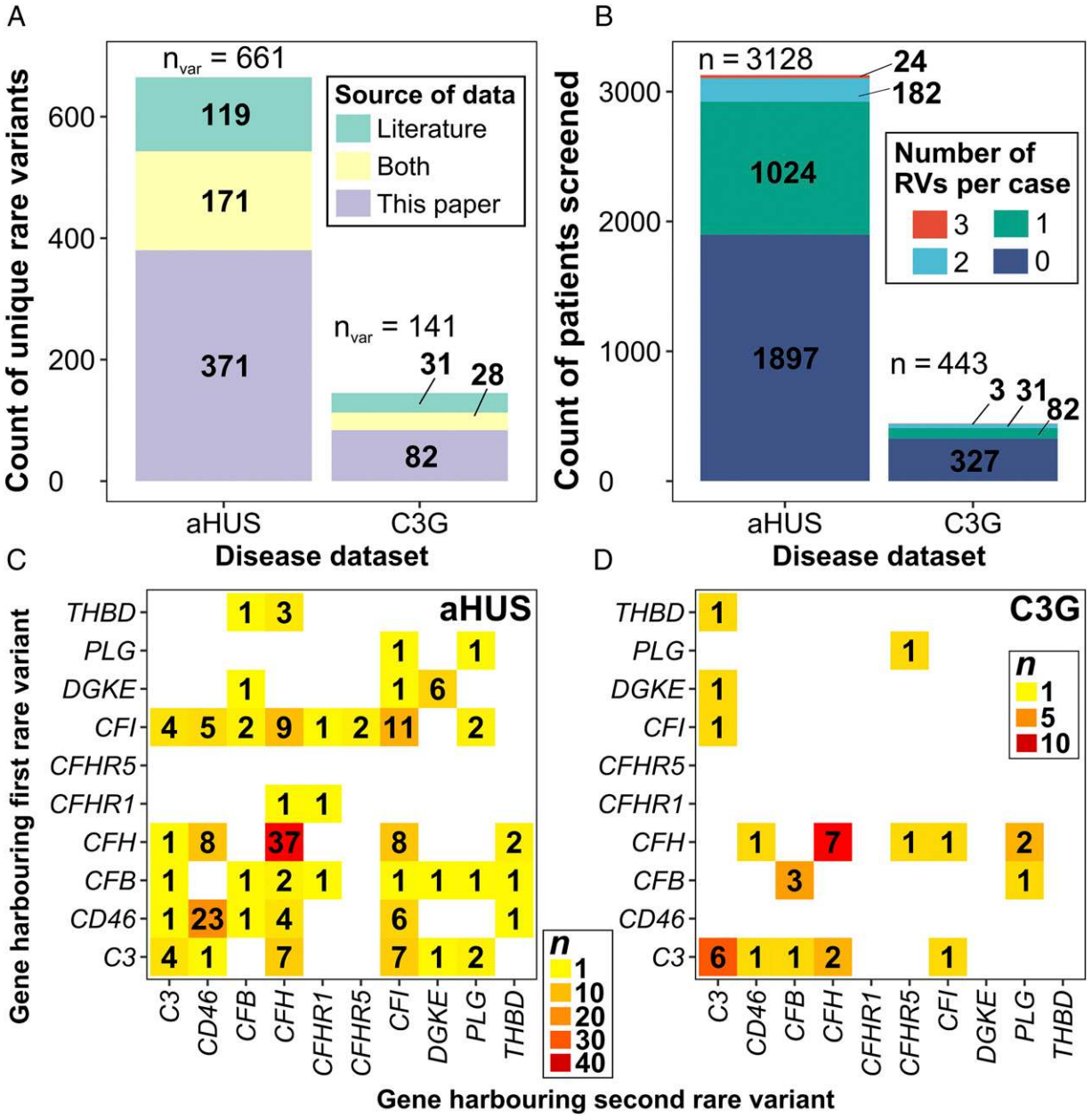

FIGURE 2. Summary of cases and variants in aHUS and C3G. (**A**) The source of the RV data in the database for aHUS and C3G. "Both" (yellow) indicates RVs that were identified in the laboratory-sourced datasets (purple) and published in the literature (green). (**B**) The number of unique RVs (0–3) per patient case in the aHUS and C3G datasets, totaling 3127 patients. For aHUS, there is an additional case with four RVs that is too small to be seen. (**C**) Matrix showing the genetic profiles of the 182 aHUS cases with compound heterozygous (in single or in two different genes) or homozygous RVs from (B). (**D**) Matrix showing the genetic profiles of the 31 C3G cases with compound heterozygous (in single or in two different genes) or homozygous RVs from (B).

homozygotes [9%]) (Fig. 2C, Table IV). For C3G, cases with two RVs in *CFH* (seven cases, including six homozygotes [30%]) were the most frequent, followed by *C3* (six cases, including two homozygotes [23%]) (Fig. 2D, Table IV). These analyses showed differences between aHUS and C3G (summarized in Fig. 2C, 2D). No *CFHR3* RV was seen in any aHUS or C3G case that had more than one RV.

Gender analyses

Finally, we examined gender dependence in aHUS and C3G cases. Of the 1231 aHUS patients with at least one identified RV, 36% (440) were female, and 29% (363) were male (Table V). In aHUS, assuming the remaining 35% of patients for which the gender was unknown showed a similar pattern, the bias toward females was significant ($p = 0.007$). However, when *CFH* was removed from the analyses, the numbers of females (244) and males (225) was not significantly different ($p = 0.38$) (Table VI). For the 116 C3G patients, no significant difference in the number of females (53) and males (63) was seen ($p = 0.35$) (Table VI). In terms of gender differences, our aHUS dataset (542 variants) showed a bias toward females for *CFH*, which is likely to be a reflection of potential triggering factors. For example, in pregnancy aHUS, most patients harbor *CFH* variants (55, 56). This trend was not observed for C3G.

RV pathogenicity classification

RVs were categorized for their pathogenicity using published experimental evidence, reference AFs, and in silico predictive

Table III. Total number of aHUS and C3G cases screened per gene

Gene	Disease	Total Number of Cases Screened by Reference Centers 1–6						
		1	2	3	4	5	6	All
C3	aHUS	410	480	286	252	409	618	2455
CD46	aHUS	524	480	286	461	578	613	2942
CFB	aHUS	395	480	286	328	350	618	2457
CFH	aHUS	662	480	286	483	578	639	3128
CFHR1	aHUS	7	250		442			699
CFHR3	aHUS		250		348			598
CFHR5	aHUS			286		31		317
CFI	aHUS	533	480	286	425	578	621	2923
DGKE	aHUS		76	286	191	150		703
PLG	aHUS			286				286
THBD	aHUS		480	286	348	214		1328
C3	C3G	115	160	104				379
CD46	C3G	142	160	104				406
CFB	C3G	115	160	104				379
CFH	C3G	179	160	104				443
CFHR5	C3G			104				104
CFI	C3G	144	160	104				408
DGKE	C3G		23	104				127
PLG	C3G			104				104
THBD	C3G		160	104				264

1, Institute of Genetic Medicine, Newcastle University; 2, Centro di Ricerche Cliniche per le Malattie Rare "Aldo e Cele Daccò," IRCCS-Istituto di Ricerche Farmacologiche "Mario Negri"; 3, Molecular Otolaryngology and Renal Research Laboratories, Carver College of Medicine, University of Iowa; 4, Department of Cellular and Molecular Medicine, Center for Biological Research and Center for Biomedical Network Research on Rare Diseases (Madrid, Spain); 5, Assistance Publique-Hopitaux de Paris, Hôpital Européen Georges Pompidou, Service d'Immunologie Biologique; 6, Department of Pediatric Nephrology, Radboud University Medical Center.

analyses, in accordance with genetic variant guidelines (1) (Fig. 3A). The average healthy genome also contains benign RVs that occur at similar AFs to disease-related RVs; therefore, AF analyses alone cannot be used as evidence for pathogenicity (43). In terms of pathogenicity, only gain-of-function RVs in the activators *C3* and *CFB* are expected to predispose for aHUS or C3G disease; in these cases, gain-of-function results from the loss of the ability to interact with an inhibitory complement regulator. The predictive tools PolyPhen-2 and SIFT were unable to predict gain-of-function phenotypes; thus, many *C3* and *CFB* RVs without experimental data could only be classified as uncertain significance (green bars, Fig. 3A). For example, the *C3* RV p.Arg161Trp was predicted to be damaging and deleterious, whereas functional studies showed a C3 gain-of-function (57). Therefore, the number of pathogenic or likely pathogenic RVs in *C3* and *CFB* was lower than for the other genes. In aHUS, the majority of RVs for *CFH*, *CFI*, *CD46*, and *DGKE* were pathogenic or likely pathogenic (red and light blue bars, Fig. 3A), which was attributed to their loss-of-function phenotypes. In C3G, the majority of RVs in all genes were of uncertain significance (green bars, Fig. 3A), with the exception of *CFH*. In terms of the nonpathogenic RVs in the aHUS and C3G datasets, a survey of all 542 RVs in aHUS and 110 RVs in C3G showed that 43 (8%) and 27 (25%) RVs, respectively, were categorized as benign and likely benign, and these were filtered out (Supplemental Table I). This suggested that the RVs in the C3G dataset may be less pathogenic compared with aHUS overall; this observation reinforced the importance of using classification guidelines prior to our protein domain hotspot analyses below. We also note that likely pathogenic RVs still require further functional studies to confirm or disprove the predicted effects and disease relevance of the variants.

RV abundance in genes

To identify differences in the frequency of unique RVs per gene between aHUS and C3G and, therefore, the molecular pathogenesis of both diseases, the abundance of RVs in each gene was analyzed. We stress that only those RVs that were not classified as benign or likely benign were analyzed (red, light blue, green, and gray bars, Fig. 3A). *CFHR1–4* genes were subject to multiple ligation-dependent probe assessment only and were not sequenced; thus, only LGRs were identified. In aHUS, *CFH* showed the most RVs (204 variants [41%]) (Fig. 3A), followed by *CFI* (82 variants [17%]), *CD46* (79 variants [16%]), and *C3* (68 variants [14%]). This outcome confirmed our 2014 analyses for *CFH*, *CFI*, *CD46*, and *C3* (9). Lesser abundances for aHUS were seen for *CFB* (20 [4%], *DGKE* (20 [4%]), *THBD* (11 [2%]), *PLG* (4 [1%]), *CFHR5* (3 [1%]), *CFHR1* (4 [1%]), *CFHR3* (2 [<1%]), and *CFP* (2 [<1%]). Only one unique variant in *CFHR4*, a *CFHR1/CFHR4* deletion, was seen in three heterozygous and one homozygous aHUS cases. This was classified under the *CFHR1* gene in the database. In C3G, *C3* (31 variants [37%]) and *CFH* (25 variants [28%]) showed the most RVs (red, light blue, green, and gray bars, Fig. 3A). Lesser abundances for C3G were seen for *CFI* (5 [6%]), *CD46* (1 [1%]), *CFB* (8 [10%]), *DGKE* (2 [2%]), *THBD* (2 [2%]), *PLG* (2 [2%]), *CFHR5* (1 [1%]), *CFHR1* (1 [1%]), *CFHR3* (1 [1%]), and *CFP* (0). It was concluded that RVs in *CFI* and *CD46* were substantially reduced in C3G compared with aHUS ($p = 0.0121$, $p < 0.0001$, respectively, two-tailed Fisher exact test). When analyzed in terms of genetic effect for aHUS and C3G, all RVs were nontruncating (yellow, Fig. 3B), with the exception of the aHUS variants in *DGKE* (mostly truncating) and LGRs in *CFHR1* and *CFHR3*. The most frequent RV in our aHUS dataset was *C3* p.Arg161Trp, which was seen at an AF of 1.16% (52 aHUS cases). However, in C3G, none of the RVs were notably more frequent than others. Data on RVs found in the complement inhibitor vitronectin and clusterin genes in the aHUS and C3G datasets were not available for analysis at the time of writing (58, 59).

Gene-based RV burden

To confirm that the amount of rare variation seen in the genes of aHUS and C3G patients was greater than in the genes of individuals without these diseases, we determined the burden of protein-altering rare variation (ExAC MAF < 0.01%) per gene for each dataset (*Materials and Methods*). These were compared with the

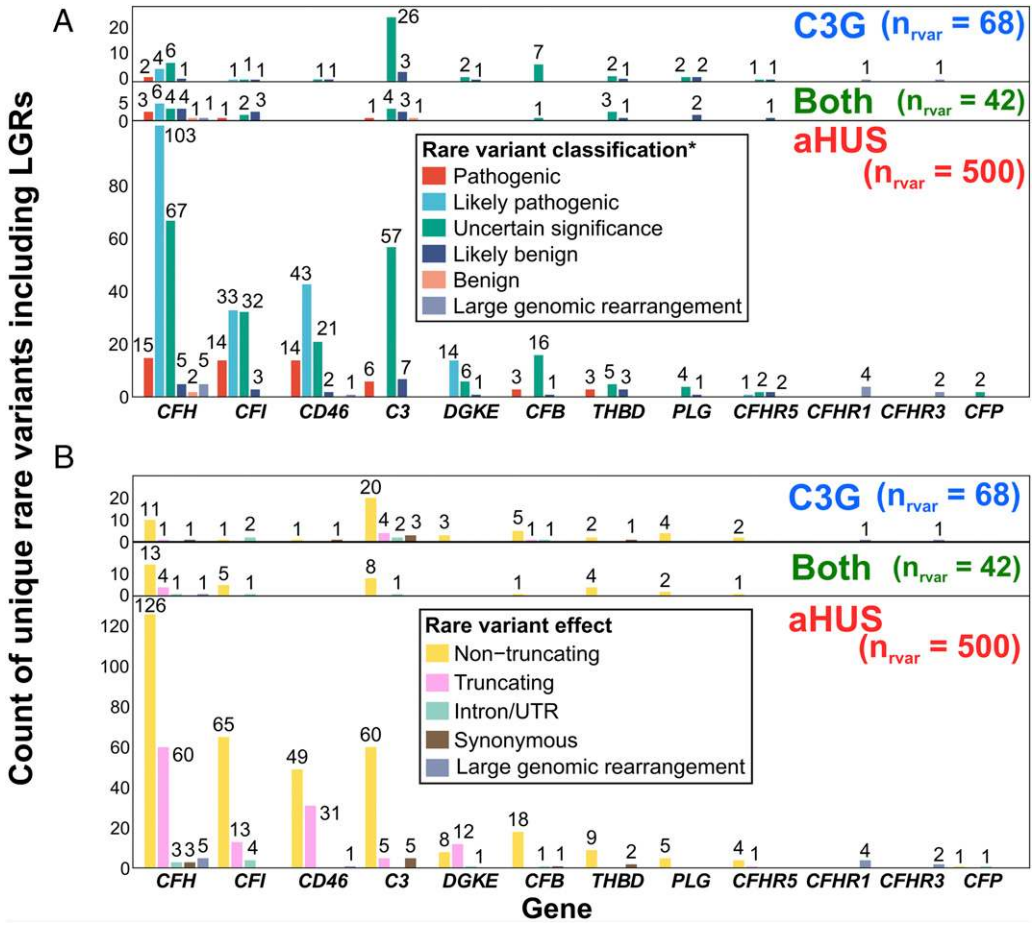

FIGURE 3. RV effects and classifications in aHUS and C3G. (**A**) In terms of pathogenicity, the total number of unique RVs for each gene and their classification, based on the pathology guidelines (*Materials and Methods*), are shown for the aHUS and C3G datasets, as well as for both datasets. (**B**) In terms of functional annotation (such as that used in ExAC), the total number of unique RVs for each gene and their effect on each protein are shown for the aHUS and C3G datasets, as well as for both datasets.

ExAC and EVS reference datasets (Fig. 4, Supplemental Tables II, III). Because ExAC and EVS did not contain data on LGRs, *CFHR1–4* were not analyzed. No aHUS or C3G AF data were available for *CFP*. This left 9 of 13 genes (*CFH*, *CFI*, *CD46*, *C3*, *DGKE*, *CFB*, *CFHR5*, *PLG*, and *THBD*) for analysis. A significantly greater burden of rare variation was revealed in patients in the aHUS dataset than in the ExAC and EVS datasets for five genes (*CFH*, *CFI*, *CD46*, *C3*, and *DGKE*, χ^2 test with the Yates correction using a Bonferroni corrected significance level of 0.0056; the five genes each gave $p < 0.0001$), as well as for *CFB* compared with the EVS dataset only (also $p < 0.0001$) (Fig. 4, Supplemental Tables II, III). No association with aHUS was observed with RVs in *THBD*, *PLG*, or *CFHR5*. The tests for these three genes showed a power > 80%; thus, their false-negative rates were expected to be very low. In the C3G dataset, *C3* ($p < 0.0001$ using ExAC, $p < 0.0001$ using EVS) and *CFH* ($p < 0.0001$ using ExAC, $p < 0.0001$ using EVS) showed a significantly greater burden of protein-altering rare variation in patients than in the ExAC and EVS datasets (χ^2 test with the Yates correction using a Bonferroni corrected significance level of 0.0056) (Fig. 4, Supplemental Tables II, III). C3G was also associated with RVs in *CFB* ($p < 0.0001$) and *THBD* ($p = 0.0052$) when only EVS was used as the reference dataset (green bar, Fig. 4, Supplemental Table III), despite both tests showing a lack of power (41 and 38%, respectively). The lack of association of *DGKE* and *PLG* with C3G may also relate to the lack of power (*DGKE*: 55% for ExAC and 68% for EVS; *PLG*: 32% for ExAC and 10% for EVS) shown in the tests for these genes.

Distribution of aHUS and C3G RVs in FH

The aHUS and C3G location and AF of the missense RVs for each gene resulted in the identification of mutational hotspots in each protein structure. For FH, as seen in the 2014 study, most aHUS missense RVs (78) occurred in the C-terminal 10 domains compared with the N-terminal 10 domains (47) ($p = 0.0042$) (Fig. 5A, Supplemental Table IV) (9). In aHUS, the total frequency of *CFH* alleles with a missense RV in the C-terminal 10 domains (3.2%) was significantly greater than for the N-terminal 10 domains (1.2%; χ^2 test with the Yates correction using a significance level of 0.05, $p < 0.0001$). In SCR-20, the total missense RV aHUS AF of 2.03% (Supplemental Table IV) was the highest for all 20 domains. These nonrandom distributions supported the functional association of SCR-20 with cell surface dysregulation in aHUS, although laboratory experimentation and functional characterization will be required to validate this result. This was still the case when normalized for the size of the FH domain (62 of 1231 residues [5.0%]), giving 37.2% (Fig. 5A, Supplemental Table IV). Most of the six FH domains (SCR-2, SCR-5, SCR-8, SCR-12, and SCR-13) with the least number of aHUS RVs did not correspond to known FH binding sites. For C3G, in contrast to aHUS, the C3G missense RVs in FH were clustered at the N-terminal C3b binding site, with comparatively few at the nonsurface-associated C-terminal domains, such as SCR-15 (Fig. 5A). No C3G clusters were seen in the heparin-binding regions of FH. The only three unique C3G missense RVs found in SCR-7 (p.Cys431Tyr; C3G

Table IV. Homozygous RVs present in aHUS and C3G

Gene	cDNA Change	Protein Change	Classification	Condition	Patients (*n*)
C3	c.3322_3333del	p.1108_1111del	Likely benign	aHUS	1
CD46	c.97G>C	p.Asp33His	Uncertain significance	aHUS	2
CD46	c.100G>A	p.Ala34Thr	Uncertain significance	aHUS	1
CD46	c.286+2T>G	—	Pathogenic	aHUS	4
CD46	c.286+1G>C	—	Pathogenic	aHUS	1
CD46	c.496T>C	p.Cys157Arg	Uncertain significance	aHUS	1
CD46	c.535G>C	p.Glu179Gln	Pathogenic	aHUS	1
CD46	c.565T>G	p.Tyr189Asp	Likely pathogenic	aHUS	1
CD46	c.718T>C	p.Ser240Pro	Pathogenic	aHUS	2
CD46	c.736T>A	p.Phe246Ile	Uncertain significance	aHUS	1
CD46	c.800_820del	—	Uncertain significance	aHUS	1
CD46	c.810T>G	p.Cys270Trp	Uncertain significance	aHUS	1
CD46	c.811_816delGACAGT	p.Asp271_Ser272del	Pathogenic	aHUS	1
CD46	c.881C>T	p.Pro294Leu	Uncertain significance	aHUS	1
CD46	c.1027+2T>C	—	Pathogenic	aHUS	1
CFH	c.79_82delAGAA	—	Likely pathogenic	aHUS	1
CFH	c.157C>T	p.Arg53Cys	Pathogenic	aHUS	2
CFH	c.158G>A	p.Arg53His	Likely pathogenic	aHUS	1
CFH	c.2697T>A	p.Tyr899[a]	Likely pathogenic	aHUS	2
CFH	c.2880delT	p.Phe960fs	Likely pathogenic	aHUS	1
CFH	c.2918G>A	p.Cys973Tyr	Uncertain significance	aHUS	1
CFH	c.3048C>A	p.Tyr1016[a]	Likely pathogenic	aHUS	2
CFH	c.3628C>T	p.Arg1210Cys	Pathogenic	aHUS	1
CFH	[c.3674A>T; 3675_3699del]	p.Tyr1225Tyrfs[a]38	Likely pathogenic	aHUS	2
CFH	c.3693_3696 delATAG	p.[a]1232Ilefs[a]38	Likely pathogenic	aHUS	5
CFI	c.341T>C	p.Val114Ala	Uncertain significance	aHUS	1
CFI	c.1357T>C	p.Cys453Arg	Uncertain significance	aHUS	1
CFI	c.1456T>C	p.Trp486Arg	Uncertain significance	aHUS	1
CFI	c.1642G>C	p.Glu548Gln	Uncertain significance	aHUS	2
DGKE	c.889-1G>A	p.IVS5-1	Likely pathogenic	aHUS	1
DGKE	c.966G>A	p.Trp322[a]	Likely pathogenic	aHUS	4
DGKE	c.1000C>T	p.Gln334[a]	Likely pathogenic	aHUS	1
DGKE	c.1608_1609del	p.His536Glnfs[a]16	Likely pathogenic	aHUS	1
PLG	c.1481C>T	p.Ala494Val	Likely benign	aHUS	1
C3	c.168_169delTG	p.Thr56Thrfs[a]16	Uncertain significance	C3G	1
C3	c.1682G>A	p.Gly561Asp	Uncertain significance	C3G	1
CFB	c.2035C>T	p.Arg679Trp	Uncertain significance	C3G	1
CFH	c.232A>G	p.Arg78Gly	Pathogenic	C3G	1
CFH	c.262C>A	p.Pro88Thr	Uncertain significance	C3G	1
CFH	c.694C>T	p.Arg232[a]	Likely pathogenic	C3G	1
CFH	c.3286T>A	p.Trp1096Arg	Uncertain significance	C3G	3

[a]Non-sense variant.

AF: 0.2%) and SCR-20 (p.Arg1210Cys, 0.1%; p.Cys1218Arg, 0.1%) in the heparin-binding regions were also seen in aHUS at similar AFs of 0.01, 0.4, and 0.01%, respectively, suggesting an overlap in phenotypes. The C3G missense RVs in FH were also missing in 11 SCR domains (SCR-4–6, SCR-8/9, SCR-11–13, SCR-16/17, and SCR-19). There was a C3G missense RV cluster at the N-terminal SCR-1/3 C3b binding region (31) (Fig. 5A, Supplemental Table IV); we infer that SCR-1–3 may be a mutational hotspot for C3G. In summary, the distribution of FH hotspots between aHUS and C3G showed clear differences between the two diseases.

Distribution of aHUS and C3G RVs in C3

For C3, the 67 aHUS C3 missense RVs occurred in 12 of the 16 C3 domains (Fig. 5B, Supplemental Table IV). Macroglobulin (MG)-2 showed the highest missense RV AF for C3 (1.3%), normalized by the proportion of domain residues (21.7%), followed by MG-6b (0.5 and 18.5%) (Fig. 5B). Thus, the MG-2

Table V. Demographics of the 1231 aHUS and 116 C3G cases showing an identified RV

Disease	Gender	Cases Aged >18 y (*n*; %)[a]	Cases Aged <18 y (*n*; %)[b]	YOB Unknown (*n*; %)	Total (*n*; %)
aHUS	Female	328 (27)	84 (7)	28 (2)	440 (36)
aHUS	Male	254 (20)	84 (7)	25 (2)	363 (29)
aHUS	Unknown	2 (<1)	1 (<1)	425 (34)	428 (35)
aHUS	All	584 (48)	169 (14)	478 (38)	1231
C3G	Female	34 (29)	12 (10)	7 (6)	53 (46)
C3G	Male	40 (35)	16 (14)	7 (6)	63 (54)
C3G	Unknown	0	0	0	0
C3G	All	74 (64)	28 (24)	14 (12)	116

[a]Born in 1997 or before.
[b]Born after 1997.
YOB, year of birth.

Table VI. Gender of the 1231 aHUS and 116 C3G cases with an identified RV

Disease	Gene	Female	Male	Unknown	Proportion of Females (Known) (%)	*p* Value[a]
aHUS	*C3*	76	61	99	55	0.20
aHUS	*CD46*	77	91	67	46	0.28
aHUS	*CFB*	13	11	11	54	0.68
aHUS	*CFH*	196	138	196	59	**0.002**
aHUS	*CFHR1*	8	13	1	38	0.28
aHUS	*CFHR3*	1	1	0	50	1.00
aHUS	*CFHR5*	4	3	0	57	0.71
aHUS	*CFI*	80	60	54	57	0.09
aHUS	*CFP*	1	1	0	50	1.00
aHUS	*DGKE*	16	9	0	64	0.16
aHUS	*PLG*	7	5	0	58	0.56
aHUS	*THBD*	16	8	10	67	0.10
aHUS	ALL[b]	440	363	428	55	**0.007**
C3G	*C3*	23	20	0	53	0.65
C3G	*CD46*	1	1	0	50	1.00
C3G	*CFB*	4	4	0	50	1.00
C3G	*CFH*	19	24	0	44	0.45
C3G	*CFHR1*	0	1	0	0	0.32
C3G	*CFHR3*	0	1	0	0	0.32
C3G	*CFHR5*	0	4	0	0	0.05
C3G	*CFI*	5	6	0	45	0.76
C3G	*DGKE*	1	2	0	33	0.56
C3G	*PLG*	2	5	0	29	0.26
C3G	*THBD*	2	7	0	22	0.10
C3G	ALL[b]	53	63	0	46	0.35

Bold text denotes a *p* value less than the Bonferroni corrected significance level.

[a]The *p* values were calculated with a two-tailed χ^2 test with Yates correction using a Bonferroni corrected significance level of 0.0042 for aHUS and 0.0045 for C3G. The "ALL" genes category was subject to a significance level of 0.05 (no Bonferroni correction).

[b]"ALL" is not equal to the gene sum, because some patients possessed RVs in more than one unique gene.

and MG-6b domains were inferred to be aHUS hotspots. The 29 C3G missense RVs occurred in 13 of the 16 C3 domains (Supplemental Table IV). In contrast to aHUS, these were spread more evenly throughout C3, and no hotspots were inferred (Fig. 5B). No aHUS or C3G variants involved the key C3 thioester residues (Glu^{991}, Cys^{998}, His^{1104}, and Glu^{1106}) or the anaphylatoxin, aNT, or β-sheet (CUB)-a domains. Again, the distribution of C3 missense RVs between aHUS and C3G showed clear differences between the two diseases.

Distribution of aHUS and C3G RVs in FI, MCP, FB, and other proteins

The 65 missense RVs in aHUS were distributed across all five domains of FI (Fig. 5C). The 45 aHUS rare missense variants were distributed across the four MCP domains (Fig. 5D). For FB, there was no aHUS or C3G missense RV in SCR-1 or SCR-3 (Fig. 5E). In FB, most aHUS missense RVs occurred in the von Willebrand factor type A domain, whereas the C3G missense RVs were spread across the SCR-2, von Willebrand factor type A, and serine protease domains. No mutational hotspots were evident for aHUS or C3G in FI, MCP, or FB.

Missense RVs mapped onto protein structures

The missense RVs for aHUS and C3G were mapped onto protein structural models for FH (48), C3 (49), FI (50), MCP (51), and FB (52). For FH, in confirmation of the above analyses, SCR-5 and SCR-12/13 were sparsely populated, implying that the variants were spatially clustered only at functional regions of FH (Fig. 6A). For C3, MG-2 showed a higher density of aHUS variants (red, Fig. 6B), whereas a higher density of C3G variants occurred at the core of C3 (yellow). For FI, the aHUS variants were distributed evenly throughout FI (Fig. 6C). For CD46, the aHUS variants were likewise distributed throughout the protein (Fig. 6D). Similar patterns were seen for FB (Fig. 6E).

MAF analyses

The AFs of each of the protein-altering RVs per gene in the aHUS and C3G datasets were compared with the corresponding AF in ExAC. The two *CFP* RVs in aHUS were not analyzed because there were no available patient AF data. *CFHR1*, *CFHR3*, and *CFHR4* were not analyzed because LGRs were not included in ExAC at the time of writing. No RVs (with ExAC MAF $< 0.01\%$) were found to be significantly more common in ExAC, than in aHUS or C3G. AFs and their significance are shown on the database for users.

Discussion

Summary of RVs in aHUS and C3G

Compared with our 2014 study (9), in which 324 aHUS- and C3G-associated genetic variants in *CFI*, *CFH*, *C3*, and *CD46* were used to identify variant hotspots in these four proteins (http://www.fh-hus.org), this more detailed study reports 610 RVs in 13 genes from 3128 aHUS and 443 C3G patients, together with their associated AFs. The expansion to 13 genes reflects new candidate genes for potential association with these diseases and clarifies the extent to which they are indeed associated. To our knowledge, this is the largest study of aHUS and C3G patients to date. The much increased totals of RVs and cases now make possible, for the first time, AF analyses and RV burden analyses of the aHUS and C3G datasets, including comparisons with three reference datasets. In turn, the AF analyses provided more detailed statistical analyses of the RV distributions in both diseases. The analyses of distinct protein domain hotspots for aHUS and C3G clarify molecular differences that rationalize the occurrence of their two phenotypes, the involvement of the RVs that are present, and the molecular mechanisms involved in both diseases. In particular, this study has reduced the earlier knowledge gap in the genetics and genotype–phenotype correlations of C3G to bring these closer to that of aHUS (1).

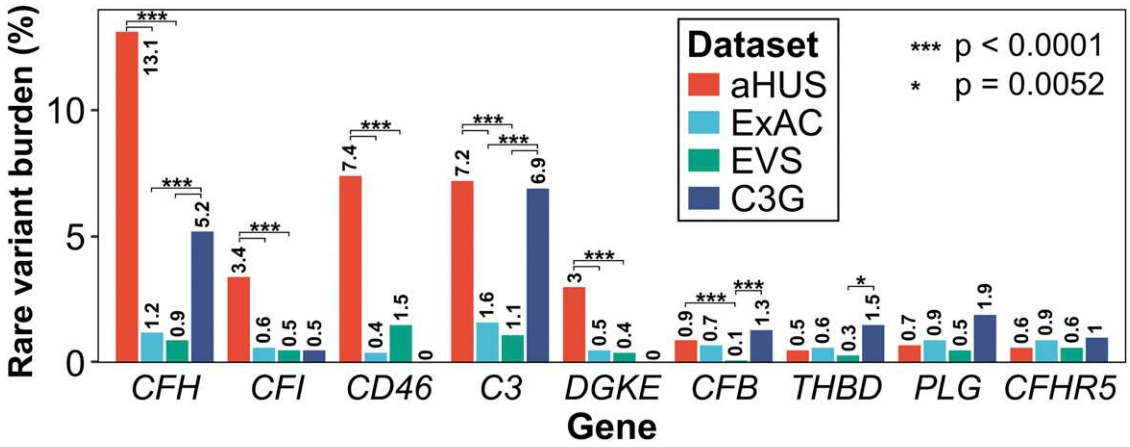

FIGURE 4. RV burden (%) per gene for the nine relevant genes in the four datasets: aHUS (allele number [AN] 634–6,256), ExAC (AN 74,194–121,246), EVS (AN 8,202–13,005), and C3G (AN 208–886). These were based on an ExAC MAF cutoff of 0.01%.

Differences between aHUS and C3G using AF analyses

AF analyses revealed three new insights into the individual RVs associated with aHUS and C3G (and not the full genes). Our analyses raised the question of what AF cutoff to use. First, our AF analyses verified the rarity of 97% of the aHUS and C3G variants compared with the ExAC reference (Fig. 1), especially given the ability of ExAC to resolve ultra-RV AFs as low as 1×10^{-5} (0.001%) (42). Such disease-predisposing alleles in aHUS and C3G are, by definition, deleterious. In theory, evolutionary pressures will maintain these alleles at very rare frequencies in the general population through negative selection (60–62). Other studies have reported that variant pathogenicity increases with rarity and low AF (43). Therefore, these results justify our focus on RVs with AF $< 1\%$ in the reference datasets. A more stringent RV AF cutoff of 0.01% is applicable to rare diseases of Mendelian inheritance (43). This 0.01% cutoff was used for the RV burden calculations to restrict them to RVs with a higher confidence of pathogenicity (Fig. 4, Supplemental Tables II, III). However, RVs observed in aHUS and C3G have reduced penetrance; disease-free individuals also harbor some of these pathogenic RVs, and the onset of aHUS or C3G disease depends upon a trigger and other factors. To enable our hotspot analyses, a less stringent reference AF cutoff $< 0.1\%$ was used, accompanied by experimental data and prediction tools, as specified in genetic variant classification

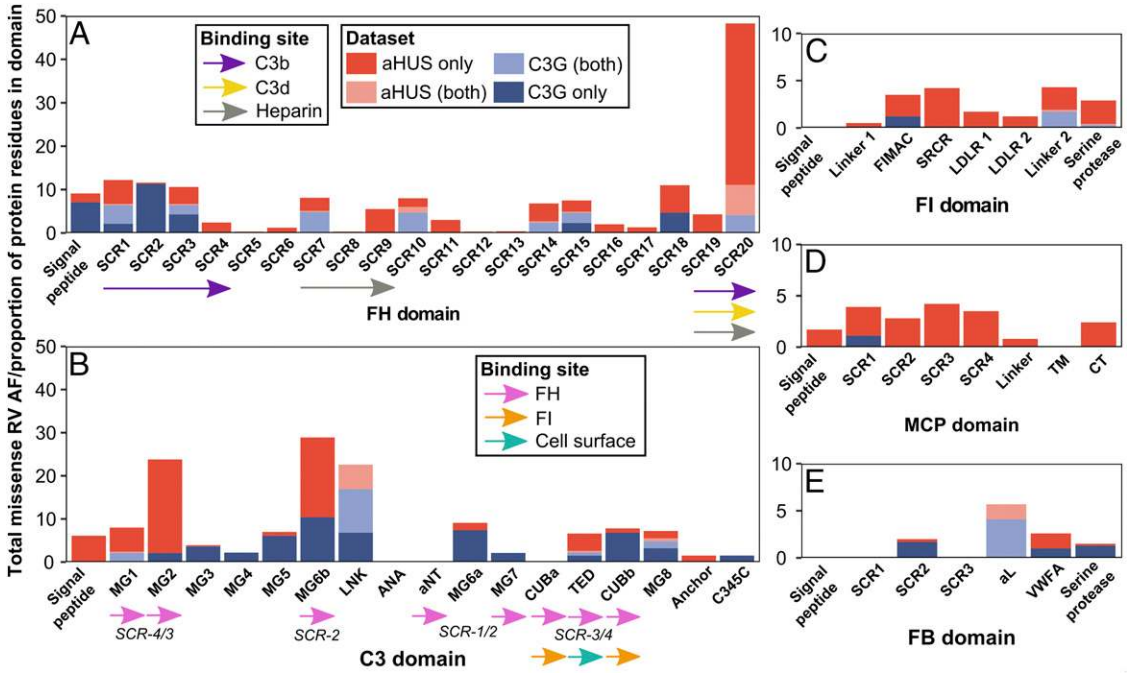

FIGURE 5. Distribution and disease AFs of nonbenign missense RVs in the domains of FH, C3, FI, MCP, and FB in the aHUS and C3G datasets. The largest bars correspond to missense RV hotspots (e.g., FH SCR-20 for aHUS; SCR-18 for C3G). Each domain missense RV AF is normalized for its size by dividing it by the proportion of residues in the protein domain. In (**A**)–(**E**), red represents the total AF missense RVs identified in the aHUS dataset, and dark blue represents the total AF missense RVs identified in the C3G dataset. For missense RVs identified in both datasets, pink represents the AFs for aHUS, and light blue represents the AFs for C3G. The domain names are shown on the *x*-axes. (A) The total AF of missense RVs in each of the 20 SCR domains in FH. The functional binding sites associated with each SCR domain are represented by colored arrows below the *x*-axis. (B) Total AF of missense RVs in each C3 domain. The functional binding sites associated with each C3 domain are represented by arrows that correspond to the SCR domains in FH (pink) or to other sites in FI or the cell surface. The C3d binding site on SCR-19/20 corresponds to the TED domain (not shown). Total AF of missense RVs in each FI (C), MCP (D), or FB (E) domain.

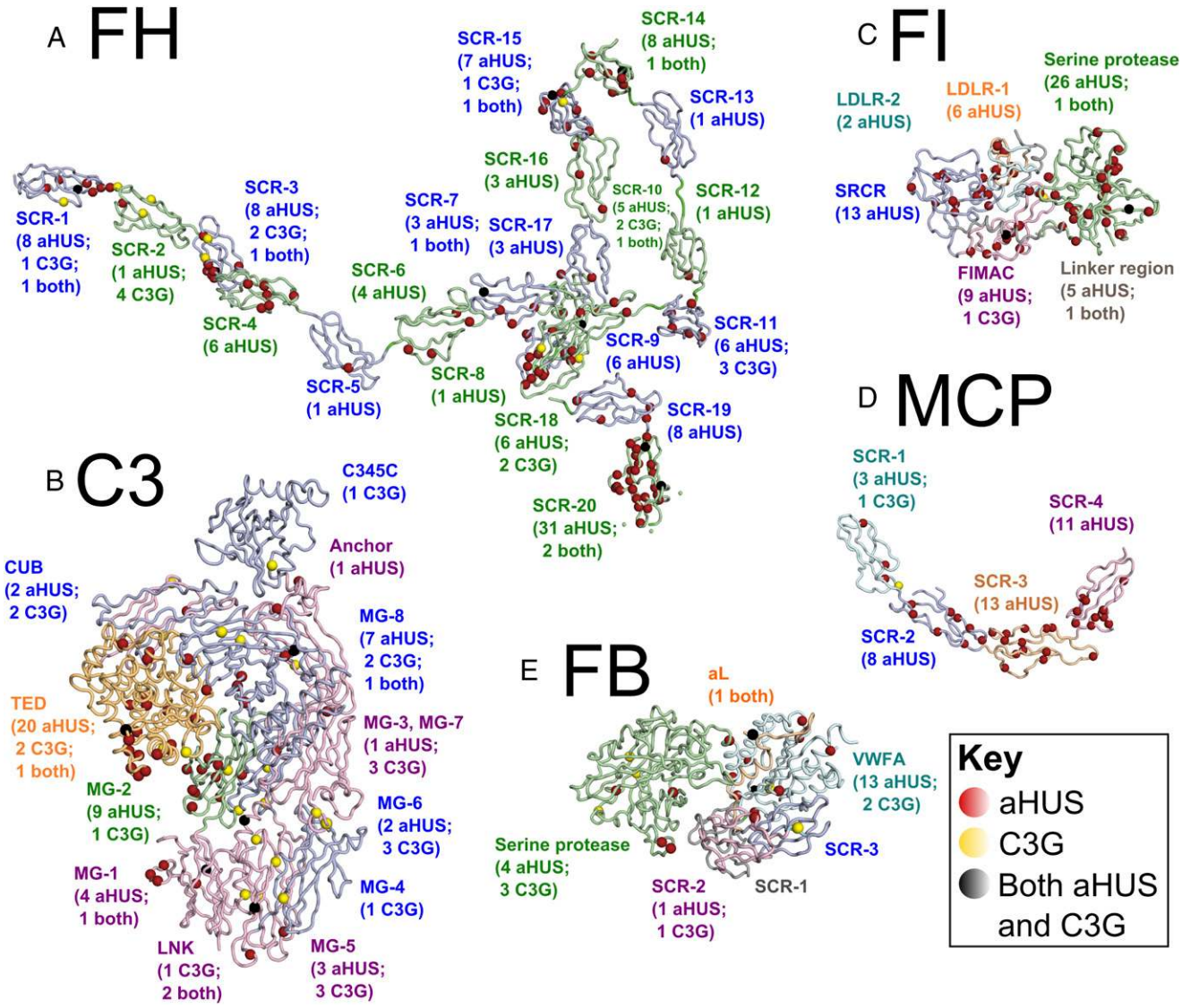

FIGURE 6. Missense RVs mapped onto protein structural models. Only the aHUS and C3G RVs that were not classified as benign or likely benign are shown. For FH (**A**), C3 (**B**), FI (**C**), MCP (**D**), and FB (**E**), the red spheres represent missense RVs identified in the aHUS dataset, the yellow spheres represent missense RVs in C3G, and the black spheres represent missense RVs in aHUS and C3G. Note that some of the spheres overlap, especially if different RVs affect the same amino acid. The protein domains are shown in alternating colors (A) or in unique colors (B–E) and are labeled in that color for clarity.

guidelines (1). Some variants show AFs > 1% (Table II). Although these may be risk factors for aHUS (e.g., *CFH* p.Val62Ile) (12), their analysis was beyond the scope of this study. All of the variants are available for viewing in the Database of Complement Gene Variants by adjusting the value of the ExAC AF filter.

The second insight involved AF differences between aHUS and C3G that relate to their different phenotypes. Significantly greater proportions of C3G RVs were identified in the reference datasets, with AFs between 0 and 1% rather than 0%, unlike aHUS (Fig. 1). This meant that the C3G RVs occurred more frequently in individuals without C3G (the reference datasets) than did the aHUS RVs. This result may explain the mostly chronic presentation of C3G that accumulates over time, as opposed to the mostly acute presentation of aHUS.

Third, AF analyses of the disease datasets (i.e., not the reference datasets) revealed differences between the most common RVs in the aHUS and C3G datasets that were likely to reflect their different phenotypes. In our aHUS dataset, although *CFH* had the most RVs, the most frequent RV was p.Arg161Trp in *C3*, corresponding to an AF of 1.16% (52 aHUS cases). *C3* p.Arg161Trp has a surface exposed position in the MG-2 domain (Fig. 6B) and forms a hyperactive C3 convertase with an increased affinity for FB, thus leading to overactivation of the AP (57). *C3* p.Arg161Trp was not seen in the reference datasets, and classification guidelines confirmed its pathogenicity (1). In our C3G dataset, *C3* had the most RVs. Further analyses (below) reveal distinct domain hotspots for aHUS and C3G.

RV burden testing

The RV burden is the proportion of screened cases with an identified protein-altering RV for which the ExAC AF was <0.01%. As opposed to the AF analyses, which look at each variant one by one, the RV burden provides insight into all of the RVs associated with each gene. In general, RV burden tests assume that all tested RVs influence the phenotype in the same direction (63). Therefore, we separated RVs into "truncating" (loss-of-function) and "nontruncating" (either loss- or gain-of-function, or neutral) to aid interpretations. RV burden tests showed clear

differences between aHUS and C3G compared with reference datasets, and this clarified the molecular mechanisms of the two diseases. Previous knowledge of experimental functional characterization of some of the RVs collectively support that aHUS is more related to surface AP dysregulation, whereas C3G is more related to fluid-phase AP dysregulation. In this article, we now extend these earlier functional results.

For aHUS, our RV burden analyses confirmed the association of rare variation in six genes: *CFH*, *CFI*, *CD46*, *DGKE* (18, 64), *C3*, and *CFB*. For the five genes of the AP, *CFH*, *CFI*, *C3*, and *CFB* are involved in cell surface and fluid-phase regulation, but *CD46* is involved only in cell surface regulation. Therefore, the RV burden analyses suggest that aHUS involves defects that result in cell surface and fluid-phase dysregulation. Different domains of C3 and FH are involved in cell surface dysregulation compared with fluid-phase dysregulation, and this is explored in terms of aHUS and C3G RV distributions in the next section. The protein encoded by the sixth gene, *DGKE*, is found in endothelium, platelets, and podocytes and normally inactivates the signaling of arachidonic acid–containing diacylglycerols that activate protein kinase C and promote thrombosis. Thus, loss of DGKE function may result in a prothrombotic state and lead to microangiopathic hemolytic anemia seen in aHUS outside the complement system (18). For three other genes, *CFHR5*, *PLG*, and *THBD*, no association was observed between the RVs in these and aHUS. Each of their tests showed a high power (>80%). This is unexpected from their known function: complement factor H-related (FHR)5 is likely to compete with FH for regulation (65, 66), whereas plasminogen and thrombomodulin (THBD) are inhibitors of thrombosis, and THBD also regulates complement (67–70).

For C3G, in contrast, RV burden analyses showed that four genes (*C3*, *CFH*, *CFB*, and *THBD*) were associated with C3G, whereas two genes (*CD46* and *CFI*) were not. This outcome suggested that C3G is not caused by defects in cell surface regulation by *CD46* or defects in cell surface or fluid-phase regulation by FI. Our results also suggested that non-LGR RVs in *CFHR5* and *PLG* are not causative for C3G. We did not analyze LGRs in *CFHR5*, such as those identified in CFHR5 glomerulopathy. Despite pathogenic RVs in known cardiac genes not being overrepresented in ExAC (39), *THBD* is involved in cardiac disease cases. Furthermore, the prevalence of RVs may differ across different research centers, especially for genes such as *CFHR5* and *PLG* in aHUS and *CFHR5*, *PLG*, and *DGKE* in C3G, which are less mutated and/or were sequenced by only one or two centers. This outcome is potentially affected by sampling bias, thus being difficult to interpret, and the results for these genes should be considered as preliminary only. In addition, the lack of association of *DGKE* and *PLG* with C3G may also be related to a lack of power (10–68%) shown by their tests.

Hotspots for missense RVs

The 2014 identification of hotspots in four complement proteins (9) can now be expanded to examine clusters of missense RVs in aHUS and C3G. Based on RVs with 0.1% reference AF cutoffs, clear differences were seen between the phenotypes of aHUS and C3G at the molecular level. In particular, RV "hotspots" were identified in FH and C3 that could be rationalized on the basis of their importance in protein–protein interactions. For example, FH is involved in fluid-phase and cell surface regulation by being a cofactor for FI, possessing C3 convertase decay-accelerating activity, and blocking the formation of the C3 convertase. Despite that FH N-terminal SCR-1/4 and C-terminal SCR-19/20 domains bind to C3b, which is required for fluid-phase and cell surface regulation, the C-terminal region of FH is critical for cell surface regulation (Fig. 6A) and is not required for fluid-phase regulation (71). Thus, surface-exposed missense RVs in *CFH* that map to individual SCR domains can be correlated with FH function. Other variants predicted to affect FH stability may lead to FH aggregation, making it unable to regulate the fluid or cell surface phase. In the fluid phase, FH is the only AP complement regulator that has cofactor activity for FI; however, at the cell surface, FH and MCP can act as the FI cofactor. Thus, if the FH C-terminal SCR-19/20 domains are compromised, wild-type MCP may save cell surface regulation. An additional scenario is that, if the FH variant is heterozygous, this would affect only half of the FH in plasma, altering the resulting phenotype. Overall, the consequence of each RV on FH function can be complex to interpret.

Different FH SCR domains were identified as hotspots in aHUS and C3G. For aHUS, AF analyses confirmed that SCR-20 with 31 RVs and an RV density of 37.2% was a notable missense hotspot (Figs. 5A, 6A, Supplemental Table IV). SCR-20 is functionally important for FH binding to C3b, C3d, heparin-like oligosaccharides, and sialic acid (9, 72–74). The occurrence of SCR-20 as a RV hotspot is well explained by the disruption of FH binding to surfaces, leading to host cell damage from excess complement activation caused by unregulated C3b. aHUS missense RVs were also identified in the remaining 19 SCR domains in FH (Supplemental Table IV). Four of the five FH domains (SCR-2, SCR-5, SCR-8, SCR-12, and SCR-13) do not correspond to known FH binding sites and have only single missense RVs. The distribution in Fig. 5A suggested that the aHUS missense RVs primarily affect FH cell surface binding.

In contrast, for C3G, the missense RVs were clustered at the N-terminal C3b binding site (SCR-2/3) (Fig. 5A). No missense RVs in C3G were now clustered at cell surface heparin binding sites in SCR-6/7 or SCR-20. The SCR-2/3 domains were identified as C3G hotspots, likely attributed to its binding to MG-2 and MG-6 in C3b (Fig. 5A) (72, 75). This different clustering best correlates our C3G variants with dysregulation of the complement AP in the fluid phase (C3, FH) and not at the cell surface (CD46).

Different C3 domains were likewise identified as hotspots in aHUS and C3G. For aHUS, the MG-2 and MG-6b domains with the highest RV density (Supplemental Table IV) were deduced to be RV hotspots (Fig. 6B). Both MG domains interact with FH SCR-2 and SCR-3 to enable C3 regulation by FH in the fluid phase and on the cell surface (75). Disruption of the MG-2 and MG-6b domains would reduce C3 regulation by FH. It is not clear whether these two C3 domains also bind MCP, thus affecting cell surface regulation further (76). Although the TED domain contained 22 RVs, its RV density, which takes into account the number of residues in the domain (300), was not as high as might be expected from its functionally important thioester group and its binding to cell surfaces.

In contrast, for C3G, too few missense RVs in C3 have been reported for a clear outcome. The RVs were distributed in 11 of its 13 domains, with no clustering seen.

Usefulness of the database

The new Database of Complement Gene Variants enhances our understanding of rare genetic variants in aHUS and C3G for clinical applications. Improvements include the use of AFs, predictive comparisons of wild-type and mutant amino acids, in silico analyses using PolyPhen-2 and SIFT, examination of evolution-conserved residues across species, and correlations with functional binding sites. These tools enable clinicians to assess RVs in disease (e.g., to investigate which variants within these genes conferred predisposition to aHUS and C3G) and to identify mutational hotspots within these protein structure. This is especially

useful for variants of uncertain significance for which no experimental data exist. Ethnicity data were recorded for <50% of the aHUS and C3G patients in our datasets. RVs are displayed on the database in comparison with ExAC ethnicity data, with a full record of ethnicity. Although the disease datasets are incomplete in this regard, the new Web database has the capacity to capture new ethnicity data for aHUS and C3G cases for future AF comparisons. Because the six renal clinics in this study are based in western Europe and the United States, which are also the source of much of the ExAC dataset, the effect of ethnicity on aHUS and C3G analyses is expected to be minimal.

Acknowledgments

We thank Pauline Bordereau for excellent technical support. S.R.d.C. is a member of the Centro de Investigaciones Biológicas Intramural Program "Molecular Machines for Better Life."

Disclosures

A.J.O., S.R.d.C., D.P.G., M.N., S.J.P., and V.F.-B. have received honoraria from Alexion Pharmaceuticals for giving lectures and participating in advisory boards. Newcastle University has received fees from Alexion Pharmaceuticals and Akari Therapeutics for lectures and consultancy undertaken by T.H.J.G. and D.K. D.K. is a director of and scientific advisor to Gyroscope Therapeutics. G.R. has consultancy agreements with AbbVie, Alexion Pharmaceuticals, Bayer Healthcare, Reata Pharmaceuticals, Novartis Pharma, AstraZeneca, Otsuka Pharmaceutical Europe, and Concert Pharmaceuticals, for which no personal remuneration was accepted; compensations were paid to his institution for research and educational activities. The other authors have no financial conflicts of interest.

References

1. Goodship, T. H., H. T. Cook, F. Fakhouri, F. C. Fervenza, V. Frémeaux-Bacchi, D. Kavanagh, C. M. Nester, M. Noris, M. C. Pickering, S. Rodríguez de Córdoba, et al; Conference Participants. 2017. Atypical hemolytic uremic syndrome and C3 glomerulopathy: conclusions from a "kidney disease: improving global outcomes" (KDIGO) controversies conference. *Kidney Int.* 91: 539–551.
2. Bu, F., N. Borsa, A. Gianluigi, and R. J. Smith. 2012. Familial atypical hemolytic uremic syndrome: a review of its genetic and clinical aspects. *Clin. Dev. Immunol.* 2012: 370426.
3. Warwicker, P., T. H. Goodship, R. L. Donne, Y. Pirson, A. Nicholls, R. M. Ward, P. Turnpenny, and J. A. Goodship. 1998. Genetic studies into inherited and sporadic hemolytic uremic syndrome. *Kidney Int.* 53: 836–844.
4. Fakhouri, F., J. Zuber, V. Frémeaux-Bacchi, and C. Loirat. 2017. Haemolytic uraemic syndrome. *Lancet* 390: 681–696.
5. Sansbury, F. H., H. J. Cordell, C. Bingham, G. Bromilow, A. Nicholls, R. Powell, B. Shields, L. Smyth, P. Warwicker, L. Strain, et al. 2014. Factors determining penetrance in familial atypical haemolytic uraemic syndrome. *J. Med. Genet.* 51: 756–764.
6. Noris, M., and G. Remuzzi. 2015. Glomerular diseases dependent on complement activation, including atypical hemolytic uremic syndrome, membranoproliferative glomerulonephritis, and C3 glomerulopathy: core curriculum 2015. *Am. J. Kidney Dis.* 66: 359–375.
7. Esparza-Gordillo, J., E. Goicoechea de Jorge, A. Buil, L. Carreras Berges, M. López-Trascasa, P. Sánchez-Corral, and S. Rodríguez de Córdoba. 2005. Predisposition to atypical hemolytic uremic syndrome involves the concurrence of different susceptibility alleles in the regulators of complement activation gene cluster in 1q32. *Hum. Mol. Genet.* 14: 703–712.
8. Bresin, E., E. Rurali, J. Caprioli, P. Sanchez-Corral, V. Frémeaux-Bacchi, S. Rodríguez de Cordoba, S. Pinto, T. H. Goodship, M. Alberti, D. Ribes, et al; European Working Party on Complement Genetics in Renal Diseases. 2013. Combined complement gene mutations in atypical hemolytic uremic syndrome influence clinical phenotype. *J. Am. Soc. Nephrol.* 24: 475–486.
9. Rodriguez, E., P. M. Rallapalli, A. J. Osborne, and S. J. Perkins. 2014. New functional and structural insights from updated mutational databases for complement factor H, factor I, membrane cofactor protein and C3. *Biosci. Rep.* 34: 635–649.
10. Nester, C. M., T. Barbour, S. Rodríguez de Cordoba, M. A. Dragon-Durey, V. Frémeaux-Bacchi, T. H. Goodship, D. Kavanagh, M. Noris, M. Pickering, P. Sanchez-Corral, et al. 2015. Atypical aHUS: State of the art. *Mol. Immunol.* 67: 31–42.
11. Goicoechea de Jorge, E., C. L. Harris, J. Esparza-Gordillo, L. Carreras, E. A. Arranz, C. A. Garrido, M. López-Trascasa, P. Sánchez-Corral, B. P. Morgan, and S. Rodríguez de Córdoba. 2007. Gain-of-function mutations in complement factor B are associated with atypical hemolytic uremic syndrome. [Published erratum appears in 2007 *Proc. Natl. Acad. Sci. USA* 104: 10749.] *Proc. Natl. Acad. Sci. USA* 104: 240–245.
12. Noris, M., and G. Remuzzi. 2017. Genetics of immune-mediated glomerular diseases: focus on complement. *Semin. Nephrol.* 37: 447–463.
13. Valoti, E., M. Alberti, A. Tortajada, J. Garcia-Fernandez, S. Gastoldi, L. Besso, E. Bresin, G. Remuzzi, S. Rodríguez de Cordoba, and M. Noris. 2015. A novel atypical hemolytic uremic syndrome-associated hybrid CFHR1/CFH gene encoding a fusion protein that antagonizes factor H-dependent complement regulation. *J. Am. Soc. Nephrol.* 26: 209–219.
14. Eyler, S. J., N. C. Meyer, Y. Zhang, X. Xiao, C. M. Nester, and R. J. Smith. 2013. A novel hybrid CFHR1/CFH gene causes atypical hemolytic uremic syndrome. *Pediatr. Nephrol.* 28: 2221–2225.
15. Venables, J. P., L. Strain, D. Routledge, D. Bourn, H. M. Powell, P. Warwicker, M. L. Diaz-Torres, A. Sampson, P. Mead, M. Webb, et al. 2006. Atypical haemolytic uraemic syndrome associated with a hybrid complement gene. *PLoS Med.* 3: e431.
16. Maga, T. K., N. C. Meyer, C. Belsha, C. J. Nishimura, Y. Zhang, and R. J. Smith. 2011. A novel deletion in the RCA gene cluster causes atypical hemolytic uremic syndrome. *Nephrol. Dial. Transplant.* 26: 739–741.
17. Frémeaux-Bacchi, V., F. Fakhouri, A. Garnier, F. Bienaimé, M. A. Dragon-Durey, S. Ngo, B. Moulin, A. Servais, F. Provot, L. Rostaing, et al. 2013. Genetics and outcome of atypical hemolytic uremic syndrome: a nationwide French series comparing children and adults. *Clin. J. Am. Soc. Nephrol.* 8: 554–562.
18. Lemaire, M., V. Frémeaux-Bacchi, F. Schaefer, M. Choi, W. H. Tang, M. Le Quintrec, F. Fakhouri, S. Taque, F. Nobili, F. Martinez, et al. 2013. Recessive mutations in DGKE cause atypical hemolytic-uremic syndrome. *Nat. Genet.* 45: 531–536.
19. Sánchez Chinchilla, D., S. Pinto, B. Hoppe, M. Adragna, L. Lopez, M. L. Justa Roldan, A. Peña, M. Lopez Trascasa, P. Sánchez-Corral, and S. Rodríguez de Córdoba. 2014. Complement mutations in diacylglycerol kinase-ε-associated atypical hemolytic uremic syndrome. *Clin. J. Am. Soc. Nephrol.* 9: 1611–1619.
20. Servais, A., L. H. Noël, L. T. Roumenina, M. Le Quintrec, S. Ngo, M. A. Dragon-Durey, M. A. Macher, J. Zuber, A. Karras, F. Provot, et al. 2012. Acquired and genetic complement abnormalities play a critical role in dense deposit disease and other C3 glomerulopathies. *Kidney Int.* 82: 454–464.
21. Bu, F., N. G. Borsa, M. B. Jones, E. Takanami, C. Nishimura, J. J. Hauer, H. Azaiez, E. A. Black-Ziegelbein, N. C. Meyer, D. L. Kolbe, et al. 2016. High-throughput genetic testing for thrombotic microangiopathies and C3 glomerulopathies. *J. Am. Soc. Nephrol.* 27: 1245–1253.
22. Iatropoulos, P., M. Noris, C. Mele, R. Piras, E. Valoti, E. Bresin, M. Curreri, E. Mondo, A. Zito, S. Gamba, et al. 2016. Complement gene variants determine the risk of immunoglobulin-associated MPGN and C3 glomerulopathy and predict long-term renal outcome. *Mol. Immunol.* 71: 131–142.
23. Tortajada, A., H. Yébenes, C. Abarrategui-Garrido, J. Anter, J. M. García-Fernández, R. Martínez-Barricarte, M. Alba-Domínguez, T. H. Malik, R. Bedoya, R. Cabrera Pérez, et al. 2013. C3 glomerulopathy-associated CFHR1 mutation alters FHR oligomerization and complement regulation. *J. Clin. Invest.* 123: 2434–2446.
24. Gale, D. P., E. G. de Jorge, H. T. Cook, R. Martinez-Barricarte, A. Hadjisavvas, A. G. McLean, C. D. Pusey, A. Pierides, K. Kyriacou, Y. Athanasiou, et al. 2010. Identification of a mutation in complement factor H-related protein 5 in patients of Cypriot origin with glomerulonephritis. *Lancet* 376: 794–801.
25. Chen, Q., M. Wiesener, H. U. Eberhardt, A. Hartmann, B. Uzonyi, M. Kirschfink, K. Amann, M. Buettner, T. Goodship, C. Hugo, et al. 2014. Complement factor H-related hybrid protein deregulates complement in dense deposit disease. *J. Clin. Invest.* 124: 145–155.
26. Athanasiou, Y., K. Voskarides, D. P. Gale, L. Damianou, C. Patsias, M. Zavros, P. H. Maxwell, H. T. Cook, P. Demosthenous, A. Hadjisavvas, et al. 2011. Familial C3 glomerulopathy associated with CFHR5 mutations: clinical characteristics of 91 patients in 16 pedigrees. *Clin. J. Am. Soc. Nephrol.* 6: 1436–1446.
27. Levy, M., L. Halbwachs-Mecarelli, M. C. Gubler, G. Kohout, A. Bensenouci, P. Niaudet, G. Hauptmann, and P. Lesavre. 1986. H deficiency in two brothers with atypical dense intramembranous deposit disease. *Kidney Int.* 30: 949–956.
28. Ault, B. H., B. Z. Schmidt, N. L. Fowler, C. E. Kashtan, A. E. Ahmed, B. A. Vogt, and H. R. Colten. 1997. Human factor H deficiency. Mutations in framework cysteine residues and block in H protein secretion and intracellular catabolism. *J. Biol. Chem.* 272: 25168–25175.
29. Martínez-Barricarte, R., M. Heurich, F. Valdes-Cañedo, E. Vazquez-Martul, E. Torreira, T. Montes, A. Tortajada, S. Pinto, M. Lopez-Trascasa, B. P. Morgan, et al. 2010. 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.
30. Chauvet, S., L. T. Roumenina, S. Bruneau, M. C. Marinozzi, T. Rybkine, E. C. Schramm, A. Java, J. P. Atkinson, J. C. Aldigier, F. Bridoux, et al. 2016. A familial C3GN secondary to defective C3 regulation by complement receptor 1 and complement factor H. *J. Am. Soc. Nephrol.* 27: 1665–1677.
31. Thomas, S., D. Ranganathan, L. Francis, K. Madhan, and G. T. John. 2014. Current concepts in C3 glomerulopathy. *Indian J. Nephrol.* 24: 339–348.
32. Durey, M. A., A. Sinha, S. K. Togarsimalemath, and A. Bagga. 2016. Anti-complement-factor H-associated glomerulopathies. *Nat. Rev. Nephrol.* 12: 563–578.
33. Kouser, L., M. Abdul-Aziz, A. Nayak, C. M. Stover, R. B. Sim, and U. Kishore. 2013. Properdin and factor H: opposing players on the alternative complement pathway "see-saw". *Front. Immunol.* 4: 93.
34. Bu, F., T. Maga, N. C. Meyer, K. Wang, C. P. Thomas, C. M. Nester, and R. J. Smith. 2014. Comprehensive genetic analysis of complement and coagulation genes in atypical hemolytic uremic syndrome. *J. Am. Soc. Nephrol.* 25: 55–64.
35. Merinero, H. M., S. P. García, J. García-Fernández, E. Arjona, A. Tortajada, and S. Rodríguez de Córdoba. 2018. Complete functional characterization of disease-

associated genetic variants in the complement factor H gene. *Kidney Int.* 93: 470–481.
36. Bennett, C. A., S. Petrovski, K. L. Oliver, and S. F. Berkovic. 2017. ExACtly zero or once: a clinically helpful guide to assessing genetic variants in mild epilepsies. *Neurol. Genet.* 3: e163.
37. Lek, M., K. J. Karczewski, E. V. Minikel, K. E. Samocha, E. Banks, T. Fennell, A. H. O'Donnell-Luria, J. S. Ware, A. J. Hill, B. B. Cummings, et al; Exome Aggregation Consortium. 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536: 285–291.
38. 1000 Genomes Project Consortium, A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, H. M. Kang, J. O. Korbel, J. L. Marchini, S. McCarthy, G. A. McVean, and G. R. Abecasis. 2015. A global reference for human genetic variation. *Nature* 526: 68–74.
39. Song, W., S. A. Gardner, H. Hovhannisyan, A. Natalizio, K. S. Weymouth, W. Chen, I. Thibodeau, E. Bogdanova, S. Letovsky, A. Willis, and N. Nagan. 2016. Exploring the landscape of pathogenic genetic variation in the ExAC population database: insights of relevance to variant classification. *Genet. Med.* 18: 850–854.
40. Rallapalli, P. M., G. Kemball-Cook, E. G. Tuddenham, K. Gomez, and S. J. Perkins. 2013. An interactive mutation database for human coagulation factor IX provides novel insights into the phenotypes and genetics of hemophilia B. *J. Thromb. Haemost.* 11: 1329–1340.
41. Löytynoja, A., and N. Goldman. 2005. An algorithm for progressive multiple alignment of sequences with insertions. *Proc. Natl. Acad. Sci. USA* 102: 10557–10562.
42. Walsh, R., K. L. Thomson, J. S. Ware, B. H. Funke, J. Woodley, K. J. McGuire, F. Mazzarotto, E. Blair, A. Seller, J. C. Taylor, et al; Exome Aggregation Consortium. 2017. Reassessment of Mendelian gene pathogenicity using 7,855 cardiomyopathy cases and 60,706 reference samples. *Genet. Med.* 19: 192–203.
43. Kobayashi, Y., S. Yang, K. Nykamp, J. Garcia, S. E. Lincoln, and S. E. Topper. 2017. Pathogenic variant burden in the ExAC database: an empirical approach to evaluating population data for clinical variant interpretation. *Genome Med.* 9: 13.
44. Richards, S., N. Aziz, S. Bale, D. Bick, S. Das, J. Gastier-Foster, W. W. Grody, M. Hegde, E. Lyon, E. Spector, et al; ACMG Laboratory Quality Assurance Committee. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* 17: 405–424.
45. McLaren, W., L. Gil, S. E. Hunt, H. S. Riat, G. R. Ritchie, A. Thormann, P. Flicek, and F. Cunningham. 2016. The Ensembl variant effect predictor. *Genome Biol.* 17: 122.
46. Adzhubei, I. A., S. Schmidt, L. Peshkin, V. E. Ramensky, A. Gerasimova, P. Bork, A. S. Kondrashov, and S. R. Sunyaev. 2010. A method and server for predicting damaging missense mutations. *Nat. Methods* 7: 248–249.
47. Kumar, P., S. Henikoff, and P. C. Ng. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* 4: 1073–1081.
48. Okemefuna, A. I., R. Nan, J. Gor, and S. J. Perkins. 2009. Electrostatic interactions contribute to the folded-back conformation of wild type human factor H. *J. Mol. Biol.* 391: 98–118.
49. Janssen, B. J., E. G. Huizinga, H. C. Raaijmakers, A. Roos, M. R. Daha, K. Nilsson-Ekdahl, B. Nilsson, and P. Gros. 2005. Structures of complement component C3 provide insights into the function and evolution of immunity. *Nature* 437: 505–511.
50. Roversi, P., S. Johnson, J. J. Caesar, F. McLean, K. J. Leath, S. A. Tsiftsoglou, B. P. Morgan, C. L. Harris, R. B. Sim, and S. M. Lea. 2011. Structural basis for complement factor I control and its disease-associated sequence polymorphisms. *Proc. Natl. Acad. Sci. USA* 108: 12839–12844.
51. Persson, B. D., N. B. Schmitz, C. Santiago, G. Zocher, M. Larvie, U. Scheu, J. M. Casasnovas, and T. Stehle. 2010. Structure of the extracellular portion of CD46 provides insights into its interactions with complement proteins and pathogens. *PLoS Pathog.* 6: e1001122.
52. Milder, F. J., L. Gomes, A. Schouten, B. J. Janssen, E. G. Huizinga, R. A. Romijn, W. Hemrika, A. Roos, M. R. Daha, and P. Gros. 2007. Factor B structure provides insights into activation of the central protease of the complement system. *Nat. Struct. Mol. Biol.* 14: 224–228.
53. Schrodinger, LLC. 2015. *The PyMOL Molecular Graphics System*, Version 1.8.
54. Dupont, W. D., and W. D. Plummer, Jr. 1990. Power and sample size calculations. A review and computer program. *Control. Clin. Trials* 11: 116–128.
55. Bruel, A., D. Kavanagh, M. Noris, Y. Delmas, E. K. S. Wong, E. Bresin, F. Provôt, V. Brocklebank, C. Mele, G. Remuzzi, et al. 2017. Hemolytic uremic syndrome in pregnancy and postpartum. *Clin. J. Am. Soc. Nephrol.* 12: 1237–1247.
56. Huerta, A., E. Arjona, J. Portoles, P. Lopez-Sanchez, C. Rabasco, M. Espinosa, T. Cavero, M. Blasco, M. Cao, J. Manrique, et al. 2018. A retrospective study of pregnancy-associated atypical hemolytic uremic syndrome. *Kidney Int.* 93: 450–459.
57. Roumenina, L. T., M. Frimat, E. C. Miller, F. Provot, M. A. Dragon-Durey, P. Bordereau, S. Bigot, C. Hue, S. C. Satchell, P. W. Mathieson, et al. 2012. A prevalent C3 mutation in aHUS patients causes a direct C3 convertase gain of function. *Blood* 119: 4182–4191.
58. van den Heuvel, L., K. Riesbeck, O. El Tahir, V. Gracchi, M. Kremlitzka, S. A. Morré, A. M. van Furth, B. Singh, M. Okrój, N. van de Kar, et al. 2018. Genetic predisposition to infection in a case of atypical hemolytic uremic syndrome. *J. Hum. Genet.* 63: 93–96.
59. Ståhl, A. L., A. Kristoffersson, A. I. Olin, M. L. Olsson, A. M. Roodhooft, W. Proesmans, and D. Karpman. 2009. A novel mutation in the complement regulator clusterin in recurrent hemolytic uremic syndrome. *Mol. Immunol.* 46: 2236–2243.
60. Zuk, O., S. F. Schaffner, K. Samocha, R. Do, E. Hechter, S. Kathiresan, M. J. Daly, B. M. Neale, S. R. Sunyaev, and E. S. Lander. 2014. Searching for missing heritability: designing rare variant association studies. *Proc. Natl. Acad. Sci. USA* 111: E455–E464.
61. Gibson, G. 2012. Rare and common variants: twenty arguments. *Nat. Rev. Genet.* 13: 135–145.
62. Rallapalli, P. M., C. A. Orengo, R. A. Studer, and S. J. Perkins. 2014. Positive selection during the evolution of the blood coagulation factors in the context of their disease-causing mutations. *Mol. Biol. Evol.* 31: 3040–3056.
63. Auer, P. L., and G. Lettre. 2015. Rare variant association studies: considerations, challenges and opportunities. *Genome Med.* 7: 16.
64. Azzi, A., D. Boscoboinik, and C. Hensey. 1992. The protein kinase C family. *Eur. J. Biochem.* 208: 547–557.
65. Goicoechea de Jorge, E., J. J. Caesar, T. H. Malik, M. Patel, M. Colledge, S. Johnson, S. Hakobyan, B. P. Morgan, C. L. Harris, M. C. Pickering, and S. M. Lea. 2013. Dimerization of complement factor H-related proteins modulates complement activation in vivo. *Proc. Natl. Acad. Sci. USA* 110: 4685–4690.
66. Csincsi, A. I., A. Kopp, M. Zöldi, Z. Bánlaki, B. Uzonyi, M. Hebecker, J. J. Caesar, M. C. Pickering, K. Daigo, T. Hamakubo, et al. 2015. Factor H-related protein 5 interacts with pentraxin 3 and the extracellular matrix and modulates complement activation. *J. Immunol.* 194: 4963–4973.
67. Delvaeye, M., M. Noris, A. De Vriese, C. T. Esmon, N. L. Esmon, G. Ferrell, J. Del-Favero, S. Plaisance, B. Claes, D. Lambrechts, et al. 2009. Thrombomodulin mutations in atypical hemolytic-uremic syndrome. *N. Engl. J. Med.* 361: 345–357.
68. Heurich, M., R. J. Preston, V. B. O'Donnell, B. P. Morgan, and P. W. Collins. 2016. Thrombomodulin enhances complement regulation through strong affinity interactions with factor H and C3b-factor H complex. *Thromb. Res.* 145: 84–92.
69. Wang, H., I. Vinnikov, K. Shahzad, F. Bock, S. Ranjan, J. Wolter, M. Kashif, J. Oh, A. Bierhaus, P. Nawroth, et al. 2012. The lectin-like domain of thrombomodulin ameliorates diabetic glomerulopathy via complement inhibition. *Thromb. Haemost.* 108: 1141–1153.
70. Conway, E. M. 2012. Thrombomodulin and its role in inflammation. *Semin. Immunopathol.* 34: 107–125.
71. Rodríguez de Córdoba, S., J. Esparza-Gordillo, E. Goicoechea de Jorge, M. Lopez-Trascasa, and P. Sánchez-Corral. 2004. The human complement factor H: functional roles, genetic variations and disease associations. *Mol. Immunol.* 41: 355–367.
72. Schmidt, C. Q., A. P. Herbert, D. Kavanagh, C. Gandy, C. J. Fenton, B. S. Blaum, M. Lyon, D. Uhrín, and P. N. Barlow. 2008. A new map of glycosaminoglycan and C3b binding sites on factor H. *J. Immunol.* 181: 2610–2619.
73. Saunders, R. E., T. H. Goodship, P. F. Zipfel, and S. J. Perkins. 2006. An interactive web database of factor H-associated hemolytic uremic syndrome mutations: insights into the structural consequences of disease-associated mutations. *Hum. Mutat.* 27: 21–30.
74. Blaum, B. S., J. P. Hannan, A. P. Herbert, D. Kavanagh, D. Uhrín, and T. Stehle. 2015. Structural basis for sialic acid-mediated self-recognition by complement factor H. *Nat. Chem. Biol.* 11: 77–82.
75. Wu, J., Y. Q. Wu, D. Ricklin, B. J. Janssen, J. D. Lambris, and P. Gros. 2009. Structure of complement fragment C3b-factor H and implications for host protection by complement regulators. *Nat. Immunol.* 10: 728–733.
76. Schramm, E. C., L. T. Roumenina, T. Rybkine, S. Chauvet, P. Vieira-Martins, C. Hue, T. Maga, E. Valoti, V. Wilson, S. Jokiranta, et al. 2015. Mapping interactions between complement C3 and regulators using mutations in atypical hemolytic uremic syndrome. *Blood* 125: 2359–2369.