

Genetic and protein structural evaluation of atypical haemolytic uremic syndrome and C3 glomerulopathy

Stephen J. Perkins*

*Department of Structural and Molecular Biology, University College London, London, United Kingdom;

Running Title: Database of Complement Gene Variants

*Address correspondence to: Prof S. J. Perkins, Department of Structural and Molecular Biology, Darwin Building, University College London, Gower Street, London, WC1E 6BT, U.K. Tel: 020-76797048; FAX: 020-7679-7193; Email: s.perkins@ucl.ac.uk.

Keywords: allele frequency; atypical haemolytic uremic syndrome; C3 glomerulopathy; complement alternative pathway; Exome Aggregation Consortium; rare variant database.

Conflicts of Interest. The author has no conflict of interest to declare.

ABSTRACT

Atypical haemolytic uraemic syndrome (aHUS) and C3 glomerulopathy (C3G) are associated with loss of regulation of the alternative pathway of complement and its resulting over-activation. As rare diseases, genetic variants leading to aHUS and C3G were previously analysed in relatively low patient numbers. To improve this analysis, data were pooled from six centres. Totals of 610 rare variants for aHUS and 82 for C3G were presented in an interactive database for 13 genes. Using allele frequency comparisons with the Exome Aggregation Consortium (ExAC) as a reference genome, the aHUS patients showed significantly more protein-altering ultra-rare variants (allele frequency <0.01%) in five genes *CFH*, *CFI*, *CD46*, *C3* and *DGKE*. In C3G patients, the corresponding association was only found for *C3* and *CFH*. Protein structure analyses of these five proteins showed distinct differences in the positioning of these variants in C3 and FH. For aHUS, variants were clustered at the C-terminus of FH and implicated changes in the binding of FH to host cell surfaces. For C3G, variants were clustered at the N-terminal C3b binding site of FH, and implicated changes in the fluid-phase regulation of C3b. We discuss the utility of the web-database as a patient resource for clinicians.

200/200 words

CLINICAL SUMMARY

- An interactive public web database is presented at <https://www.complement-db.org/> for 610 rare genetic variants in 13 mostly complement genes from >3500 patients with aHUS and C3G.
- The database showed that genetic variants in *CFH*, *CFI*, *CD46*, *C3* and *DGKE* most often lead to aHUS, and those in *C3* and *CFH* most often lead to C3G.
- Protein three-dimensional structural analyses explained the occurrence of these two distinct diseases in terms of localised hot spots in functional regions of these proteins.
- This web resource provides clinicians with an improved tool for patient diagnosis and management.

INTRODUCTION

Atypical haemolytic uraemic syndrome (aHUS) and C3 glomerulopathy (C3G) are rare kidney diseases with severe consequences. Both diseases involve a failure to regulate the alternative pathway in the complement system of innate immunity. In terms of the onset of both diseases, both familial (85%) and sporadic forms of aHUS and C3G are observed. Rare genetic abnormalities in the alternative pathway and thrombosis-related genes are present in ~60% of aHUS cases,¹⁻³ and in ~20% of C3G patients.⁴ Anti-CFH autoantibodies are involved in 5-13% of cases.^{4,5} The penetrance of a disease-causing mutation is the proportion of individuals with the mutation who exhibit clinical symptoms. The penetrance of genetic aHUS is ~50% and is determined by additional rare variants, the *CFH* and *CD46* haplotype, and a trigger.^{4,6-8}

Most aHUS cases are attributed to genetic variants in the complement proteins, namely *CFH* (25-30% of cases), followed by *CD46* (8-10%), *CFI* and *C3* (4-8% each) and *CFB* (1-4%).^{4,9} These genes correspond to CFH, CD46, and CFI which are complement regulatory proteins while C3 and CFB are complement activators. Rare variants in thrombomodulin (*THBD*) additionally account for 3-4% of aHUS,⁴ however none were detected in the French aHUS cohort.¹⁰ Rare variants in diacylglycerol kinase epsilon (*DGKE*) account for ~27% of aHUS presenting under the age of one year, and <4% under two years.^{11,12} Rare copy number variation (CNV) in the *CFH-CFHR* region, leading to rearrangements such as the *CFH/CFHR1* and *CFH/CFHR3* hybrid genes, have been identified in aHUS.¹³⁻¹⁶ In contrast, familial C3G differs in that this is most often linked to highly penetrant heterozygous CNV in the *CFHR1-5* genes, as well as homozygous *CFH* deficiency and heterozygous gain-of-function mutation in *C3*.¹⁷⁻²⁴ Other rare complement gene variants have been identified in ~20% of sporadic cases of C3G.²⁵⁻²⁷ For both aHUS and C3G, unaffected carriers of these rare variants are seen, meaning that the genetic variation only predisposes for the manifestation of both diseases.²⁸

In relation to these rare variants and genes in complement, their link with aHUS and C3G involve two issues. First, prior to our recent work, the rare variants were not known to be significantly more common in large patient populations than in large reference populations. To clarify a robust association with these diseases, we have statistically analysed a comparatively large number of 610 rare genetic variants from 3571 patients in six renal centres.²⁹ The results are presented in an interactive web database. The resulting comparison with 60,706 genomic reference sequences from the Exome Aggregation Consortium (ExAC)³⁰ enabled the rare variant allele frequency, the gene-based rare variant burden, and mutational hotspots to be assessed. Second, the effects of genetic variants on protein function are often revealed by mapping the location of the variants onto the three-dimensional structure of the protein in which they occur. Such a comparison is available from the interactive database since its inception, but is now updated by the inclusion of more recent protein structures.^{29,31,32} This reveals the functional significance of the mutational hotspots in the disease-related proteins and the molecular mechanisms of aHUS and C3G. The interactive database thus provides two new insights into patient management, and a simplified account of both these are provided here. In addition, we also illustrate how the database can be used to analyse variants, including new ones of unknown significance.

CONCISE METHODS

The Database of Complement Gene Variants is available at <https://www.complement-db.org> and includes variant search tools, allele frequency analyses, protein structural views, species conservation of residues, predictions of protein damage caused by missense variants (PolyPhen-2 and SIFT), and literature searches.²⁹ The aHUS and C3G variant data for 13 genes

were sourced from six centres and tested for duplicates. The rare variant burden per gene was computed for the aHUS, C3G and ExAC datasets for the nine genes.

ALLELE FREQUENCIES IN AHUS AND C3G

Specifically the allele frequency is the fraction of all chromosomes in the population that carry that allele. In order to amass sufficient variant data for statistical analysis, a total of 610 variants were identified in 13 genes from 3128 aHUS and 443 C3G patients based in six laboratories in six countries.²⁹ The 13 genes comprised ten complement ones (*CFH*, *CFI*, *C3*, *CD46*, *CFB*, *CFHR1*, *CFHR3*, *CFHR4*, *CFHR5*, *CFP*), and three non-complement ones (*THBD*, *DGKE*, *PLG*). When compared with three reference datasets (the 1000 Genomes Project, the Exome Variant Server, and ExAC), 610 variants in the 13 genes were identified with an allele frequency (AF) of <1%, showing that these were indeed rare variants (Figure 1A). A further 14 variants were found to be more common with allele frequencies of $\geq 1\%$ in at least one reference dataset, or were rare copy number variants not covered by the reference datasets. In the aHUS dataset, 32% (174 of 542) of the rare variants, when compared with ExAC as reference, showed an allele frequency $>0\%$ (orange, yellow, green, Figure 1A). This proportion was higher in the C3G dataset, being 58% (67 of 110) (orange, yellow, green, Figure 1A). Overall, 542 and 110 rare variants were identified for aHUS and C3G respectively, with an overlap of 42 variants between both. This total of 610 genetic variants²⁹ (Osborne et al 2018) showed a continued increase when compared to 12 variants originally reported for *CFH* in aHUS in 2002,³³ then 167 variants for *CFH*, *MCP* and *CFI* in aHUS in 2006,³² and 324 variants in *CFH*, *MCP*, *CFI* and *C3* in aHUS in 2014.³¹

Analyses of the variant datasets from these 3128 aHUS and 443 C3G patients showed that most rare variants corresponded to aHUS cases (1231; 91%) when compared to C3G (116; 9%).²⁹ Note that the extent of genetic screening differed between the six laboratories that contributed to this project. For example, for aHUS, *CFH* was screened in 3128 patients, *CD46* in 2942 and *CFI* in 2923, while for C3G, *CFH* was screened in 443 patients, *CD46* in 406, and *CFI* in 408. Overall there were 0.44 unique rare variants per case in aHUS, compared to 0.95 in C3G. This suggested that aHUS cases were more likely to share the same rare variant, whereas almost every C3G case had a different, unique rare variant. In the 1231 aHUS and 116 C3G cases, 1024 and 82 patients respectively had a single rare variant on one of the 13 genes. Two rare variants were identified in 182 aHUS and 31 C3G cases, and three in 24 and 3 respectively. A small bias towards females was noticed for aHUS cases but not for C3G cases, but the removal of *CFH* variants from this dataset removed this aHUS gender bias.

The pathogenicity of each rare variant was typically evaluated using (i) the reference allele frequency, (ii) protein structural predictions and (iii) published experimental functional studies.²⁹ Because the average healthy genome contains benign rare variants that occur at similar allele frequencies to disease-related rare variants, allele frequencies alone cannot be used as evidence for pathogenicity. Because only gain-of-function variants in the activators C3 and CFB will predispose for aHUS or C3G, a damaging and over-activated C3 or CFB variant will result from the loss of interaction with an inhibitory complement regulator FH, FI or MCP. This effect is distinct from FH, FI or MCP variants where the ability of these proteins to interact with C3 or CFB to regulate these is damaged, leaving C3 and CFB to become more active. This difference prevents predictive software from predicting gain-of-function phenotypes, thus the significance of many C3 and CFB variants is unclear. Thus, for aHUS, the majority of rare variants in *CFH*, *CFI*, *CD46* and *DGKE* were predicted as either ‘pathogenic’ or ‘likely pathogenic’ for reason of their predicted loss-of-function phenotypes. In C3G, the majority of rare variants in all genes were predicted to be of ‘uncertain significance’ except for *CFH*. The

identification of ‘likely pathogenic’ rare variants will require further functional studies to confirm or disprove the disease relevance of the variants.

In aHUS and C3G, most rare variants were missense ones, in which one amino acid is replaced by another. For example, of the 216 rare unique FH variants, 64% were missense. In C3G, of the 32 rare unique FH variants, 75% were missense. Only the variants that were not classified as ‘benign’ or ‘likely benign’ by protein structural predictions were analysed here. In aHUS with ~490 such variants, four proteins occurred the most frequently. *CFH* showed the most rare variants with 204, followed by *CFI* with 82, *CD46* with 79, and *C3* with 68. This outcome replicated our 2014 analyses for *CFH*, *CFI*, *CD46* and *C3*.³¹ Lesser abundances for aHUS were seen for *CFB* with 20, *DGKE* with 20, *THBD* with 11, and *PLG*, *CFHR5*, *CFHR1*, *CFHR3* and *CFP* with 2-4 variants in each. The most frequent variant in our aHUS dataset was *C3* p.Arg161Trp with an allele frequency of 1.16% in 52 aHUS patients. In C3G with ~85 variants, only two proteins, *C3* with 31 and *CFH* with 25, showed the most rare variants. Other genes showed much reduced abundances for C3G in *CFI* (5), *CD46* (1), *CFB* (8), *DGKE* (2), *THBD* (2), *PLG* (2), *CFHR5* (1), *CFHR1* (1), and *CFHR3* (1). For C3G, none of the rare variants were notably more frequent than others.

Allele frequency differences between aHUS and C3G relate to their different phenotypes. Significantly greater proportions of C3G rare variants, unlike aHUS, were identified in the reference datasets, with allele frequencies between 0-1% rather than 0%, (Figure 1A). This meant that our C3G rare variants occurred more frequently in individuals without C3G (the reference datasets) than for our aHUS rare variants. Allele frequency differences also revealed differences between the most common rare variants in the aHUS and C3G datasets that were likely to reflect their different phenotypes.

The allele frequency cut-off to employ when using the database depends on the context in which it is used, and can be adjusted in the database. Our allele frequency analyses verified the rarity of 97% of the aHUS and C3G variants when compared to the ExAC reference (Figure 1A), especially given the ability of ExAC to resolve ultra-rare variant allele frequencies as low as 1×10^{-5} (0.001%).³⁴ This result employed an allele frequency <1% in the reference datasets. A more stringent rare variant allele frequency cut-off of 0.01% was used for the rare variant burden calculations in order to restrict them to rare variants with a higher confidence of pathogenicity. However, rare variants observed in both aHUS and C3G have reduced penetrance, and disease-free individuals also harbour these pathogenic rare variants. For these, an intermediate reference allele frequency cut-off of <0.1% was thus applicable when required.

RARE VARIANT BURDEN TESTS IN AHUS AND C3G

The rare variant burden is the proportion of screened cases with an identified protein-altering rare variant for which the ExAC allele frequency was <0.01%. As opposed to the allele frequency analyses that look at each variant one-by-one, the rare variant burden now provides insight into all the rare variants associated with each gene. In general, rare variant burden tests assume that all tested rare variants influence the phenotype in the same direction.³⁵ We therefore separated rare variants into ‘truncating’ (loss-of-function) and ‘non-truncating’ (either loss- or gain-of-function, or neutral). Rare variant burden tests showed clear differences between aHUS and C3G when compared to reference datasets. Prior knowledge of experimental functional characterization indicated that aHUS is more related to surface alternative pathway dysregulation, while C3G is more related to fluid phase alternative pathway dysregulation.

Rare variant burden tests confirmed that the rare variation seen in the genes of aHUS and C3G patients was greater than in the genes of disease-free individuals, thus associating these rare variants with these diseases.²⁹ The tests were made possible by the (i) sufficient availability of 610 rare variants in 13 genes from 3128 aHUS and 443 C3G patients, together with (ii) the corresponding information from reference datasets such as that of ExAC. Because not all the required information was available for the 13 genes, only nine could be analysed (*CFH*, *CFI*, *CD46*, *C3*, *DGKE*, *CFB*, *CFHR5*, *PLG*, *THBD*). For aHUS, a significantly greater burden of rare variation was revealed in patients than in the ExAC dataset for six genes, namely *CFH*, *CFI*, *CD46*, *C3*, *CFB* and *DGKE*, and no such association was seen for rare variants in *THBD*, *PLG* and *CFHR5*. For the five genes of the alternative pathway, *CFH*, *CFI*, *C3* and *CFB* are functionally involved in both cell surface and fluid phase regulation, but *CD46* is only involved in cell surface regulation. Therefore, the rare variant burden analyses suggested that aHUS involved defects that result in both cell surface and fluid phase dysregulation. For C3G, just *C3* and *CFH* showed a significantly greater burden of protein-altering rare variation in patients than in the ExAC dataset. 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. These comparisons were based on a minor allelic frequency cutoff of <0.01% per gene, thus giving a higher confidence of pathogenicity in these tests.

LOCATION OF THE RARE VARIANTS IN THE COMPLEMENT PROTEINS

The knowledge of the domains in the sequences in each of FH, FI, MCP, C3 and FB (Figure 2), plus known crystal structures for the individual proteins and several of the co-complexes formed by the interacting proteins, provides key molecular insights into the causes of aHUS and C3G when combined with the known missense variants.²⁹ The distribution of the missense variants within each protein structure showed whether or not mutational hotspots that possess increased numbers of variants occurred, or whether the variants were spread out more evenly across the protein. However, the consequence of each rare variant on complement protein function is not always straightforward to interpret.

For FH, the most aHUS variants (currently 51) dominate in the C-terminal SCR-20 domain (Figure 1B; defined in Figure 2), compared to the other 19 SCR domains, and this result has been known since 2006.^{31,32} For aHUS, the total missense allele frequency of 2.03% in SCR-20 was the highest of all 20 domains. No notable localisation of variants occurred in the other FH domains, although more were seen in the N-terminal four domains SCR-1/4 (i.e. SCR-1 through SCR-4), and in the C-terminal domains SCR-14/19. This varied localisation of variants correlates well with the functional association of the C3b binding site at SCR-1/3 (which correlate with FH binding to C3b at SCR-1/4) and SCR-20 with crystal structures of C3d and sialic acid binding to SCR-19/20 and cell surface dysregulation of FH binding.^{36,37}

For FH, the C3G rare missense variants were clustered at the N-terminal C3b binding site at SCR-1/3, and also in the non-surface associated C-terminal domain SCR-18 (Figure 1B). In contrast to aHUS, rare missense C3G variants were not found in twelve SCR domains (Figure 1B). For C3G, SCR-2 showed the greatest frequency of missense rare variants. It is noteworthy that p.Cys431Tyr (SCR-9) and p.Cys1218Arg (SCR-20) were seen in both aHUS and C3G patients. Overall, the distribution of FH hotspots between aHUS and C3G showed clear differences between the two diseases.

For C3, the aHUS C3 missense rare variants occurred in 9 of the 13 C3 domains (pink; Figure 1C). The MG2 domain (MG: macroglobulin) showed the highest missense allele frequency for C3 (1.3%), normalised by the proportion of domain residues, followed by the

MG6 domain (0.5%). Both the MG2 and MG6 domains were thus aHUS hotspots. The 29 C3G missense rare variants occurred in 9 of the 13 C3 domains. In contrast to aHUS, these were spread more evenly throughout C3 and no hotspots were inferred (blue; Figure 1C). No known aHUS or C3G variants involved the crucial C3 thioester residues at the active site in the TED domain that form covalent links with the immune targets of C3 (Glu991, Cys998, His1104 and Glu1106).

For FI and CD46, the missense rare variants in aHUS were distributed across all five domains of FI and all four CD46 domains (Figure 2). No mutational hotspots were evident for either aHUS or C3G in both proteins.

STRUCTURAL ANALYSES HELP RATIONALISE aHUS AND C3G VARIANTS

The effect of predisposing variants in aHUS and C3G involved assessments of their impacts on protein function and structure. From the above, for aHUS and C3G, the most common variants involved non-truncating, missense changes in the regulators FH, FI and MCP, and the activators C3 and FB. Their molecular correlations with aHUS and C3G were assessed by mapping the variants onto three-dimensional structures for unbound FH, FI, MCP, C3, and FB.²⁹ For FH, the missense rare variants for aHUS and C3G were initially mapped onto protein structural models from atomistic scattering modelling³⁸, there being no crystal structure for full-length FH, while crystal structures were available for C3,³⁹ FI,⁴⁰ and MCP.⁴¹ Compared to previous FH modelling,³⁸ the more recent FH modelling started from 17 high resolution SCR structures and advanced atomistic modelling to generate full-length FH structures, including the eight FH glycan chains.⁴² Two distinct FH molecular structures were determined that showed either an extended N-terminal domain with a folded-back C-terminus, or an extended C-terminus and folded-back N-terminus. The extended N-terminal conformation accounted for C3b fluid phase regulation (Figure 3A), the extended C-terminal conformation accounted for C3d binding and cell surface regulation (Figure 3B), and both conformations accounted for bivalent FH binding to anionic saccharides at target cell surfaces. These two new full-length FH structures are displayed superimposed onto crystal structures for fragments of FH in complexes with its C3b-FI and C3d ligands (Figure 3A,B).^{43,44}

Residues involved in molecular interactions are mainly surface-associated whereas residues involved in protein structural stabilisation are typically buried. Residue accessibilities were calculated from the new modelled FH structures.⁴⁵ Of the 1213 residues in FH, 1171 (97%) are surface accessible and exposed to solvent, while just 42 are surface inaccessible. This outcome resulted from the most elongated FH structure which is much exposed to solvent. Of a total of 107 FH rare variants associated with aHUS, 98 involved surface accessible residues, and 9 involved buried residues. For a total of 17 FH rare variants associated with C3G, 16 involved surface accessible residues, and 1 was buried. For FH, 128 aHUS missense rare variants mostly affected three types of residues, namely 13 case alleles in the 42 buried residues (31%), 290 case alleles in the 1045 non-binding surface-accessible residues (28%), and 12 case alleles in the 92 residues at C3b-binding interfaces (13%). None were found at the FH-FI interface. In contrast, for 19 C3G variants in FH, these three types of residues totalled 2% (one case allele), 3% (26 case alleles) and 5% (five case alleles) respectively, plus two case alleles (6%) at the FH-FI binding interface. SCR-20 with 31 rare variants and a rare variant density of 37.2% was a notable missense hotspot. This is well explained by the involvement of SCR-20 in FH binding to surfaces, leading to host cell damage from excess complement activation caused by unregulated C3b. aHUS missense rare variants were also identified in the remaining 19 SCR domains in FH. However several FH domains (SCR-5, SCR-8, SCR-12, and SCR-13) do not correspond to known FH binding sites and have only single missense rare

variants. Inspection of these best-fit FH structures confirmed indeed that SCR-5 and SCR-12/13 were sparsely populated and distant from any functional sites (Figure 3A,B).

While many rare variants were spatially clustered near functionally-relevant binding sites, the existence of variants that are located distant from these functional regions implies that another molecular mechanism is responsible for damaging the FH protein structure. A prominent group of rare variants turned out to be buried Cys residues involved in disulphide bridges. The disruption of a structural disulphide bond often leads to loss of function via protein misfolding, thus these events are predicted by the occurrence of missense rare variants in Cys residues. Only those in the regulators FH, MCP and FI were associated with aHUS, while only those in FH and C3 were associated with C3G. The SCR domain is the most abundant domain type in complement, occurring in FH, MCP and FB, and stabilised by two disulphide bridges within two pairs of conserved Cys residues. Of the 96 Cys residues in FH, 29 (30%) are Cys residues that were replaced in aHUS variants, and 3 more in C3G variants. By creating a consensus view of the rare variants superimposed onto a typical SCR domain, it is seen that many rare variants involve all four Cys positions (purple, Figure 4) compared to other rare variants, and therefore these rare variants would cause disease by misfolding of the SCR domain. Misfolding is expected either to destabilise its folded protein structure through the loss of the disulphide bridge or, in the case of surface-accessible residues, alters the arrangement of its folded-back SCR domain structure by modifying the linker conformation between two adjacent SCR domains. This interpretation would account for the rare variants seen outside the functionally-important binding residues in FH (Figure 4). Previously, for the SCR consensus domain for FH and MCP, the so-called “hypervariable loop” in SCR domains was suggested to have a functional association involving aHUS and C3G disease variants.³¹ Given the increased number of available variants for analysis,²⁹ our consensus SCR domain showed relatively fewer 24 aHUS variants (13%) in a total of 192 found in FH and MCP, plus two more in C3G variants. The comparison with the 30% of Cys variants noted above shows that the “hypervariable loop” is less significant than previously thought (Figure 3A,B).

In the protein core of C3, interestingly, the TED and MG2 domains showed more aHUS variants (pink, Figure 1C) while MG5, MG6 and CUB β showed more C3G variants (blue, Figure 1C). In the crystal structure of C3b, the aHUS and C3G variants in C3 showed different spatial distributions (Figure 3C), with those for C3G being located further away from the FH-C3b interface, unlike those for aHUS which are seen at this interface. The MG2 and MG6 domains in C3b have the highest rare variant density. Both MG domains interact with FH SCR-2 and SCR-3 to enable C3 regulation by FH in both the fluid phase and on the cell surface. The disruption of the MG2 and MG6 domains would reduce C3 regulation by FH. Interestingly, while the TED domain contained 22 rare variants, its rare variant density, which considers 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. For FI, the aHUS variants were distributed evenly throughout FI, and not just at the interface between FI and FH (Figure 3A). For CD46, the aHUS variants were likewise distributed throughout the protein (Figure 2), but C3G variants were almost all missing in MCP.

UTILITY OF THE aHUS AND C3G WEB SITES

In order to analyse genetic variants identified in aHUS and other complement diseases, an interactive FH-HUS web-database (<http://www.fh-hus.org>) was originally set up at University College London in 2006 for both clinicians and researchers.³² By 2014, 193 *CFH*, 130 *CFI*, 86 *CD46* and 64 *C3* variants were reported for aHUS and other complement diseases.³¹ However, major omissions from the 2006 FH-HUS database included not only new

complement genes and variants associated with aHUS and related complement diseases, but also the inclusion of variant allele frequency data. The recent Database of Complement Gene Variants has incorporated these new variant and allele frequency data, which were used to verify the association of rare variants in the complement and related genes with aHUS and C3G, as well as more recent crystal structures.

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. The six-centre analysis of 610 rare genetic variants in 13 mostly complement genes provided data on 3571 aHUS and C3G patients. This resulted in 371 novel rare variants for aHUS and 82 for C3G. The resulting Database of Complement Gene Variants was used to extract allele frequency data for these 13 genes using the ExAC server as the reference genome, and presented for users. Because the renal clinics are based in Western Europe and the US, the effect of ethnicities not well represented in ExAC will be minimal. For aHUS, significantly more protein-altering rare variation was found in five genes *CFH*, *CFI*, *CD46*, *C3* and *DGKE* than in ExAC (allele frequency <0.01%), thus correlating these with aHUS. For C3G, an association was only found for rare variants in *C3* and the N-terminal C3b-binding or C-terminal non-surface-associated regions of *CFH*. The rare variant analyses showed non-random distributions over the affected proteins, and different distributions were observed between aHUS and C3G that clarified their phenotypes. Importantly, unlike previously, these results were able to clarify how changes in the same biochemical pathway and/or proteins result in the different pathologies observed with aHUS and C3G. Furthermore, this work has reduced the earlier knowledge gap in the genetics and genotype-phenotype correlations of C3G to bring these closer to that of aHUS.⁴⁶

In conclusion, the web Database of Complement Gene Variants enhances the understanding of rare genetic variants in aHUS and C3G for clinical applications. We illustrate how the database can be used by clinical nephrologists. For variants that are already in the database, new improvements include the use of a user-selected ExAC allele frequency cut-off in the search tools (Figure 5), the display of allele frequency data for the disease and reference datasets, predictive comparisons of wild-type and mutant amino acids, protein predictions using PolyPhen-2 and SIFT in application to a damaging variant Cys431Tyr in SCR-7, which is visualised (Figure 6), examination of evolution-conserved residues across species, and correlations with functional binding sites. These tools enable clinicians to assess rare variants in disease, for example, to investigate which variants within these genes conferred predisposition to aHUS and C3G, and to identify mutational hotspots within these protein structures. For new variants of unknown significance, for which no experimental data exist, it is possible to identify its sequence location in the protein via the “New variant” option under the “Variants” tab, together with an assessment of its likely effect on protein function, and its location viewed in a protein structure model. Through the use of allele frequencies and burden testing in the database, the new database informs patient management by enabling clinicians to interpret known variants in terms of their associations with aHUS and C3G. The differences between aHUS and C3G reflect their different pathologies. The structural location of a newly-discovered variant may clarify the occurrence of aHUS or C3G in patients.

ACKNOWLEDGMENTS

This work was supported in part by Complement UK and Alexion through a grant to SJP. The author thanks Dr Amy J. Osborne, Dr Daniel P. Gale and Professor Timothy H. J. Goodship for collaborations and many useful discussions, and our clinical collaborators for generously contributing their variant data from patients.

REFERENCES

1. Bu F, Borsa N, Gianluigi A, Smith RJ. Familial atypical hemolytic uremic syndrome: a review of its genetic and clinical aspects. *Clin Dev Immunol* 2012;370426.
2. Warwicker P, Goodship TH, Donne RL, Pirson Y, Nicholls A, Ward RM, Turnpenny P, Goodship JA. Genetic studies into inherited and sporadic hemolytic uremic syndrome. *Kidney Int* 1998;53(4):836-844.
3. Fakhouri F, Zuber J, Fremeaux-Bacchi V, Loirat C. Haemolytic uraemic syndrome. *Lancet* 2017; 390(10095):681-696.
4. Noris M, Remuzzi G. Glomerular diseases dependent on complement activation, including atypical hemolytic uremic syndrome, membranoproliferative glomerulonephritis, and C3 glomerulopathy: core curriculum. *Am J Kidney Dis* 2015; 66(2):359-375.
5. Durey MA, Sinha A, Togarsimalemath SK, Bagga A. Anti-complement-factor H-associated glomerulopathies. *Nat Rev Nephrol* 2016;12(9):563-578.
6. Sansbury FH, Cordell HJ, Bingham C, Bromilow G, Nicholls A, Powell R, Shields B, Smyth L, Warwicker P, Strain L, Wilson V, Goodship JA, Goodship TH, Turnpenny PD. Factors determining penetrance in familial atypical haemolytic uraemic syndrome. *J Med Genet* 2014;51(11):756-764.
7. Esparza-Gordillo J, Goicoechea de Jorge E, Buil A, Carreras Berges L, Lopez-Trascasa M, Sanchez-Corral P, Rodriguez de Cordoba S. 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* 2005;14(5):703-712.
8. Bresin E, Rurali E, Caprioli J, Sanchez-Corral P, Fremeaux-Bacchi V, Rodriguez de Cordoba S, Pinto S, Goodship TH, Alberti M, Ribes D, Valoti E, Remuzzi G, Noris M, European Working Party on Complement Genetics in Renal Diseases. Combined complement gene mutations in atypical hemolytic uremic syndrome influence clinical phenotype. *J Am Soc Nephrol* 2013;24(3):475-486.
9. Goicoechea de Jorge E, Harris CL, Esparza-Gordillo J, Carreras L, Arranz EA, Garrido CA, Lopez-Trascasa M, Sanchez-Corral P, Morgan BP, Rodriguez de Cordoba S. Gain-of-function mutations in complement factor B are associated with atypical hemolytic uremic syndrome. *Proc Natl Acad Sci USA* 2007;104(1):240-245.
10. Fremeaux-Bacchi V, Fakhouri F, Garnier A, Bienaime F, Dragon-Durey MA, Ngo S, Moulin B, Servais A, Provot F, Rostaing L, Burtey S, Niaudet P, Deschenes G, Lebranchu Y, Zuber J, Loirat C. Genetics and outcome of atypical hemolytic uremic syndrome: a nationwide French series comparing children and adults. *Clin J Am Soc Nephrol* 2013;8(4):554-562.
11. Lemaire M, Fremeaux-Bacchi V, Schaefer F, Choi M, Tang WH, Le Quintrec M, Fakhouri F, Taque S, Nobili F, Martinez F, Ji W, Overton JD, Mane SM, Nurnberg G, Altmuller J, Thiele H, Morin D, Deschenes G, Baudouin V, Llanas B, Collard L, Majid MA, Simkova E, Nurnberg P, Rioux-Leclerc N, Moeckel GW, Gubler MC, Hwa J, Loirat C, Lifton RP. Recessive mutations in DGKE cause atypical hemolytic-uremic syndrome. *Nat Genet* 2013;45(5):531-536.
12. Sanchez Chinchilla D, Pinto S, Hoppe B, Adragna M, Lopez L, Justa Roldan ML, Pena A, Lopez Trascasa M, Sanchez-Corral P, Rodriguez de Cordoba S. Complement mutations in diacylglycerol kinase-epsilon-associated atypical hemolytic uremic syndrome. *Clin J Am Soc Nephrol* 2014;9(9):1611-1619.
13. Valoti E, Alberti M, Tortajada A, Garcia-Fernandez J, Gastoldi S, Besso L, Bresin E, Remuzzi G, Rodriguez de Cordoba S, Noris M. 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* 2015;26(1):209-219.
14. Eyler SJ, Meyer NC, Zhang Y, Xiao X, Nester CM, Smith RJ. A novel hybrid CFHR1/CFH gene causes atypical hemolytic uremic syndrome. *Pediatr Nephrol* 2013;28(11):2221-2225.
 15. Venables JP, Strain L, Routledge D, Bourn D, Powell HM, Warwicker P, Diaz-Torres ML, Sampson A, Mead P, Webb M, Pirson Y, Jackson MS, Hughes A, Wood KM, Goodship JA, Goodship TH. Atypical haemolytic uraemic syndrome associated with a hybrid complement gene. *PLoS Med* 2006;3(10):e431.
 16. Maga TK, Meyer NC, Belsha C, Nishimura CJ, Zhang Y, Smith RJ. A novel deletion in the RCA gene cluster causes atypical hemolytic uremic syndrome. *Nephrol Dial Transplant* 2011;26(2):739-741.
 17. Tortajada A, Yebenes H, Abarrategui-Garrido C, Anter J, Garcia-Fernandez JM, Martinez-Barricarte R, Alba-Dominguez M, Malik TH, Bedoya R, Cabrera Perez R, Lopez Trascasa M, Pickering MC, Harris CL, Sanchez-Corral P, Llorca O, Rodriguez de Cordoba S. C3 glomerulopathy-associated CFHR1 mutation alters FHR oligomerization and complement regulation. *J Clin Invest* 2013;123(6):2434-2446.
 18. Gale DP, de Jorge EG, Cook HT, Martinez-Barricarte R, Hadjisavvas A, McLean AG, Pusey CD, Pierides A, Kyriacou K, Athanasiou Y, Voskarides K, Deltas C, Palmer A, Fremeaux-Bacchi V, de Cordoba SR, Maxwell PH, Pickering MC. Identification of a mutation in complement factor H-related protein 5 in patients of Cypriot origin with glomerulonephritis. *Lancet* 2010;376(9743):794-801.
 19. Chen Q, Wiesener M, Eberhardt HU, Hartmann A, Uzonyi B, Kirschfink M, Amann K, Buettner M, Goodship T, Hugo C, Skerka C, Zipfel PF. Complement factor H-related hybrid protein deregulates complement in dense deposit disease. *J Clin Invest* 2014;124(1):145-155.
 20. Athanasiou Y, Voskarides K, Gale DP, Damianou L, Patsias C, Zavros M, Maxwell PH, Cook HT, Demosthenous P, Hadjisavvas A, Kyriacou K, Zouvani I, Pierides A, Deltas C. Familial C3 glomerulopathy associated with CFHR5 mutations: clinical characteristics of 91 patients in 16 pedigrees. *Clin J Am Soc Nephrol* 2011;6(6):1436-1446.
 21. Levy M, Halbwachs-Mecarelli L, Gubler MC, Kohout G, Bensenouci A, Niaudet P, Hauptmann G, Lesavre P. H deficiency in two brothers with atypical dense intramembranous deposit disease. *Kidney Int* 1986;30(6):949-956.
 22. Ault BH, Schmidt BZ, Fowler NL, Kashtan CE, Ahmed AE, Vogt BA, Colten HR. Human factor H deficiency. Mutations in framework cysteine residues and block in H protein secretion and intracellular catabolism. *J Biol Chem* 1997;272(40):25168-25175.
 23. Martinez-Barricarte R, Heurich M, Valdes-Canedo F, Vazquez-Martul E, Torreira E, Montes T, Tortajada A, Pinto S, Lopez-Trascasa M, Morgan BP, Llorca O, Harris CL, Rodriguez de Cordoba S. Human C3 mutation reveals a mechanism of dense deposit disease pathogenesis and provides insights into complement activation and regulation. *J Clin Invest* 2010;120(10):3702-3712.
 24. Chauvet S, Roumenina LT, Bruneau S, Marinozzi MC, Rybkine T, Schramm EC, Java A, Atkinson JP, Aldigier JC, Bridoux F, Touchard G, Fremeaux-Bacchi V. A familial C3GN secondary to defective C3 regulation by complement receptor 1 and complement factor H. *J Am Soc Nephrol* 2016;27(6):1665-1677.
 25. Servais A, Noel LH, Roumenina LT, Le Quintrec M, Ngo S, Dragon-Durey MA, Macher MA, Zuber J, Karras A, Provot F, Moulin B, Grunfeld JP, Niaudet P, Lesavre P, Fremeaux-Bacchi V. Acquired and genetic complement abnormalities play a critical

- role in dense deposit disease and other C3 glomerulopathies. *Kidney Int* 2012;82(4):454-464.
26. Bu F, Borsa NG, Jones MB, Takanami E, Nishimura C, Hauer JJ, Azaiez H, Black-Ziegelbein EA, Meyer NC, Kolbe DL, Li Y, Frees K, Schnieders MJ, Thomas C, Nester C, Smith RJ. High-throughput genetic testing for thrombotic microangiopathies and C3 glomerulopathies. *J Am Soc Nephrol* 2016;27(4):1245-1253.
 27. Iatropoulos P, Noris M, Mele C, Piras R, Valoti E, Bresin E, Curreri M, Mondo E, Zito A, Gamba S, Bettoni S, Murer L, Fremeaux-Bacchi V, Vivarelli M, Emma F, Daina E, Remuzzi G. Complement gene variants determine the risk of immunoglobulin-associated MPGN and C3 glomerulopathy and predict long-term renal outcome. *Mol Immunol* 2016;71:131-142.
 28. Thomas S, Ranganathan D, Francis L, Madhan K, John GT. Current concepts in C3 glomerulopathy. *Indian J Nephrol* 2014;24(6):339-348.
 29. Osborne AJ, Breno M, Borsa NG, Bu F, Fremeaux-Bacchi V, Gale DP, van den Heuvel LP, Kavanagh D, Noris M, Pinto S, Rallapalli PM, Remuzzi G, Rodriguez de Cordoba S, Ruiz A, Smith RJH, Vieira-Martins P, Volokhina E, Wilson V, Goodship THJ, Perkins SJ. Statistical validation of rare complement variants provides insights into the molecular basis of atypical hemolytic uremic syndrome and C3 glomerulopathy. *J Immunol* 2018;200(7):2464-2478.
 30. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won HH, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG, Exome Aggregation C: Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;536(7616):285-291.
 31. Rodriguez E, Rallapalli PM, Osborne AJ, Perkins SJ. New functional and structural insights from updated mutational databases for complement factor H, Factor I, membrane cofactor protein and C3. *Bioscience Reports* 2014;34(5)e00146.
 32. Saunders RE, Goodship TH, Zipfel PF, Perkins SJ. An interactive web database of factor H-associated hemolytic uremic syndrome mutations: insights into the structural consequences of disease-associated mutations. *Human Mutation* 2006;27(1):21-30.
 33. Perkins SJ, Goodship TH. Molecular modelling of mutations in the C-terminal domains of factor H of human complement: a correlation between haemolytic uraemic syndrome and a predicted heparin binding site. *J Mol Biol* 2002;316(2):217-224.
 34. Walsh R, Thomson KL, Ware JS, Funke BH, Woodley J, McGuire KJ, Mazzarotto F, Blair E, Seller A, Taylor JC, Minikel EV, Exome Aggregation C, MacArthur DG, Farrall M, Coe SA, Watkins H. Reassessment of Mendelian gene pathogenicity using 7,855 cardiomyopathy cases and 60,706 reference samples. *Genet Med* 2017;19(2):192-203.
 35. Auer PL, Lettre G. Rare variant association studies: considerations, challenges and opportunities. *Genome Med* 2015;7(1),16.

36. Wu J, Wu YQ, Ricklin D, Janssen BJ, Lambris JD, Gros P. Structure of complement fragment C3b-factor H and implications for host protection by complement regulators. *Nat Immunol* 2009;10(7):728-733.
37. Blaum BS, Hannan JP, Herbert AP, Kavanagh D, Uhrin D, Stehle T. Structural basis for sialic acid-mediated self-recognition by complement factor H. *Nat Chem Biol* 2015;11(1):77-82.
38. Okemefuna AI, Nan R, Gor J, Perkins SJ. Electrostatic interactions contribute to the folded-back conformation of wild type human factor H. *J Mol Biol* 2009;391(1):98-118.
39. Janssen BJ, Huizinga EG, Raaijmakers HC, Roos A, Daha MR, Nilsson-Ekdahl K, Nilsson B, Gros P. Structures of complement component C3 provide insights into the function and evolution of immunity. *Nature* 2005; 437(7058): 505-511.
40. Roversi P, Johnson S, Caesar JJ, McLean F, Leath KJ, Tsiftoglou SA, Morgan BP, Harris CL, Sim RB, Lea SM. Structural basis for complement factor I control and its disease-associated sequence polymorphisms. *Proc Natl Acad Sci USA* 2011;108(31):12839-12844.
41. Persson BD, Schmitz NB, Santiago C, Zocher G, Larvie M, Scheu U, Casasnovas JM, Stehle T. Structure of the extracellular portion of CD46 provides insights into its interactions with complement proteins and pathogens. *PLoS Pathog* 2010;6(9):e1001122.
42. Osborne AJ, Nan R, Miller A, Bhatt JS, Gor J, Perkins SJ. Two distinct conformations of factor H regulate discrete complement-binding functions in the fluid phase and at cell surfaces. *J Biol Chem* 2018;293(44),17166-17187.
43. Xue X, Wu J, Ricklin D, Forneris F, Di Crescenzo P, Schmidt CQ, Granneman J, Sharp TH, Lambris JD, Gros P. Regulator-dependent mechanisms of C3b processing by factor I allow differentiation of immune responses. *Nat Struct Mol Biol* 2017;24(8):643-651.
44. Kolodziejczyk, R., Mikula, K.M., Kotila, T., Postis, V.L.G., Jokiranta, T.S., Goldman, A., Meri, T. Crystal structure of a tripartite complex between C3dg, C-terminal domains of factor H and OspE of *Borrelia burgdorferi*. *PLoS One* 2017;12(11):e0188127.
45. Osborne AJ. *PhD thesis*, University College London. 2018.
46. Goodship TH, Cook HT, Fakhouri F, Fervenza FC, Fremeaux-Bacchi V, Kavanagh D, Nester CM, Noris M, Pickering MC, Rodriguez de Cordoba S, Roumenina LT, Sethi S, Smith RJ, Conference P. Atypical hemolytic uremic syndrome and C3 glomerulopathy: conclusions from a "Kidney Disease: Improving Global Outcomes" (KDIGO) Controversies Conference. *Kidney Int* 2017;91(3),539-551.

FIGURE LEGENDS

Figure 1. Pie chart analyses of the reference allele frequencies of the rare variants identified in the aHUS and C3G datasets, and the distribution and disease allele frequencies of non-benign missense rare variants in the FH and C3 domains.

A. The total number of unique rare variants are shown in the bar chart. These are broken down in terms of allelic frequencies in the two pie charts for each of aHUS and C3G. The aHUS and C3G datasets have allele numbers of 634-6256 and 208-886 respectively, to be compared to the ExAC dataset with an allele number of 14,708-121,142.

B. The 20 SCR domains in FH. The functional binding sites associated with each SCR domain are shown by labelled coloured arrows. Each domain missense rare variant allele frequency is normalised for its size by dividing it by the proportion of residues in the protein domain. Pink represents the total missense rare variants identified in the aHUS dataset, blue for C3G, and green for both aHUS and C3G.

C. The 13 domains in C3. The aNT region and anchor peptide are shown also. The functional binding sites associated with each C3 domain are shown by labelled coloured arrows to correspond to the SCR domains in FH (pink), FI domains (orange) and the cell surface (cyan). The C3d binding site on SCR-19/20 is equivalent to the TED domain in C3. The colours follow those in B. Adapted from Ref. 29.

Figure 2. Domain structures of Factor H, Factor I, Membrane Cofactor Protein and C3.

The schematic cartoon (not drawn to scale) with residue numbering represents the 20 SCR domains of Factor H, the five FIMAC, CD5, LDLr1, LDLr2 and serine protease domains of Factor I, the four SCR domains of Membrane cofactor protein, and the 13 domains of complement C3, including eight MG domains, and the LINK, ANA, CUB, TED and C345C domains. Crystal structures are known for all these domains except for nuclear magnetic resonance-determined structures for FH SCR-5, SCR-10/13 and SCR-15/16 (green) and homology models for FH SCR-9, SCR-14 and SCR-17 (orange). The signal peptides are shown as grey boxes at the N-terminus (left). The totals of reported genetic variants in each domain are shown in blue above each domain. Abbreviations: SCR, short complement regulator/short consensus repeat; MG, macroglobulin.

Figure 3. Rare missense variants mapped onto structures for FH in its complexes with C3b, FI and C3d.

A. The N-terminal extended FH model, from atomistic scattering modelling, shown as a blue ribbon cartoon, was aligned with structures for FI (green), C3b (tan) and SCR-1/4 (cyan) seen in their co-crystal structures. B. The C-terminal extended FH structure was aligned with the co-crystal structure for C3d (green) and FH SCR-19/20 (not shown). Missense rare variants for aHUS, C3G and both aHUS and C3G were mapped as red, yellow and black spheres, respectively. For FH, 128 aHUS (109 unique residues) and 19 C3G (18 unique residues) missense rare variants are shown, of which eight in each dataset are common to both aHUS and C3G (eight unique residues). For C3, 61 aHUS (56 unique residues) and 24 C3G (24 unique residues) missense rare variants are shown, of which five in each dataset are common to both aHUS and C3G (five unique residues). Adapted from Ref (45). C. To show the distribution of aHUS and C3G rare variants in the co-crystal structure of C3b (bronze) and FH SCR-1/4 (cyan), two views of this structure are shown, rotated by 90° (PDB code 2WII). The 67 aHUS missense rare variants (red) in C3b were mostly located to the MG2 and MG6 domains that were adjacent to the SCR-1/4 domains of FH in their complex (denoted by a dashed red ellipse), and also by residues in the TED domain (such as ¹¹³⁴DICEEQ¹¹³⁹). In distinction, the 29 C3G missense rare variants (yellow) were spread more evenly throughout the C3b structure.

Figure 4. Front and back view of a secondary structure cartoon of the consensus SCR structure. The N-terminus and C-terminus are denoted by N and C respectively. Four of the six β -strands ($\beta 2$, $\beta 4$, $\beta 6$ and $\beta 7$) are visible in this cartoon. The number of missense variants at each position in the sequence are highlighted as spheres. The size of the individual spheres indicates the number of variants at that position, and ranges from one to 13. The four conserved Cys residues (5-13 variants) and one Trp residue (7 variants) are highlighted in purple and red. Five SCR residue positions that show 5-6 variants are highlighted as orange spheres. The position of the so-called hypervariable loop between Ser13 and Lys14 is arrowed.

Figure 5. The genetic search results web-page in the Database for a FH variant p.Cys431Tyr. The original interactive database^{31,32} for the four CFH, CFI, MCP and C3 complement genes had 27,608 hits in March 2018 at the point when it was replaced by the new database for 13 genes at <https://www.complement-db.org/>.²⁹ Since then, the total of hits has increased to 31,503 in February 2020. The interactive nature of the web database enables searches for variant data based on search terms that includes the source of the variant data, the disease condition, the number of variants per case and an Exome Aggregation Consortium (ExAC) allele frequency cut-off value for filtering based on a reference allele frequency of 0.01 that can be adjusted. For the *CFH* variant, c.1292G>A p.Cys431Tyr in the SCR-7 domain, which is associated with both aHUS and C3G, the estimated allele frequencies of 0.00016 for aHUS and 0.002257 for C3G are shown. These values are compared with those derived from the 1000 Genomes Project (1000GP), the Exome Aggregation Consortium (ExAC) and the Exome Variant Server (EVS). For the ExAC values, these allele frequencies are further categorised by ethnicity. Links are also provided to this variant in two other databases of single nucleotide polymorphisms (dbSNP) and ClinVar. Similar websites for the evolutionarily-related coagulation proteins Factors VII, VIII, IX and XI have been created in the author's laboratory (<http://www.factorvii.org/>; <http://www.factorviii-db.org/>; <http://www.factorix.org/> and <http://www.factorxi.org/>). The website copyright for all these sites is retained by S.J. Perkins and University College London.

Figure 6. The structural search results web-page in the Database for the variant FH p.Cys431Tyr. Using as example the same *CFH* variant c.1292G>A p.Cys431Tyr as in Figure 5, the structural and functional implications of this genetic variant in the database are presented as follows. The physicochemical properties of the wild-type and mutated residue are first compared, then the results from the *in silico* predictive tools PolyPhen-2, SIFT and Provean are presented. Because the missense variant affects a Cys residue, this variant is predicted to be damaging. The variant (red text) is visualised on the three-dimensional protein structure using the JSMol tool. SCR-1/7 are labelled here for reason of clarity, although these are not displayed on the web site.

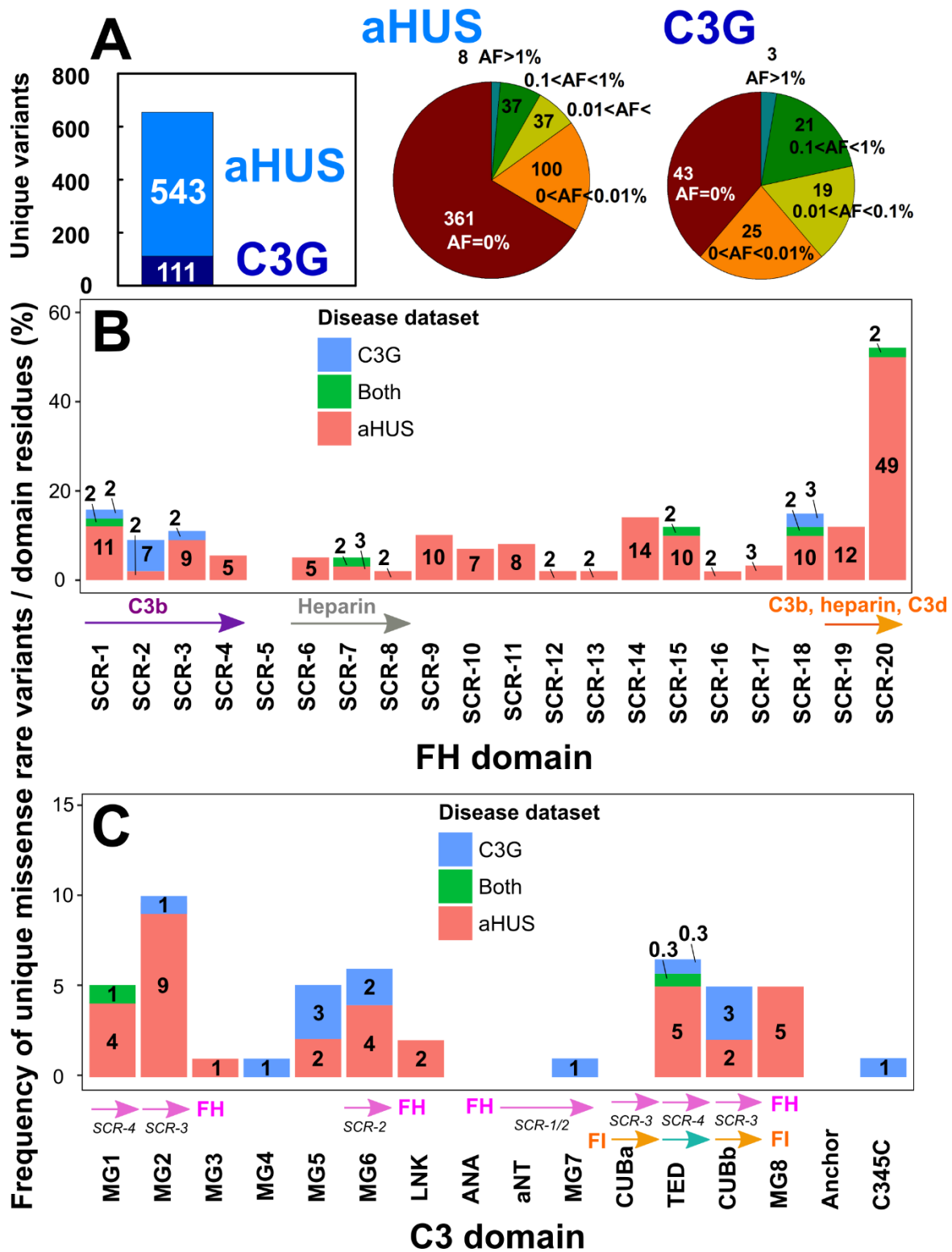


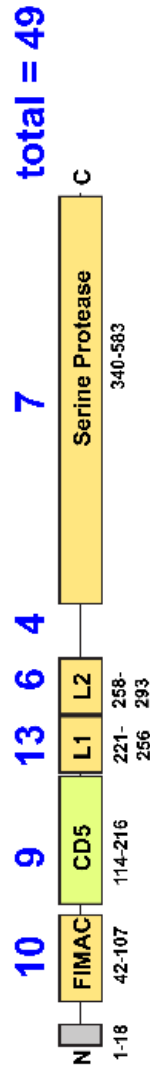
Figure 1. Pie chart analyses of the reference allele frequencies of the rare variants identified in the aHUS and C3G datasets, and the distribution and disease allele frequencies of non-benign missense rare variants in the FH and C3 domains.

Complement Factor H (FH)



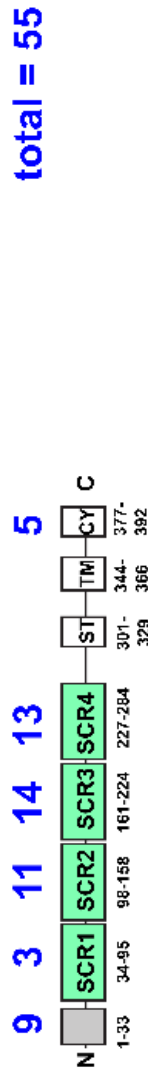
total = 190

Complement Factor I (FI)



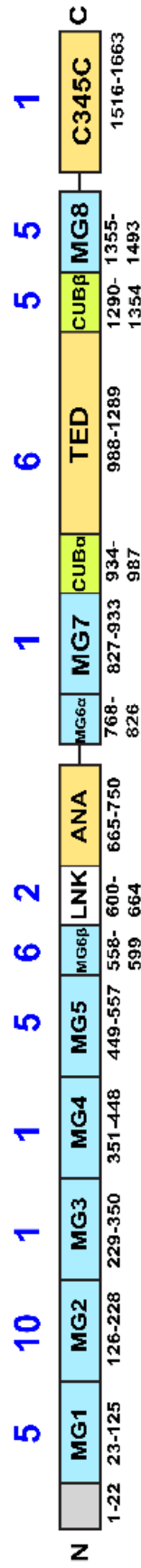
total = 49

Membrane Cofactor Protein (MCP)



total = 55

Complement C3



total = 48

Figure 2. Domain structures of Factor H, Factor I, Membrane Cofactor Protein and C3

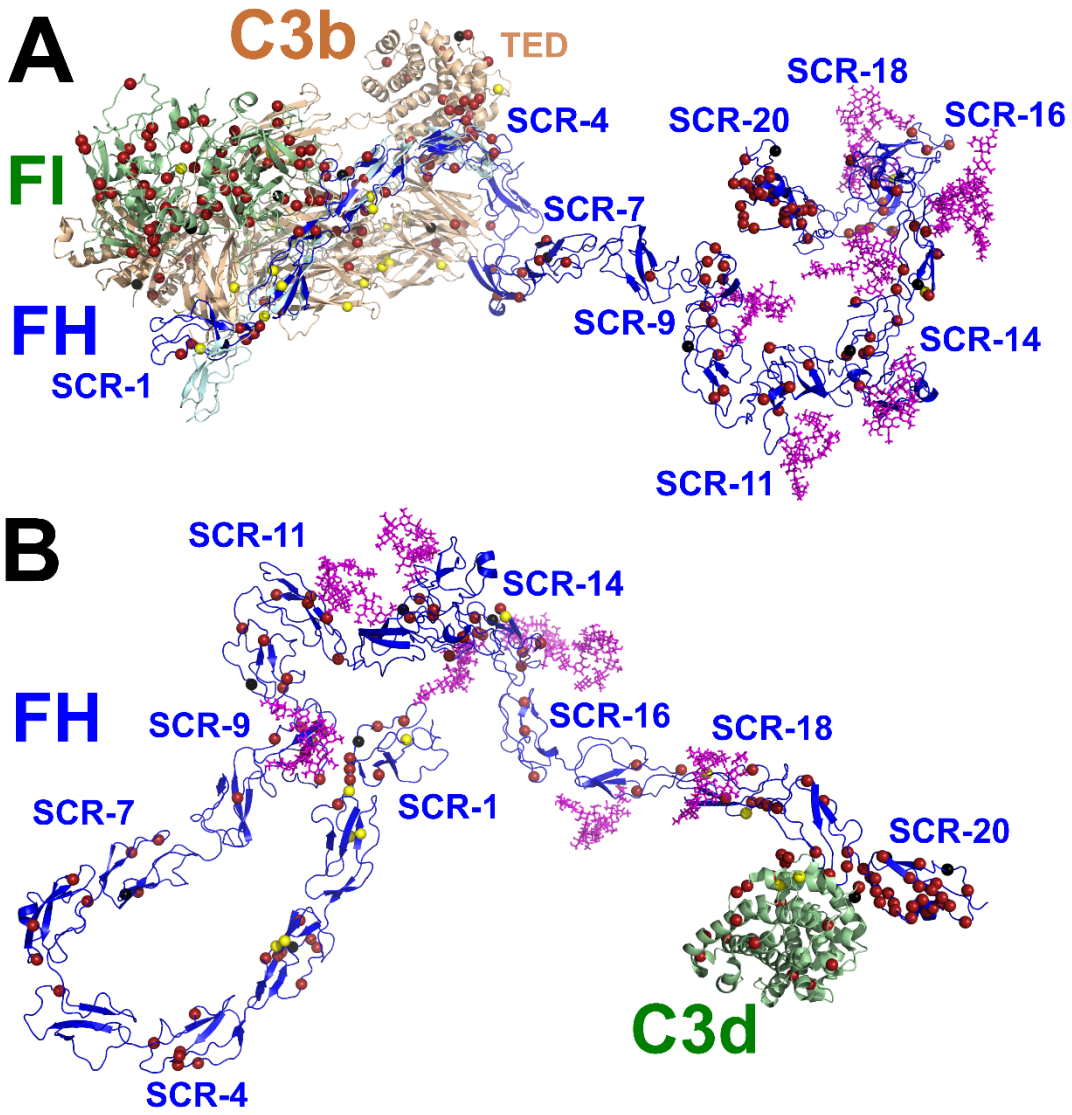


Figure 3AB (legend below)

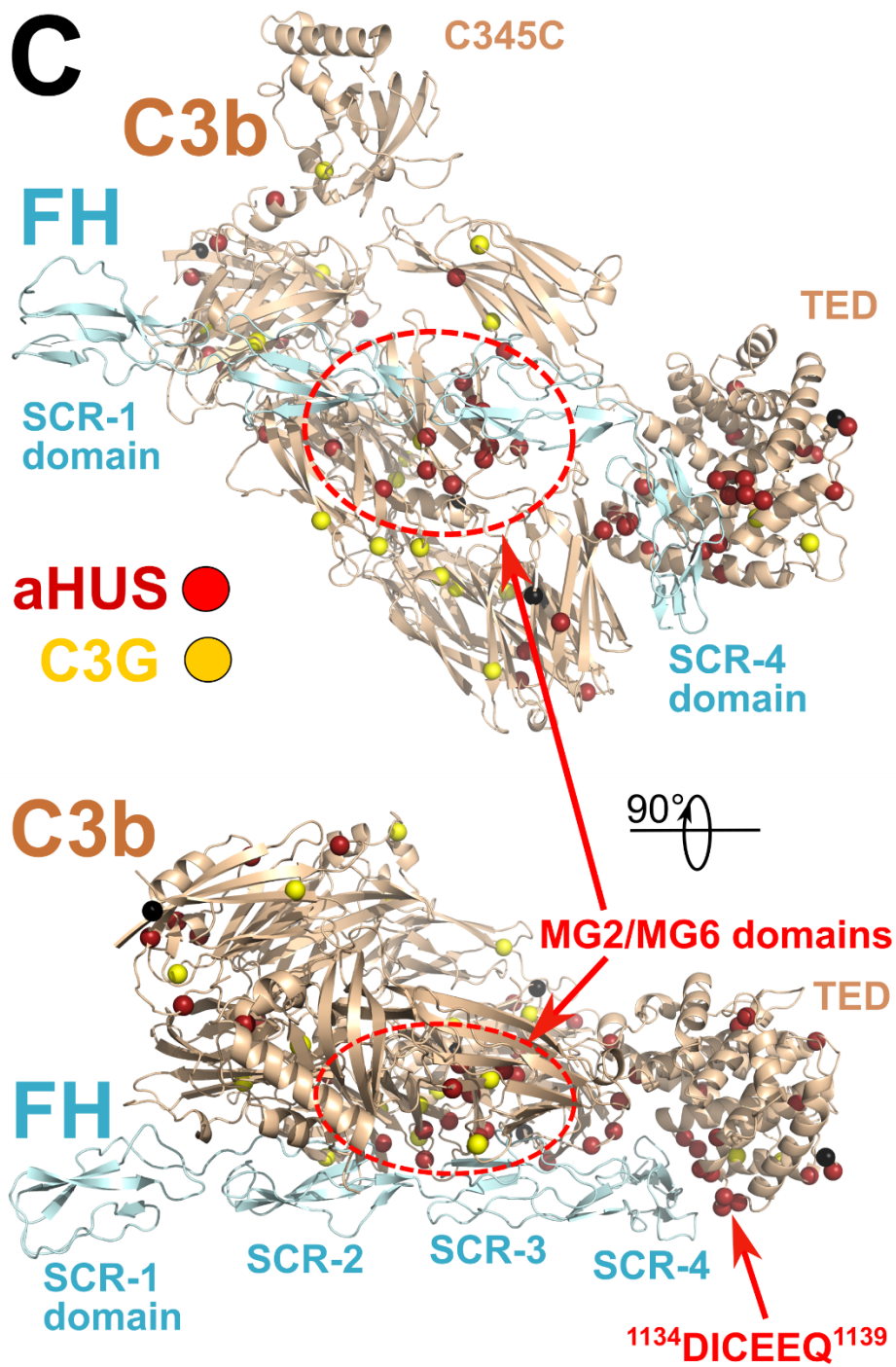


Figure 3C. Rare missense variants mapped onto structures for FH in its complexes with C3b, FI and C3d.

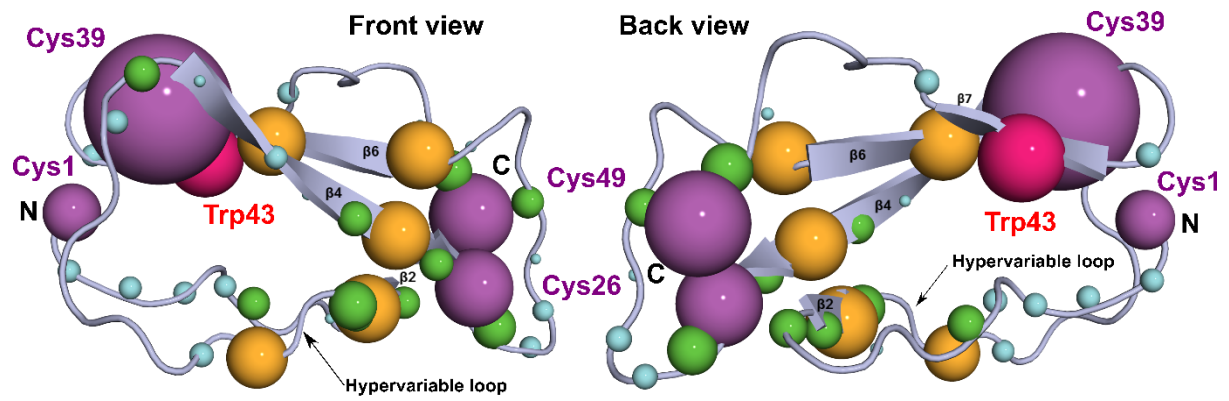


Figure 4. Front and back view of a secondary structure cartoon of the consensus SCR structure.

Database of complement gene variants

Home
Advanced Search
Variants ▼
Structures
AA Alignments
Resources ▼
Site Map
Log in

Search Results

Criteria:

- Source: Osborne et al, 2018
- Gene(s): *CFH*
- Amino acid no.: 431

1 unique variant retrieved.

Data Export Options:

UNIQUE (without patient data) : EXCEL/CSV TABULAR

MULTIPLE (with patient data) : EXCEL/CSV TABULAR

CFH (FH) c.1292G>A p.Cys431Tyr (413) The 1000 Genomes Project AF: 0

Location: Exon(9) Variant Type: Point Variant Effect: Missense
 Domain: SCR7 Transcript: [ENST00000367429](#) Genomic: 1: 196659325 [GRCh37]
 Phenotype ▶

Number	FH level	C3 level	FI level	MCP level	AntiFH Ab	Zygoty	Disease inheritance	Other variants	Condition	Reference
1257						Heterozygous	Recurrent		C3G	Osborne et al, 2018
1263						Heterozygous	Recurrent		C3G	Osborne et al, 2018
1547	59	239	92	N		Heterozygous	Sporadic		aHUS	Osborne et al, 2018

Allele Frequency ▶

aHUS population (based on [Osborne et al, 2018](#) only)

Allele Count (AC)	No. of patients screened for CFH	Allele Number (AN)	Allele Frequency (AF) estimate
1	3128	6256	0.00016

C3G population (based on [Osborne et al, 2018](#) only)

Allele Count (AC)	No. of patients screened for CFH	Allele Number (AN)	Allele Frequency (AF) estimate
2	443	886	0.002257

Reference populations

dbSNP: Not found.
ClinVar entry
The 1000 Genomes Project (1000GP):

1000GP Alleles	1000GP AC	1000GP AN	1000GP AF	P (aHUS)	P (C3G)
-	0	Mean: 5008	0	0.1 < P < 0.4875	0.005 < P < 0.01

ExAC:

ExAC Ethnicity	ExAC AC	ExAC AN	ExAC AF	P (aHUS)	P (C3G)
-	0	Mean: 118177	0	0.0125 < P < 0.025	< 0.00005

Exome Variant Server (EVS):

EVS Ethnicity	EVS Alleles	EVS Allele	EVS AC	EVS AN	EVS AF	P (aHUS)	P (C3G)
-	-	-	0	Mean: 12745	0	0.1 < P < 0.4875	< 0.00005

Structure and Function ▶


Study Title	Summary of findings for <i>CFH</i> p.Cys431Tyr	Reference
Genetics and Outcome of Atypical Hemolytic Uremic Syndrome: A Nationwide French Series Comparing Children and Adults	Type I mutation (protein level below the normal range).	Fremeaux-Bacchi et al, 2013

Please click [here](#) to view the mapping of this variant onto the latest FH structure and view the functional analysis by PolyPhen-2 and SIFT.

Variant classification
 (using guidelines in ACMG and Goodship et al, 2017): Likely pathogenic

Structural Immunology Group · University College London · Gower Street, London WC1E 6BT

Figure 5. The genetic search results web-page in the Database for a FH variant p.Cys431Tyr.



Database of complement gene variants

Home Advanced Search Variants ▼ Structures AA Alignments Resources ▼ Site Map Log in

In Depth Variation Analysis: **c.1292G>A (p.Cys431Tyr)**

CFH (FH) c.1292G>A p.Cys431Tyr (413) The 1000 Genomes Project AF: 0

Location: Exon(9) **Variant Type:** Point **Variant Effect:** Missense
Domain: SCR7 **Transcript:** [ENST00000367429](#) **Genomic:** 196659325 [\[GRCh37\]](#)

Residue Information:

	Name	Size	Charge	Hydrophobicity	Preferred position	Type	Class
Wild Type	Cys	medium	neutral	hydrophilic	buried	-	-
Mutated	Tyr	large	neutral	hydrophobic	surface	-	aromatic

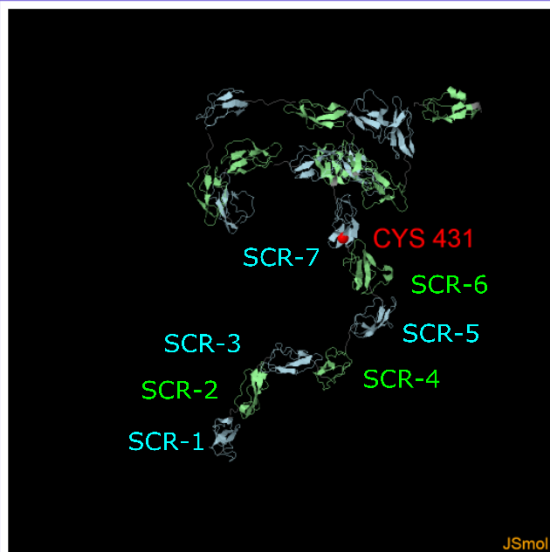
- Structural Implications:
- **Cys431** is shown in the structural model below as a red sphere.
 - The substitution involves a conserved cysteine residue that forms a disulphide bridge. Breakage of this disulphide bridge may have serious effects on the stability of the domain.
 - **Cys431** is in a proposed C3c binding site [Schmidt \(2008\)](#).

Tool	Prediction	Binding site	PPH2	DSSP (SecStr)
PolyPhen-2 ; HumDiv	probably damaging	DISULFID	deleterious	-
SIFT	Damaging (score: 0.00)	-	-	-
Provean	Deleterious (score: -8.57)	-	-	-

Available structures:

- **3N0J**: Full length FH (original: [3N0J.pdb](#))

The default structure is shown below.



Controls

The molecule can be rotated with the mouse holding the left button and the image can be magnified using the middle mouse button. Right Click on the JMOL screen for more options.

Options

Structure 3N0J

Spacefill OFF ON

Cartoon OFF ON

Wireframe OFF ON

Trace OFF 0.4 0.8 structure

Backbone OFF ON

Spin OFF ON

Background Black White

Disulphides OFF ON

Colour OFF Domains Structure Amino acid Binding sites

Labels OFF Domains Mutation Binding sites

[Please click here to view the Database Statistics](#)

Figure 6. The structural search results web-page in the Database for the FH variant p.Cys431Tyr.