Database of Complement Gene Variants: a comprehensive database providing insights on function, structure and allele frequency for genetic variants identified in complement-mediated diseases

Amy J. Osborne[1], Santiago Rodriguez de Cordoba[2], Veronique Fremeaux-Bacchi[3], Marina Noris[4], Richard J. Smith[5], Bert van den Heuvel[6], Timothy H. J. Goodship[7], Stephen J. Perkins[1]

[1]Department of Structural & Molecular Biology, University College London, UK; [2]Departamento de Inmunologia, Centro de Investigaciones Biologicas (CSIC), Madrid, Spain; [3]Service d'Immunologie Biologique, Hopital Europeen Georges Pompidou, Paris, France; [4]Laboratory of Immunology and Genetics of Rare Diseases and Transplantation, Mario Negri Institute for Pharmacological Research, Italy; [5]Iowa Institute of Human Genetics, University of Iowa, Iowa City, US; [6]Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; [7]Institute of Genetic Medicine, Newcastle University, UK

Atypical haemolytic uremic syndrome (aHUS) and C3 glomerulopathy (C3G) are associated with dysregulation of the alternative pathway (AP) of complement. Recently, we described an updated compilation of 324 genetic variants identified in four AP genes encoding the proteins factor H (FH), factor I (FI), membrane cofactor protein (MCP, CD46) and C3 in our interactive FH-HUS web-database (www.fh-hus.org). Our database has been updated further to a new version, the 'Database of Complement Gene Variants' (www.complement-db.org), with variant data contributed by six centres (Newcastle, Paris, Bergamo, Madrid, Iowa City, and Nijmegen) from >1000 patients with aHUS and C3G. This laboratory dataset comprises 601 variants (allele frequency < 1%) in 13 genes (*CFH, CFI, CD46, C3, CFB, CFHR1, CFHR3, CFHR5, CFP, PLG, DGKE, THBD, ADAMTS13*). The database provides enhanced search tools and the ability to assess variant pathogenicity using:

- Comparison of the estimated disease-associated variant allele frequencies to those of matched variants in 3 control datasets, namely "The 1000 Genomes Project," the "ExAC" and the "Exome Variant Server (EVS)" using Chi-square analyses.
- The mapping of missense variants onto protein structural models.
- Determination of the pathogenicity of each variant from its residue position, amino acid property, binding sites and residue conservation using PolyPhen-2, PROVEAN and SIFT.
- Measurement of the degree to which amino acid properties at any residue position are conserved across evolution, both within humans, and across other species, by multiple sequence alignment methods.

Thus, the tools in this new version of the database enable us to identify rare variants of the complement proteins enriched in disease which are worthy of further functional analysis. We will describe their occurrence in these 13 proteins, and discuss the utility of this new database for characterising complement-mediated diseases.

(1949 characters)