Immunogenetics (1995) 41: 139–143

BRIEF COMMUNICATION

Fernando Pardo-Manuel de Villena
Santiago Rodríguez de Córdoba

C4BPAL2: a second duplication of the *C4BPA* gene in the human *RCA* gene cluster

Received: 9 September 1994

We report here the characterization of the *C4BPAL2* gene, a new member of the human regulator of complement activation (*RCA*) gene cluster that arose from the duplication of the gene coding for the A chain of the complement component C4b-binding protein (C4BPA). We postulate that *C4BPAL2* is the human homolog of the pig *ApoR* gene.

The human *RCA* gene cluster is located in the long arm of chromosome 1 (1q32; Weiss et al. 1987), and spans over several kilobases (kb) of DNA. It includes the functional genes *CR1, CR2, DAF, MCP, C4BPA, C4BPB, HF,* and *F13B* (Bora et al. 1989; Carroll et al. 1988; Pardo-Manuel et al. 1990; Rey-Campos et al. 1987, 1988; Rodríguez de Córdoba et al. 1985, 1988), and three apparently nonfunctional genes, *CR1L1, MCPL1*, and *C4BPAL1* (Hourcade et al. 1990, 1992; Sánchez-Corral et al. 1993). These three genes are incomplete genomic duplications of the *CR1, MCP,* and *C4BPA* genes, respectively. Classical genetic studies suggested the existence of two separated regions within the human *RCA* gene cluster, namely the *C4BP-CR1* and the *H* regions (Rodríguez de Córdoba et al. 1988). Two-color fluorescence in situ hybridization analyses support this conclusion and demonstrate that these two regions are physically distant and separated by a region that contains the *REN* and *LCA* (*CD45*) genes (F. Pardo-Manuel de Villena and S. Rodríguez de Córdoba, unpublished data). Pulsed field gel electrophoresis has been used to determine the gene organization within both regions of the *RCA* gene cluster (Bora et al. 1989; Carroll et al. 1988; Hourcade et al. 1990, 1992; Pardo-Manuel et al. 1990; Rey-Campos et al. 1988, 1990; Sánchez-Corral et al. 1993). Thus, the genes within the *C4BP-CR1* region were shown to be arranged in the order: *5′-C4BPB-C4BPA-C4BPAL1-DAF-CR2-CR1-MCPL1-CR1L1-MCP-3′*.

The nucleotide sequence data reported in this paper have been submitted to the EMBL nucleotide sequence database and have been assigned the accession numbers X81360, X81361, and X81362

F. Pardo-Manuel de Villena · S. Rodríguez de Córdoba (✉)
Unidad de Immunologia, Centro de Investigaciones Biológicas (CSIC), Velazquez 144, 28006 Madrid, Spain

Most of the *RCA* genes code for proteins involved in the control of activation of the complement system (reviewed by Hourcade et al. 1989). They all belong to a multigene superfamily characterized by the presence of a 60-amino-acid repeat called SCR domain (Reid et al. 1986). The *RCA* genes present common structural features and a general pattern of genomic organization that supports the concept that they were generated by multiple events of gene duplication from a single ancestor (Farries and Atkinson 1991; Hourcade et al. 1989). There are, however, other structural peculiarities among the *RCA* genes, mostly located at their 3′ ends, which are only shared by some members of the gene cluster. Interestingly, genes sharing this second level of similarities map together within specific subregions of the *RCA* gene cluster. We have previously suggested that these subregions are still active sites for gene duplication and have postulated important differences in the genes encoded at these locations among different mammalian species (Pardo-Manuel et al. 1990; Rodríguez de Córdoba et al. 1994; Sánchez-Corral et al. 1993).

Apolipoprotein R (ApoR) is a 23000 M_r protein of unknown function that is found on very low-density lipoprotein (VLDL), on chylomicrons, and in the d >1.21 gr/mL fraction of pig plasma. ApoR is also a member of the SCR multigene superfamily. It is composed of two SCR and a C-terminal domain that are homologous to SCR-7, SCR-8 and the C-terminal region of the human C4BPα polypeptide (Cooper and Attie 1992). Interestingly, no protein homologous to ApoR has been identified in humans and other mammals (Cooper and Attie 1992).

We have recently cloned the entire *C4BP* subregion of the human *RCA* gene cluster in overlapping YAC clones (Sánchez-Corral et al. 1993). Whole yeast DNA from three of these clones, y87G11, y402C8, and y404G11 (codes from the CEPH YAC library, Paris (Albertsen et al. 1990), was partially digested with the restriction enzyme *Sau* 3A and subcloned in the *Bam* HI site of the λEMBL-3 vector. Most of the λEMBL-3 clones that contained human DNA were isolated from these libraries, using total human DNA as a probe, and were ordered in a contig that spans

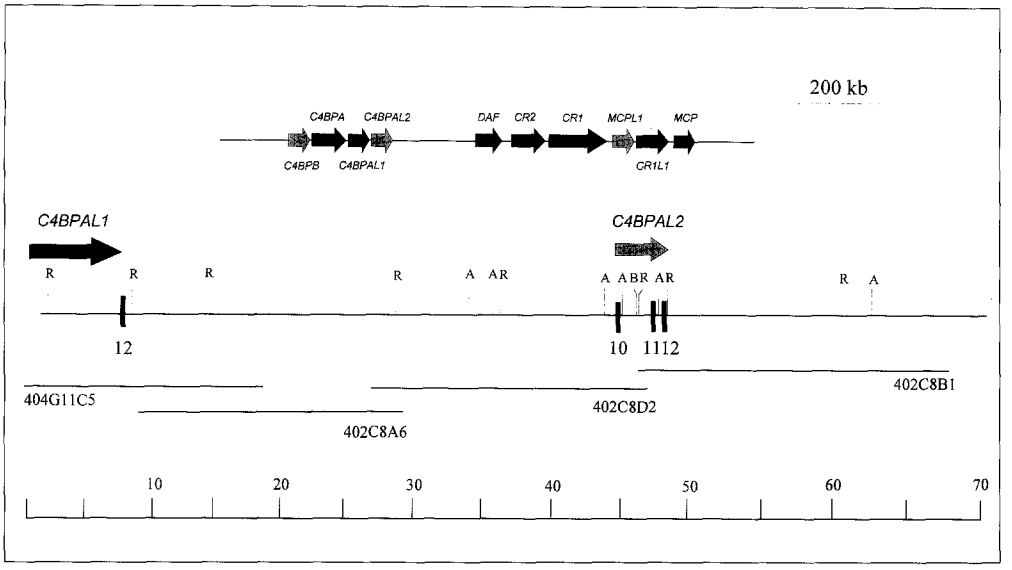

Fig. 1 Restriction map of the *C4BPAL2* region. The alignment of the 404G11C5, 402C8A6, 402C8D2, and 402C8B1 genomic clones is presented to show the overall organization of *C4BPAL2* and its location in relation to other genes of the *RCA* gene cluster. Numbering of the exon-like regions in *C4BPAL2* and *C4BPAL1* refers to the exons of the human *C4BPA* gene to which they are homologous. At the *top* is presented the physical map of the *C4BP-CR1* region of the *RCA* gene cluster. Codes for restriction enzymes are A, *Apa* I; B, *Bam* HI; R, *Eco* RI. The *scale* at the *bottom* of the Figure expresses kilobases of DNA

350 kb of DNA. this contig, which includes the *C4BPB, C4BPA*, and *C4BPAL1* genes, plus 100 kb of DNA both upstream and downstream of these genes, is now being sequenced and will be described elsewhere. Four clones of this contig (404G11C5, 402C8A6, 402C8D2, and 402C8B1), spanning 70 kb downstream of the C4BPAL1 gene are described here, since they demonstrate the existence of a novel gene in this region. The alignment of these four clones and their location within the human *RCA* gene cluster is shown in Figure 1.

Clone 404G11C5 contains the 3′ end sequences of the *C4BPAL1* gene and consequently shows hybridization with a *C4BPA* cDNA probe. Surprisingly, clones 402C8D2 and 402C8B1, which are further downstream, also show hybridization with the *C4BPA* cDNA probe in low-stringency conditions (55 °C, 2 × standard sodium citrate). Southern blot analyses of these clones demonstrate the presence of three separate regions hybridizing with the *C4BPA* cDNA probe (Fig. 1). Restriction fragments of the 402C8D2 and 402C8B1 genomic clones that hybridize with the *C4BPA* cDNA probe were isolated, subcloned into the plasmid vector pBluescript SK+ (Stratagene, La Jolla, CA), and sequenced. The alignment of the nucleotide sequences of these restriction fragments and the sequence of the *C4BPA* cDNA, revealed three exon-like sequences homologous to the *C4BPA* exons 10, 11, and 12 (SCR-7, SCR-8, and C-terminal region (Rodríguez de Córdoba et al. 1991; Fig. 2). These three exons are in the correct relative orientation and present an arrangement that resembles the structure of the *C4BPA* gene. These findings indicate the existence of a second genomic duplication of the *C4BPA* gene, named *C4BPAL2*. The *C4BPAL2* gene is located 40 kb downstream from the *C4BPAL1* gene and it is in the same 5′ to 3′ orientation found for all *RCA* genes within the *C4BP-CR1* region. Further searches in this genomic region for additional sequences homologous to the *C4BPA* or *C4BPAL1* exons were unsuccessful, indicating that either no more exon-like sequences are present in *C4BPAL2* or that they do not have sufficient similarity to hybridize with the *C4BPA* cDNA probes.

The nucleotide sequence of the *C4BPAL2* gene is clearly different from that of *C4BPA* or *C4BPAL1* (Fig. 2), indicating that *C4BPAL2* is not the result of some kind of rearrangement in the YAC clones. Furthermore, Southern blot analyses of human genomic DNA from four different individuals, using genomic probes from the 3′ end of the *C4BPA, C4BPAL1*, and *C4BPAL2* genes, demonstrate that these probes hybridize with different restriction fragments in each of the four DNAs (not shown).

The nucleotide sequence of the *C4BPAL2* exons presents many nucleotide insertions compared with that of the *C4BPA* exons. These insertions result in frameshifts, which generate early termination codons in all three *C4BPAL2* exons (Fig. 2). These findings strongly suggest that *C4BPAL2* is a pseudogene. It is interesting, however, that similar to the situation found in *C4BPAL1* (Sánchez-Corral et al. 1993), the *C4BPA* exon sequences have not been equally conserved in *C4BPAL2* and present a pattern of sequence conservation that parallels that found between the human and mouse *C4BPA* genes. The *C4BPAL2* exons are also considerably better conserved than their flanking introns (Table 1). These observations denote the existence of some kind of positive selection to maintain the coding sequences in *C4BPAL2*, which suggests that it was a functional gene in the past.

Phylogenetic analyses indicate that *C4BPAL2* originated as a duplication of the human *C4BPA* gene and that this duplication preceded the duplication that originated *C4BPAL1* (Fig. 3). The relative location of the *C4BPA, C4BPAL1*, and *C4BPAL2* genes in the *RCA* gene cluster (Fig. 1) fits with this conclusion and suggests that *C4BPA, C4BPAL1*, and *C4BPAL2* originated by gene duplication involving unequal recombination.

Our conclusions that *C4BPAL2* originated early in the evolution of the human lineage, just after the mammalian radiation, and that it was a functional gene, imply that *C4BPAL2* may still be a functional gene in species that split from the human lineage several million years (MY) ago. In this context, the striking similarities between *C4BPAL2* and the pig gene coding for ApoR are very suggestive. The pig

Fig. 2 Nucleotide sequence of the *C4BPAL2* gene. The Figure shows the alignment of the nucleotide sequence of the three exons of the *C4BPAL2* gene with their homologous exons in the human *C4BPA*, human *C4BPAL1*, and pig *ApoR* genes. Sequences were aligned with the program CLUSTAL V (Higgins et al. 1992). The first 6 nucleotides flanking the 5′ and 3′ splice sites are also included in the comparison. *Dashes* represent identity in the nucleotide position with respect to the *C4BPAL2* sequence. The gaps introduced to maximize homology are represented by *asterisks*. *Asterisks* are also used to indicate that the exon coding SCR-7 is missing in the *C4BPAL1* gene. The pig *ApoR* cDNA sequence is also included, since, as discussed in the text, we postulate that it is the pig homolog of the human *C4BPAL2* gene

SCR-7

```
C4BPAL2    ttcaag TT TGC AGA ATG GTT CCT CCC ATT GGC CAT GGG TCC TAT GAA GAT GTG AGA TCA TTT
C4BPA      c----- -- --- -AT T-T CC- --- AAA --- -C- --- --- CAT --- A-- C-A TCT -GT --- -AC
ApoR             G- --- GAT -AT CC- --- GTG G-- -C- --- --A CAT C-- AC- C-A A-T -TT GGG C-A
C4BPAL1    ****** ** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** ***

C4BPAL2    TTT ATT CTC TAC TAC TCT GGT GCA C** TGT CAC TGT GAT GAT GGA TAT GTT CTA ATT GGA GAG
C4BPA      AGC T-- T-- A-A G-A GAG AT- AT- *** -A- G-A --- --- A-A --C --C A-- --G G-C --- C--
ApoR       --- GGA A-G A-A A-A GA- -AG -TT GTA -A- A-A --- --- --A --- --C AC- --G G-- --- ---
C4BPAL1    *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** ***

C4BPAL2    GCT GAA GTC ACC TGC AGA AAT TCA CAT TGG TCA GCT CCA GCC CCT CAA TGT AAA G gttaact
C4BPA      --G A-- C-- T-- --- --T T-- --- --C --- --- --- --- --- --- --- --- --- - --*----
ApoR       -AC AG- C-- T-- --- --T C-T TC- --- -GC --- --- C-- G-- --- --- --- --- --- -
C4BPAL1    *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** * *******
```

SCR-8

```
C4BPAL2    ttatag CT CTG TGT CTG AAA CCA GAG ACA GAA AAT GGA AAG CTG TCT GGG **T AAG GAT CTG
C4BPA      ---c-- -- --- --- -G- --- --- --A TT- -TG --- --- -G- T-- --- -T- GA- --- --- -A-
ApoR             -A T-- --- -C- --- --- C-- -T- --T CG- --- --- T-A --- -T- GA- C-- --- GAA
C4BPAL1    ---c-- -- --- --- --- --- --- --A -T- -TG --- --- -G- --- --- -T- GA- --- --- -A-

C4BPAL2    TAT GTT GAA CTT GAA AAT GTC ACC ATC CAG TGT GAC TCT GGC TAT AAA GTG GTT GGT CCT CAA
C4BPA      --- --- --- -C- --- --- --G --- --- --A --- --T --- --- --- GGT --- --- --- --C ---
ApoR       --- A-- --G TC- --- --C --- -TT G-- --- --- -G- --- --- --- GGT T-- --- --- --C A--
C4BPAL1    --- --- --G -C- --- --- --T C-- --T G-A --- --T --- --- --- GGT --- --- --- -TC A--

C4BPAL2    AAT ATC ACT TTGGTTGGTCCTCAAAATATCACTTGA TCA GAG CAC AGA ACATGGC AGG CAC CCA GAG GCT
C4BPA      -G- --- --- -**************************-T --T -G- A-- --- --****- T-- T-- --- --- -TG
ApoR       -T- --- --- -**************************-C A-- --A G-- G-- --****- T-- --- --- CG- -TG
C4BPAL1    -G- --- --- -**************************-C --- --- A-G --- --****- T-- T-- --- --A -TG

C4BPAL2    C*G AGT TGT GAG TGG gtaagt
C4BPA      -CC -AG --- --- --- ------
ApoR       -CC -AG --- --A ---
C4BPAL1    GC- --A --- --- CAG ------
```

C-Terminal

```
C4BPAL2    atttag GTA ATC TCT GAA GGT TTG GAG CAA TGT GCC ACG GGT AGA AAG CTC ATG CAG TGT CTC
C4BPA      ttac-- -AG -C- C-C --- --C -GT --A --- GTG CT- --A --C -A- -GA --- --- --- --- ---
ApoR             -AG TA- C-C --- -AC -GT --- --- GTG CAT GAA --C -A- --A --- --- --- --- ---
C4BPAL1    ------ -AG TC- C** *** *** **T --G --- GCA CT- C-A --C --- --A --- --- --- --- ---

C4BPAL2    CCA CAC CCA GAG GAT TCG AAA ATG ACC CTG GAG GTG TAA AAG CTG TCT CTG GAG ATT AAA CGA
C4BPA      --- A-- --- --- --- GT- --- --- G-- --- --- --A -AT --- --- --- --- --A --- G-- -A-
ApoR       --- AC- -TG --- --G ATA --- T-- G-- --- --- C-- --T --- --- --C --- --- -C- --- -T-
C4BPAL1    --- A-- --- --- A-- GT- --- --- G-- --- --- --- --T --- --- --- -CA --- --- G-C -T-

C4BPAL2    CTG GAA CAC GTG TGA GAC TT* G*A AGA CGAACAA TGC CAG TTA GAA CCA GTT AGG AAG
C4BPA      --- --- -TA CA- A-- --- AGC -C- --- -****-- -C- ACT --G --T AA- -AA CTA T-A
ApoR       --- --G -TT -AC AT- --T AAG -A- -AG A****-- GC- AAA GCG A-G TAC TCA -TA T-G
C4BPAL1    --- --- -TA **- A-- T-- AAG -G- --- T****G- -C- ACT C-G --- AA- TAA CTA T-A
```

ApoR and the human *C4BPAL2* genes are composed of precisely the same exons and show an overall 66% nucleotide sequence identity. Phylogenetic analysis, depicted in Figure 3, supports the conclusion that both genes are homologs and that the human gene evolved to a pseudogene after the separation of human and pig lineages.

The results shown in this report further illustrate the differences between the *RCA* gene cluster of the different mammalian species with respect to the number of genes and their individual evolution. Just within the *C4BP* gene family the differences are remarkable. Humans have four genes, namely *C4BPB, C4BPA, C4BPAL1*, and *C4BPAL2*, but only *C4BPB* and *C4BPA* appear to be functional. We show here that *C4BPAL2* is probably the human homolog of a functional gene in pigs, the *ApoR* gene. A similar situation may be true for the *C4BPAL1* gene. We have recently postulated that the *C4BPAL1* gene is functional in some of the species that split from the human lineage less than 50 MY ago (Sánchez-Corral et al. 1993). In mouse only two *C4BP* genes have been characterized thus far, *C4BPB* and *C4BPA* genes. Interestingly, *C4BPB* has evolved to a pseudogene in this species (Rodríguez de Córdoba et al. 1994) and this may also be the case for rabbits (He and Dahlbäck 1994).

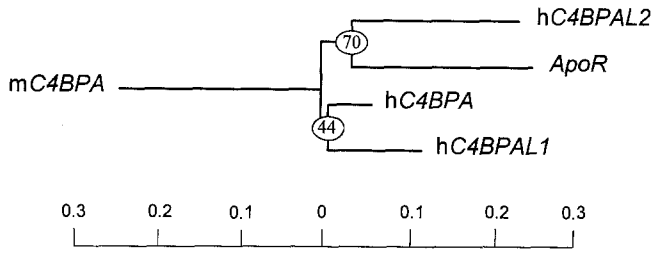

Fig. 3 Phylogenetic relationships of the *C4BPAL2* gene. Sequences of the SCR-8 and C-terminal region exons of the human *C4BPA* (h*C4BPA*), human *C4BPAL1* (h*C4BPAL1*), human *C4BPAL2* (h*C4BPAL2*), mouse *C4BPA* (m*C4BPA*), and pig ApoR (*ApoR*) genes were used in this analysis. The phylogenetic tree depicted in the Figure was obtained with the CLUSTAL V package, deleting all sites with gaps in any sequence, and using the Kimuras 2 parameters model distances correction (Higgins et al. 1992). The *scale* at the *bottom* represents the divergence values calculated by the neighbour-joining method (CLUSTAL V package). Numbers in the nodes are the percentages of bootstrapping trials (n = 1000) in which an identical node was produced, and are thus a measure of the robustness of the data generating that particular node. The phylogenetic analysis was restricted to the SCR-8 and C-terminal region exons, because these are the only sequences shared by the five genes

Table 1 Percentage of sequence identity between the exons and introns of *C4BPAL2* and their homologous regions in the human genes *C4BPA* and *C4BPAL1* and the pig *ApoR* gene. For each region the length of the sequence considered in the analysis is indicated between *parentheses*. Exon sequences are complete, but note that for the introns these figures do not match the real length of the introns. They are only to indicate the length of the sequence available for comparison

	C4BPAL2				
	SCR-7 (171)	Intron I (152)	SCR-8 (174)	Intron II (313)	C-ter (174)
C4BPA	65	55	82	39	69
C4BPAL1	–	–	79	45	62
ApoR	59	–	74	–	64

In summary, the data presented emphasize that the role gene duplication has played is the generation of new *RCA* genes after mammalian radiation, and they underline the importance of the comparative analysis of this region between different mammalian species.

Acknowledgments We thank Dr. A. Garcia-Pardo, Dr. M. A. Peñalva, and Dr. J. Perez Martin for helpful suggestions. We appreciate the excellent technical assistance of Ms. S. Vara de Rey. This work was supported by grants from the Spanish Fondo de Investigaciones Sanitarias (FIS-92/0889) and Comunidad Autonoma de Madrid (CAM-92/0155).

References

Albertsen, H. M., Abderahim, H., Cann, H. M., Dausset, J., Le Paslier, D., and Cohen, C. Construction and characterization of a yeast artificial chromosome library containing seven haploid genome equivalents. *Proc Natl Acad Sci USA 87:* 4256–4250, 1990

Bora, N. S., Lublin, D. L., Kumar, B. V., Hockett, R. D., Holers, V. M., and Atkinson, J. P. Structural gene for human membrane cofactor protein (MCP) of complement maps to within 100 kb of the 3′ end of the C3b/C4b receptor gene. *J Exp Med 169:* 597–602, 1989

Carroll, M. C., Alicot, E. M., Katzman, P. J., Klickstein, L. B., Simth, J. A., and Fearon, D. T. Organization of the genes encoding complement receptors type 1 and 2, decay accelerating factor and C4b-binding protein in the RCA locus on human chromosome 1. *J Exp Med 167:* 1271–1280, 1988

Cooper, S. C. and Attie, A. D. Pig Apolipoprotein R: a new member of the short consensus repeat family of proteins. *Biochemistry 31:* 12328–12336, 1992

Farries, T. C. and Atkinson, J. P. Evolution of the complement system. *Immunol Today 12:* 295–300, 1991

He, X. and Dahlbäck, B. Rabbit plasma, unlike its human counterpart, contains no complex between protein S and C4B-binding protein. *Thromb Haemostas 71:* 446–451, 1994

Higgins, D. G., Bleasy, A. J., and Sharp, P. M. CLUSTAL V: improved software for multiple sequence alignment. *CABIOS 8:* 189–191, 1992

Hourcade, D., Holers, V. M., and Atkinson, J. P. The regulators of complement activation (RCA) gene cluster. *Adv Immunol 45:* 381–416, 1989

Hourcade, D., Miesner, D. R., Bee, C., Zeldes, W., and Atkinson, J. P. Duplication and divergence of the amino-terminal coding region of complement receptor 1 (CR1) gene. *J Biol Chem 265:* 974–980, 1990

Hourcade, D., Garcia, A. D., Post, T. W., Taillon-Miller, P., Holers, V. M., Wagner, L. M. Bora, N. S., and Atkinson, J. P. Analysis of the human regulators of complement activation (RCA) gene cluster with yeast artificial chromosomes (YACs). *Genomics 12:* 289–300, 1992

Pardo-Manuel, F., Rey-Campos, J., Hillarp, A., Dahlbäck, B., and Rodríguez de Córdoba, S. Human genes for the α and β chains of complement C4b-binding protein are closely linked in a head-to-tail arrangement. *Proc Natl Acad Sci USA 87:* 4529–4532, 1990

Reid, K. B. M., Bentley, D. R., Campbell, R. D., Chung, L. P., Sim, R. B., Kristensen, T., and Tack, B. F. Complement system proteins which interact with C3b or C4b. A superfamily of structurally related proteins. *Immunol Today 7:* 230–234, 1986

Rey-Campos, J., Rubinstein, P., and Rodríguez de Córdoba, S. Decay accelerating factor: genetic polymorphism and linkage to the RCA (regulator of complement activation) gene cluster in humans. *J Exp Med 166:* 246–252, 1987

Rey-Campos, J., Rubinstein, P., and Rodríguez de Córdoba, S. A physical map of the human regulator of complement activation gene cluster linking the complement genes CR1, CR2, DAF and C4BP. *J Exp Med 167:* 664–669, 1988

Rey-Campos, J., Baeza-Sanz, D., and Rodríguez de Córdoba, S. Physical linkage of the human genes for complement factor H and coagulation factor XIII B subunit. *Genomics 7:* 644–646, 1990

Rodríguez de Córdoba, S., Lublin, D., Rubinstein, P., and Atkinson, J. P. Human genes for three complement components that regulate the activation of C3 are tightly linked. *J Exp Med 161:* 1189–1195, 1985

Rodríguez de Córdoba, S., Rey-Campos, J., Dykes, D. D., McAlpine, P. J., Wong, P., and Rubinstein, P. Coagulation factor XIII B subunit is encoded by a gene linked to the regulator of complement activation (RCA) gene cluster in man. *Immunogenetics 28:* 452–454, 1988

Rodríguez de Córdoba, S., Sánchez-Corral, P., and Rey-Campos, J. Structure of the gene coding for the α polypeptide chain of the human complement component C4b-binding protein. *J Exp Med 173:* 1073–1082, 1991

Rodríguez de Córdoba, S., Pérez-Blas, M., Ramos-Ruiz, R., Sánchez-Corral, P., Pardo-Manuel de Villena, F., and Rey-Campos, J. The gene coding for the β-chain of C4b-binding protein (C4BPB) has become a pseudogene in the mouse. *Genomics 21:* 501–509, 1994

Sánchez-Corral, P., Pardo-Manuel de Villena, F., Rey-Campos, J., and Rodríguez de Córdoba, S. *C4BPAL1*, a member of the human regulator of complement activation (RCA) gene cluster that resulted from the duplication of the gene coding for the α-chain of C4b-binding protein. *Genomics 17:* 185–193, 1993

Weiss, J. H., Morton, C. C., Bruns, G. A., Weis, J. J., Klickstein, L. B., Wong, W. W., and Fearon, D. T. A complement receptor locus: genes encoding C3b/C4b receptor and C3d/Epstein-Barr virus receptor map to 1q32. *J Immunol 138:* 312–319, 1987