This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: http://orca.cf.ac.uk/107348/

This is the author’s version of a work that was submitted to / accepted for publication.

Citation for final published version:
Boshra, Hani, Zelek, Wioleta, Hughes, Timothy, Rodriguez de Cordoba, Santiago and Morgan, B. Paul 2017. Absence of CD59 in Guinea Pigs: Analysis of the Cavia porcellus Genome suggests the evolution of a CD59 Pseudogene. Journal of Immunology , 1701238. 10.4049/jimmunol.1701238

Publishers page: http://dx.doi.org/10.4049/jimmunol.1701238
<http://dx.doi.org/10.4049/jimmunol.1701238>

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher’s version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.
Absence of CD59 in guinea pigs: analysis of the *Cavia Porcellus* genome suggests the evolution of a *CD59* pseudogene.

Hani Boshra¹⁺, Wioleta M. Zelek¹, Timothy R. Hughes¹, Santiago Rodriguez de Cordoba² and B. Paul Morgan¹.

**Running title:** Guinea pigs lack CD59.

¹ Systems Immunity Research Institute, School of Medicine, Cardiff University, Cardiff CF14 4XN, United Kingdom; ²Department of Cellular and Molecular Medicine and Center for Biological Research and Center for Biomedical Network Research on Rare Diseases, Madrid, Spain.

⁺ current address: Structural Biology Unit, CIC bioGUNE, Bizkaia Technology Park, Derio, Spain, 48160

Address correspondence to B.P. Morgan at the above address. Email: morganbp@cardiff.ac.uk.

Phone: (44) 2920687096.

**Footnote:** This work was funded by a Wellcome Trust Value in People Award to HB and BPM.

**Keywords:** Rodent; Complement; Gene rearrangement; Comparative immunology/evolution.
Abstract:

CD59 is a membrane-bound regulatory protein that inhibits the assembly of the terminal membrane attack complex (C5b-9; MAC) of complement. From its original discovery in humans almost thirty years ago, CD59 has been characterized in a variety of species, from primates to early vertebrates such as teleost fish. CD59 is ubiquitous in mammals; however, we have described circumstantial evidence suggesting that guinea pigs (*Cavia porcellus*) lack CD59, at least on erythrocytes. Here, we have used a combination of phylogenetic analyses with syntenic alignment of mammalian *CD59* genes to identify the only span of genomic DNA in *Cavia porcellus* that is homologous to a portion of mammalian *CD59*, and show that this segment of DNA is not transcribed. We describe a pseudogene sharing homology to exons 2 through 5 of human *CD59* present in the *Cavia porcellus* genome. This pseudogene was flanked by *Cavia porcellus* homologs of two genes, *FBXO3* and *ORF91*, a relationship and orientation that was consistent with other known mammalian *CD59* genes. Analysis using RNA-Seq confirmed that this segment of chromosomal DNA was not transcribed. We conclude that guinea pigs lack an intact gene encoding CD59, the first report of a mammalian species that does not express a functional CD59. The pseudogene we describe is likely the product of a genomic deletion event during its evolutionary divergence from other members of the rodent order.
Introduction:

The complement system plays an essential role in innate and adaptive immunity in all known vertebrate animals (1, 2). Complement immunity is initiated through antigen-antibody interactions (i.e. the classical pathway) (3), the activation and amplification of C3 in the presence of activating surfaces (i.e. alternative pathway or amplification loop) (4), or the recognition of sugar-moieties (i.e. lectin pathway) (5), all ultimately leading to the cleavage and activation of C3 (6). This in turn, initiates the formation of the membrane attack complex (MAC), composed of components C5b-C9 (7). Once formed, the MAC causes perforation and lysis of target cell membranes.

In vertebrates, complement regulators (CReg) have been well-characterized (8). CReg are soluble or membrane-bound proteins that are responsible for the protection of host cells from complement-mediated damage (9, 10). One of the most ubiquitously expressed CReg is CD59, an 18-20 kDa glycosylphosphatidylinositol (GPI)-anchored membrane-bound protein, expressed on almost all cell types in mammals, and preventing MAC assembly by inhibition of the interaction of the C5b-8 complex with C9 (11). While initially characterized in humans, CD59 has since been identified across vertebrate species, including mammals, birds, reptiles, amphibians, and boned fish (8, 12-14). Structural conservation of this molecule across animal phyla, as well is its expression across all types, suggest a high degree of importance of CD59 in protecting the host from complement-mediated damage.

Early work using recombinantly expressed or native human CD59 suggested that the MAC inhibitory activity was species specific (15, 16). However, subsequent work comparing the activities of human, rat, pig and sheep CD59 analogs against different species sources of complement found a high degree of cross-species activity (17). Notable in this latter study was
that guinea pig erythrocytes were uniquely sensitive to MAC-induced lysis regardless of the
species source of the MAC proteins, including homologous MAC proteins, and were uniquely
protected from complement-mediated lysis by exogenously added CD59 from the various
species. These observations provoked the suggestion that guinea pig erythrocytes lacked an
endogenous CD59 analog (17).

Although this unique susceptibility of guinea pig erythrocytes to complement-mediated
hemolysis has been known for several decades, there is still no definitive explanation as to why
this is the case. Components of the guinea pig complement system have been purified and
characterized from as early as the 1960’s (18, 19), with all components of the activation
pathways and terminal pathway characterized (20, 21). Guinea pig complement efficiently lyses
“classical” targets such as antibody-sensitised sheep erythrocytes, demonstrating that there is not
an underlying complement deficiency. Other CReg, including decay accelerating factor (DAF;
CD55) and membrane cofactor protein (MCF; CD46) have been identified in and cloned from
guinea pig (22, 23), but CD59 remains unidentified and an enigma.

Here we describe a combined immunochemical, phylogenetic and genomic approach to
elucidate whether guinea pigs express CD59. Our data further strengthen the contention that
guinea pigs do not express CD59. We show that the gene encoding for CD59 in the Cavia
porcellus genome has become a pseudogene. Together, the data demonstrate that there is no
functional gene encoding CD59 in guinea pigs; these findings provide an explanation for the
unique hemolytic properties attributed to guinea pig erythrocytes.
Materials and Methods:

Expression of CD59 on guinea pig erythrocytes

The expression of CD59 on guinea pig erythrocytes was assessed using six different antibodies; two rabbit polyclonal antibodies generated against mouse CD59a and known to be cross-reactive across species, including human CD59, and two monoclonal antibodies specific for human CD59 and mouse CD59a respectively. Erythrocytes from normal mice (C57/Bl), CD59a-deficient mice and human were used as positive and negative controls. Erythrocytes from each source were washed three times in isotonic phosphate-buffered saline (PBS) with centrifugation at 2000 g for 10 minutes to pellet after each wash, then resuspended to 0.1% (v/v) in flow cytometry buffer (FCB; PBS pH 7.4 supplemented with 1% bovine serum albumin (w/v)). Aliquots (100µl) were incubated with 1µg of the appropriate anti-CD59 antibody at 4°C for one hour, washed twice in FCB then incubated for one hour at 4°C with the appropriate fluorescein (FITC)-conjugated secondary; donkey anti-mouse IgG (Jackson Immunoresearch, 715-096-151), goat anti-rabbit IgG (Oxford Biotechnology L320-NC51Z), or rabbit anti-rat IgG (Sigma, F-1763) at the recommended dilution in FCB. Stained erythrocytes were washed as above and analysed by flow cytometry using a Becton Dickinson FACScalibur. Erythrocytes from each species yielded a single population when analysed for forward and side scattered and for each, the entire population was gated for analysis of fluorescent staining.

Gene Alignments

Human CD59 gene (Ensembl transcript ID ENST00000351554.7) localized to Chromosome 11 (bases 33,708,865-33,736,445, translated from the negative strand) was used as the basis for syntenic and codon-based alignments. The five exons from this region were used for alignment
with the *Cavia porcellus* genome from the Ensembl annotation of the Broad Institute’s CavPor3 assembly, version 90 (24).

**Syntenic Alignments**

As the human *CD59* gene and other mammalian *CD59* genes are flanked by genes *FBXO3* and *C11orf91*, alignments were performed to a chromosomal region of the *Cavia porcellus* genome flanked by their respective gene homologs. This region, characterized by Ensembl as Scaffold DS562947.1 (bases 5,004,217 to 5,025,988), was used for alignment using individual human *CD59* exons.

**Exon Alignments**

All five exons of human *CD59* (Ensembl accession numbers ENSE00001326619, ENSE00001364419, ENSE00000824399, ENSE00000710249, ENSE00002187503, respectively) were aligned with the Ensembl annotated CavPor3 genome assembly, focusing on Scaffold DS562947.1 as described in the previous section. Each alignment was performed using the default parameters of the genomic alignment, changing only the default start/stop sequences in order to match the precise beginning and end of each exon (as indicated in Table I). Sequence similarity was calculated using the ClustalW aligned sequences and the PRABI suite of the Pole Bioinformatique Lyonnais website (https://prabi.ibcp.fr) (25).

**Phylogenetic Analysis**

In order to determine the evolutionary relationship of the *Cavia porcellus CD59* pseudogene with other vertebrate *CD59* genes, a phylogenetic tree was constructed using the putative cDNA
encoding CD59 from 43 species (listed in Supplementary Table I). All cDNA sequences were aligned using ClustalW, and the phylogenetic tree generated using the Neighbor-Joining method in the MEGA6 program suite (26), with bootstrap values n=1000.

Another phylogenetic tree was constructed, based on the taxonomic organization of each of the species described in Supplementary Table I. In this case, the taxonomic tree was generated using the PhyloT (http://phylot.biobyte.de) web-based phylogenetic tree generator (27), and the taxonomic NCBI identifiers for each species.

**RNA-Seq Analysis**

To determine whether any transcriptional products were derived from the identified CD59 pseudogene, alignments were performed using RNA-Seq data, acquired from the Broad Institute’s CavPor 3 assembly (Genbank assembly no. GCF_000151735.1). The reads were analyzed using NCBI’s Genome Data Viewer, using the default aggregate settings of the *Cavia porcellus* annotation release 102. This setting included reads from brain, liver, lung, skeletal muscle, kidney and cervix; tracks displaying RNA-Seq exon coverage were used as a measure of expression, compared with other previously characterized guinea pig complement regulators CD46, CD55 and Crrp (complement receptor-related protein, Crry homolog).
Results:

Flow cytometric analysis of CD59 expression in erythrocytes.

Erythrocytes from guinea pig, normal mouse, CD59a-deficient mouse and human were stained with monoclonal and polyclonal antibodies against mouse CD59a or human CD59. Human erythrocytes were strongly positive with the monoclonal anti-human CD59 MEM43, and moderately strongly with two different polyclonal antibodies raised against mouse CD59a (Figure 1; plate III); mouse erythrocytes stained positively with the rat monoclonal anti-mouse CD59a and with the two polyclonal anti-CD59a antibodies (Figure 1; plate IV); CD59a-deficient mouse erythrocytes used here as a specificity control were negative for all antibodies tested (Figure 1; plate II). Guinea pig erythrocytes were also negative for all antibodies tested (Figure 1; plate I).

Syntenic alignment of human CD59 exons with Cavia porcellus

Initial attempts to identify guinea pig CD59 cDNA utilised BLAST searches against the Cavia porcellus genome, using default parameters derived from mammalian CD59 cDNAs and amino acid sequences. The only homologies identified were members of the Ly6 protein family, leukocyte surface protein lacking the conserved residues and motifs that characterise vertebrate CD59 (28, 29).

The second strategy to locate a putative CD59 gene in the Cavia porcellus genome involved syntenic alignment with the region of human chromosome 11 encoding the CD59 gene. The CD59 gene is located between the FBXO3 and C11orf91 genes in multiple mammalian genomes. Individual human CD59 exons were aligned on a chromosomal scaffold of Cavia porcellus, limited to a region between the FBXO3 and C11orf91 homologs. The results are shown in Table
I. No DNA sequence homologous with human CD59 exon 1 was found, but homologs of human CD59 exons 2-5 were located within the 27.3 kb region, annotated by Ensembl as Scaffold DS562947.1 in the Cavia porcellus genome. Individual alignments of human exon 2, 3 and 4 showed nucleotide identities of 55.8%, 62.3% and 57.8%, respectively. The homolog of human exon 5 showed only 25.8% identity, primarily because the exon 5 homolog was truncated compared to its human counterpart. A schematic diagram of this alignment is shown in Figure 2.

Alignment of human CD59 ORF-encoding exons with Cavia porcellus homologs.

In order to determine the degree of conservation at the protein level between human CD59 and a potential product of the identified guinea pig exon homologs, nucleotide alignments were performed between the open reading frame (ORF)-encoding regions of human CD59 (exons 3, 4 and 5) and the exon homologs identified in the Cavia porcellus genome. As seen in Figure 3, the protein product of human exon 3, encoding the majority of the N-terminal leader peptide, shared a high degree of amino acid identity and similarity to the predicted product of the Cavia porcellus homolog, without any frameshifts (i.e. the sequence translated in a +1 ORF). In particular, the leucine-rich motifs, phenylalanine and cysteine residues were conserved in the predicted Cavia porcellus CD59 leader sequence.

Conservation at the (predicted) protein level was also observed when aligning human CD59 exon 4 and its Cavia porcellus counterpart without any frameshift (i.e. +1 ORF). In the human sequence, exon 4 encodes the last three amino acids of the leader peptide and the first 30 amino acids of the mature protein (starting LQCY); these 30 amino acids include five cysteine residues essential for protein stability, four of which are conserved in the predicted Cavia porcellus
sequence. The first cysteine residue is displaced into the predicted leader sequence in the guinea pig.

Divergence between the two genomes were most pronounced when comparing the protein product of human CD59 exon 5 with the predicted Cavia porcellus exon 5 product. There was some sequence conservation through the first 24 amino acids (residues 31 to 54 in the mature human protein) with conservation of the two cysteine residues in this segment but not the tryptophan residue (W40 in mature human CD59) known to be a key functional residue in multiple species CD59 analogs; indeed, in the selected ORF this (W40 in human) codon is replaced by a stop codon.

Taken together, these results demonstrate that the genomic sequence identified as the sole homolog of CD59 in the Cavia porcellus genome is a pseudogene incapable of producing a functional CD59 protein.

Phylogenetic and comparative analysis of Cavia porcellus CD59 pseudogene with other vertebrate CD59 cDNAs

In order to determine whether the identified Cavia porcellus CD59 pseudogene shares a phylogenetic relationship with other vertebrate CD59 homologs, a phylogenetic tree was generated from CD59 cDNA sequences from diverse species (Figure 4; Supplementary Table I), using a Neighbor-Joining method with bootstrap confidence intervals of n=1000. These results were compared to a pruned phylogenetic tree, based on NCBI taxonomy (not shown); the trees displayed a high degree of similarity between clades, with CD59 conservation distributed along phylogenetic classes (i.e. more similar CD59 sequences clustering along avian, reptilian, fish and mammalian orders). Furthermore, within mammals, CD59 clusters were distributed consistently
Pairwise nucleotide alignment of Cavia porcellus CD59 pseudogene and Heterocephalus glaber CD59 cDNA

A pairwise alignment between the Cavia porcellus CD59 pseudogene and the protein-encoding CD59 cDNA of its most taxonomically related species, the naked mole rat demonstrated a high-degree of conservation throughout the sequence, although a 15-bp gap and two 3-bp gaps were observed in the Cavia porcellus CD59 pseudogene, all consistent with keeping the sequence in-frame.

RNA-Seq Analysis of Cavia porcellus CD59 Pseudogene Transcription.

In order to test whether the Cavia porcellus CD59 pseudogene was transcribed, the predicted sequences from all 4 characterized “exon” domains were aligned with RNA-Seq data acquired from Genbank, based on the Broad Institute’s CavPor3 assembly; and included reads from brain, liver, lung, skeletal muscle, kidney and cervix. Expression was visualized using the NCBI Genome Browser. This analysis was also performed with other known Cavia porcellus complement regulators, CD46, CD55 and Crrp. As seen in Figure 6, the three characterized complement regulators had significant RNA expression throughout the gene, with signal intensities reaching nearly 17000 for CD59, 530 for CD46 and 300 for Crrp. For each of the CD59 pseudogene “exons”, there was essentially no signal for expression, except for a weakly defined 80 bp region (corresponding to Ensembl scaffold DS562947.1: 5065147-5065227),
reaching a signal intensity of \(~10\); this segment corresponds to an intronic area between the regions homologous to exons 4 and 5 of human \(CD59\).
Discussion:

The propensity for guinea pig erythrocytes to be lysed by a variety of different animal sera, and to be effectively protected from lysis by incorporation of CD59 from other species led us to propose that guinea pig erythrocytes lacked a functional CD59 protein rendering them susceptible to lysis by homologous or heterologous complement (17). Although we had no data on other tissues and cell types we further speculated that guinea pigs might have a genetic lack of CD59, an anomaly given the wide species distribution of CD59 analogs.

Here we set out to address this anomaly and provide evidence of the fate of CD59. First, we stained guinea pig erythrocytes with a panel of anti-mouse CD59a and anti-human CD59 antibodies, including two polyclonal antibodies raised against mouse CD59a that showed broad species cross-reactivity, staining both mouse and human erythrocytes; specificity was demonstrated by showing that CD59a-deficient mouse erythrocytes were negative for all antibodies. Guinea pig erythrocytes showed no staining, supporting the contention that they lacked a CD59 analog.

We then conducted an extensive genomic study using an annotated *Cavia porcellus* genome. BLAST alignments using other mammalian *CD59* genes identified only limited homology to genes of the Ly6 superfamily of molecules; no gene possessing the conserved domains of *CD59* was found in the *Cavia porcellus* genome (28, 29). Syntenic alignments were then performed, looking for the *Cavia porcellus* genomic region corresponding to the region of human chromosome 11 containing the *CD59* gene. In mammalian species, the *CD59* gene is flanked by the *FBXO3* and *C11orf91* genes; the *Cavia porcellus* *FBXO3* and *C11orf91* homologs were identified; however, no intact *CD59* gene was located between these two genes. Individual alignments of the five human *CD59* exons across this region identified regions of homology with
exons 2-5, exons 2-4 sharing a high degree of sequence similarity between human and guinea pig with putative in-frame ORFs. In contrast, alignments with human exon 5, which codes for the majority of the mature protein sequence, identified putative regions of homology comprised of two adjacent, truncated regions (together, approximately 60% the size of its human counterpart) that did not contain a complete ORF but contained both non-sense mutations and deletions. In humans, this region encodes for a Cysteine-Asparagine combination that appears to be highly conserved across vertebrate Ly6 family proteins (30). The loss of this highly conserved region and the presence of nonsense and deletion mutations support the assertion that this segment of chromosomal DNA is no longer under functional constraint, but instead represents an inactive pseudogene prone to genetic drift.

RNA-Seq analysis confirmed that no portion of any of the CD59 exon homologs were transcribed. Therefore, from an evolutionary perspective, it is likely that elements associated with the promoter have been lost. The loss of this region, along with the loss of a human exon 1-homologous region would suggest that a deletion event at the 5’-end of the gene may have contributed to the inactivation of a previously functional gene. In contrast, expression of other Cavia porcellus CReg proteins was readily detectable from RNA-Seq, thereby indicating that other complement regulators were still functional, and that these components are likely sufficient to prevent MAC-associated autoimmune damage.

The absence of a region homologous to exon 1 of human CD59 in the guinea pig genome may imply either high genetic drift, eliminating any significant similarity to human CD59 exon 1, or a deletion event removing the exon 1-like region. The second possibility is of particular interest, since it may imply an ancestral region prone to genetic recombination. This might help explain gene duplication events associated with the CD59 gene; CD59 gene duplication has been
described in mice (31), and a putative $CD59$ gene duplication event is reported in Chinese softshell turtle (Ensembl access number ENSPSIG00000007164). In both cases, the gene duplication products are adjacent to each other, while still being syntenic to the $FBXO3$ gene. In the mouse, the two $CD59$ genes are differentially expressed with $CD59a$ broadly distributed in all tissues and $CD59b$ expressed only in male genital tissue (31). The $CD59$ gene duplication product in Chinese softshell turtle lacks the conserved regions associated with a functional $CD59$. Nevertheless, these two instances of gene duplication, combined with the putative gene deletion event in guinea pigs suggests a dynamic genetic element which may exist in $CD59$ genes across all vertebrates.

The fact that intact and functional $CD59$ genes have been found in all other available mammalian genomes suggests that loss of CD59 expression was caused by a mutation unique to guinea pigs (or an ancestral species). From an evolutionary standpoint, the lack of a functional CD59 in guinea pigs adds to the debate on the importance of this molecule in protecting vertebrates from autoimmune damage. CD59 is reported to be ubiquitously expressed on (almost) all cells of mammals (32) and is considered to be indispensable for immune homeostasis. Studies using CD59a-knockout mice demonstrate that CD59 deficiency is compatible with life; the mice showed increased erythrocyte turnover and a compensated anemia, and increased pathology in a number of complement-mediated disease models (33-36). CD59 deficiency in humans has serious consequences; isolated CD59 deficiency in humans was first described over 25 years ago in a Japanese patient presenting in his 20s with symptoms resembling paroxysmal nocturnal hemoglobinuria (37). More recently, CD59 deficiency was reported in several families of North African Jewish ethnicity, all cases presenting in infancy with chronic hemolysis and episodes of peripheral neuropathy resembling Guillain-Barre syndrome (38). All cases were homozygous for...
a missense mutation in the CD59 gene, pCys89Tyr, and had no detectable CD59 on cell surfaces. Disease course was severe with high mortality. Other cases with remarkably similar clinical features but different CD59 mutations were subsequently described (39,40); importantly, treatment with the anti-C5 antibody eculizumab markedly improved both hemolytic and neurological symptoms.

The relatively mild phenotype in CD59a knockout mice has been ascribed to the presence in rodents of another widely expressed complement regulator, Cry; this was shown to be essential for murine erythrocyte protection from complement attack, while CD59 was dispensable (41). To date, no Cry homolog has been reported in guinea pigs, although a truncated, homologous protein, Crp has been described (42). The gene encoding for Crp aligns with murine Cry and may serve a similar function, reducing the impact of the absence of CD59 in guinea pigs. Overall, the redundancy of membrane-bound complement regulators in rodents likely compensates for the lack of functional CD59 in guinea pigs. Nevertheless, from an evolutionary perspective, the results presented in this analysis represent the first example of a species in which the CD59 gene has become a pseudogene, and demonstrates the capacity of a rodent species to regulate complement activation on self-cells independent of this molecule.
Acknowledgements:

The authors would like to thank Dr. Ana M. Aransay (CIC bioGUNE, Derio, Spain) for providing technical advice in the preparation of this manuscript. This work was supported in part by the Wellcome Trust through ISSF funding awarded to Cardiff University.
References:


42. Moore, F. D. Jr. 1991. CRRP: a guinea pig protein, identified by sequence homology to human CR1, which contains two short consensus repeat motifs and appears not to be transmembrane or secreted. *J. Immunol.* 147: 3615-3622.
Figure Legends:

Figure 1. Flow cytometric analysis of CD59 expression on erythrocytes.

Guinea pig (I), CD59 deficient mouse (-CD59) (II), human (III) and wild-type mouse (+CD59) (IV) erythrocytes were stained with monoclonal anti-human CD59 MEM43 (E), monoclonal rat anti-mouse CD59a (F) or two different polyclonal anti-CD59a (G = anti CD59a.1, H = anti CD59a.2) followed by appropriate fluorescein FITC-labelled secondary antibodies, then analyses by flow cytometry. Controls are: A = unstained cells, B = donkey anti-Mouse IgG, C = rabbit anti-rat IgG, D = goat anti-Rabbit IgG; all are overlapping and negative on each of the histograms. Positive staining for CD59 was observed on human erythrocytes and on wild-type mouse erythrocytes (F, G, H) with the species-specific monoclonal and the two polyclonal antibodies (E, G, H). In contrast, guinea pig erythrocytes and erythrocytes from CD59a-deficient mice were negative for all antibodies tested.

Figure 2. Alignment of human CD59 exons to the Cavia porcellus genome.

Nucleotide alignments from the human CD59 gene were compared to a sequenced scaffold annotated from the Ensembl guinea pig browser using the Broad Institute’s CavPor3 genome assembly (see Materials and Methods). The precise degree of homology between segments is described in Table 1.

Figure 3. Nucleotide alignments of the protein-encoding regions of human CD59.

Exons 3, 4 and 5 of human CD59 were compared to homologous chromosomal regions of the Cavia porcellus genome. It should be noted that the codon encoding the terminal glycine in exon 3 is composed of the two terminal nucleotides of exon 3, followed by the first nucleotide of
exon 4. For all exons, the nucleotide sequences were translated into a +1 ORF. Identity (*) or similarity (. or :) between individual amino acid residues is indicated.

Figure 4. *CD59* phylogenetic tree based on NCBI taxonomic classification.
*CD59* coding regions of 44 vertebrate species were compared. The taxonomy ID numbers are indicated in supplemental Table 1. The *CD59* phylogenetic tree was generated using a neighbor-joining method, with bootstrap values from n=1000 indicated. The box indicates that the species most phylogenetically related to guinea pig is the naked mole-rat.

Figure 5. Nucleotide alignment of the *CD59* ORF in naked mole rat and the putative *Cavia porcellus* *CD59* pseudogene.
The nucleotide sequences and derived amino acid sequences for naked mole rat (top) and guinea pig (bottom) *CD59* are shown. Identity (*) or similarity (. or :) between individual amino acid residues is shown.

Figure 6. RNA Seq profiles of guinea pig complement regulator genes.
The location of the *Cavia porcellus* homologs of CD46, CD55 and Crrp were identified in the Ensembl annotation of the Broad Institute’s CavPor3 assembly; and RNA-Seq profiles for these and the genomic area corresponding to the *Cavia porcellus* *CD59* pseudogene were obtained using NCBI Genome Browser viewer. The location of each gene is as follows: CD46-scaffold_DS562867.1: 15,332,888-15,365,235; CD55- scaffold_DS562867.1: 15,598,604-15,622,439; Crrp- scaffold_DS562867.1: 15,425,329-15,451,739; *CD59* pseudogene-scaffold_DS562947.1:5,042,265-5,069,597. Signal intensities are indicated in the diagram.
Taxonomic relationships between species are presented in Figure 4, where a rooted taxonomic tree was generated and compared to a phylogenetic tree based on their respective CD59 cDNAs (see Materials and Methods). All CD59 sequences were obtained from Ensembl, with the exception of the Large Yellow Croaker, Rainbow Trout and

Supplemental Table I - Identification Of Different Animal CD59 cDNAs Used In This Study

<table>
<thead>
<tr>
<th>Common Name</th>
<th>Species Name</th>
<th>NCBI ID</th>
<th>CD59 Acc No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpaca</td>
<td>Vicugna pacos</td>
<td>30538</td>
<td>ENSVPAT0000007804.1</td>
</tr>
<tr>
<td>Bushbaby</td>
<td>Otolemur garnetti</td>
<td>30611</td>
<td>ENSOGAT0000003326.1</td>
</tr>
<tr>
<td>Cat</td>
<td>Felis catus</td>
<td>9685</td>
<td>ENSFCAT00000010141.2</td>
</tr>
<tr>
<td>Chicken</td>
<td>Gallus gallus</td>
<td>9031</td>
<td>ENSGALT00000054780.1</td>
</tr>
<tr>
<td>Chimpanzee</td>
<td>Pan troglodytes</td>
<td>9598</td>
<td>ENSPTRT00000006570.4</td>
</tr>
<tr>
<td>Chinese Softshell Turtle</td>
<td>Pelodiscus sinensis</td>
<td>13735</td>
<td>ENSPSIT00000007520.1</td>
</tr>
<tr>
<td>Cow</td>
<td>Bos taurus</td>
<td>9913</td>
<td>ENSBTAT00000002967.2</td>
</tr>
<tr>
<td>Dog</td>
<td>Canis lupus familiaris</td>
<td>9615</td>
<td>ENSCAFT000000044159.1</td>
</tr>
<tr>
<td>Dolphin</td>
<td>Tursiops truncatus</td>
<td>9739</td>
<td>ENSTRT000000016371.1</td>
</tr>
<tr>
<td>Duck</td>
<td>Anas platyrhynchos</td>
<td>8839</td>
<td>ENSAIPL00000006990.1</td>
</tr>
<tr>
<td>Ferret</td>
<td>Mustela putorius furo</td>
<td>9669</td>
<td>ENSMPUT00000001389.1</td>
</tr>
<tr>
<td>Flycatcher</td>
<td>Ficedula albicollis</td>
<td>59894</td>
<td>ENSFALT00000001389.1</td>
</tr>
<tr>
<td>Gibbon</td>
<td>Nomascus leucogenys</td>
<td>61853</td>
<td>ENSNLET000000032353.1</td>
</tr>
<tr>
<td>Gorilla</td>
<td>Gorilla gorilla gorilla</td>
<td>9595</td>
<td>ENSGGOT000000016902.2</td>
</tr>
<tr>
<td>Guinea Pig</td>
<td>Cavia porcellus</td>
<td>10141</td>
<td>Not Available</td>
</tr>
<tr>
<td>Hedgehog</td>
<td>Erinaceus europaeus</td>
<td>9365</td>
<td>ENSEUT000000014945.1</td>
</tr>
<tr>
<td>Human</td>
<td>Homo sapiens</td>
<td>9606</td>
<td>ENST00000351554.7</td>
</tr>
<tr>
<td>Hyrax</td>
<td>Procavia capensis</td>
<td>9813</td>
<td>ENSPCAT00000000528.1</td>
</tr>
<tr>
<td>Kangaroo Rat</td>
<td>Dipodomys ordii</td>
<td>10020</td>
<td>ENSDOTR000000008867.1</td>
</tr>
<tr>
<td>Large Yellow Croaker</td>
<td>Larimichthys crocea</td>
<td>215358</td>
<td>DQ649025.1</td>
</tr>
<tr>
<td>Macaque</td>
<td>Macaca mulatta</td>
<td>9544</td>
<td>ENSMUMUT00000070699.1</td>
</tr>
<tr>
<td>Marmoset</td>
<td>Callithrix jacchus</td>
<td>9483</td>
<td>ENSCJAT000000020355.1</td>
</tr>
<tr>
<td>Megabat</td>
<td>Pteropus vampyrus</td>
<td>132908</td>
<td>ENSPVAG00000008331.1</td>
</tr>
<tr>
<td>Microbat</td>
<td>Myotis lucifugus</td>
<td>59463</td>
<td>ENSLMTU000000013129.2</td>
</tr>
<tr>
<td>Mouse (a)</td>
<td>Mus musculus</td>
<td>10090</td>
<td>ENSMUST000000168176.7</td>
</tr>
<tr>
<td>Mouse (b)</td>
<td>Mus musculus</td>
<td>10090</td>
<td>ENSMUST000000090429.7</td>
</tr>
<tr>
<td>Mouse Lemur</td>
<td>Microcebus murinus</td>
<td>30608</td>
<td>ENSMICT000000053152.1</td>
</tr>
<tr>
<td>Naked Mole Rat</td>
<td>Heterocephalus glaber</td>
<td>10181</td>
<td>GEF01006011.1</td>
</tr>
<tr>
<td>Olive Baboon</td>
<td>Papio anubis</td>
<td>9555</td>
<td>ENSPANT00000018547.1</td>
</tr>
<tr>
<td>Opossum</td>
<td>Monodelphis domestica</td>
<td>13616</td>
<td>ENSMODT000000035940.1</td>
</tr>
<tr>
<td>Orangutan</td>
<td>Pongo abelii</td>
<td>9601</td>
<td>ENSPPT00000004019.2</td>
</tr>
<tr>
<td>Panda</td>
<td>Aihiropoda melanoleuca</td>
<td>9646</td>
<td>ENSAMET00000008032.1</td>
</tr>
<tr>
<td>Pig</td>
<td>Sus scrofa</td>
<td>9823</td>
<td>ENSSCT000000026205.2</td>
</tr>
<tr>
<td>Pika</td>
<td>Ochotona princeps</td>
<td>9978</td>
<td>ENSOPRT00000009673.1</td>
</tr>
<tr>
<td>Platypus</td>
<td>Ornithorhynchus anatinus</td>
<td>9258</td>
<td>ENSOANT00000032522.1</td>
</tr>
<tr>
<td>Rabbit</td>
<td>Oryctolagus cuniculus</td>
<td>9986</td>
<td>ENSUCUT000000026578.1</td>
</tr>
<tr>
<td>Rainbow Trout</td>
<td>Oncorhynchus mykiss</td>
<td>8022</td>
<td>AY593999</td>
</tr>
<tr>
<td>Rat</td>
<td>Rattus norvegicus</td>
<td>10116</td>
<td>ENSRNOT000000067085.3</td>
</tr>
<tr>
<td>Sheep</td>
<td>Ovis aries</td>
<td>9940</td>
<td>ENSOART000000019487.1</td>
</tr>
<tr>
<td>Sloth</td>
<td>Choloepus hoffmanni</td>
<td>9358</td>
<td>ENSCHT00000001515.1</td>
</tr>
<tr>
<td>Squirrel</td>
<td>Ictidomys tridecemlineatus</td>
<td>43179</td>
<td>ENSSTOT00000013193.1</td>
</tr>
<tr>
<td>Turkey</td>
<td>Meleagris gallopavo</td>
<td>9103</td>
<td>ENSMGAT00000011417.1</td>
</tr>
<tr>
<td>African Green Monkey</td>
<td>Chlorocebus sabaeus</td>
<td>60711</td>
<td>ENSCAT00000003384.1</td>
</tr>
<tr>
<td>Zebra Finch</td>
<td>Taeniopygia guttata</td>
<td>59729</td>
<td>ENSTGUT00000005810.1</td>
</tr>
<tr>
<td>Zebras</td>
<td>Danio rerio</td>
<td>7955</td>
<td>ENSDART000000126737.3</td>
</tr>
</tbody>
</table>
Naked Mole Rat, which were obtained from NCBI’s Genbank. The deduced sequence of *Cavia porcellus CD59* pseudogene is presented in this manuscript.

Figure 1.
Figure 2.

I

II

III

IV

V

Human CD59 Exons

II

III

IV

V

Guinea Pig CD59 Pseudogene Exon-Homologues

V
Figure 1.
Human CD59 Exons

Guinea Pig CD59 Pseudogene Exon-Homologues
Figure 3
Figure 4
Figure 5
**Figure 6**

RNA-seq exon coverage, aggregate (filtered), NCBI Cavia porcellus Annotation Release 102 - log 2 scaled