Published in final edited form as:

Virus Res. 2011 September ; 160(1-2): 120–127. doi:10.1016/j.virusres.2011.05.022.

The E3 CR1-gamma gene in human adenoviruses associated with epidemic keratoconjunctivitis

Christopher M. Robinson^a, Jaya Rajaiya^a, Xiaohong Zhou^a, Gurdeep Singh^a, David W. Dyer^b, and James Chodosh^{a,*}

^aHowe Laboratory, Massachusetts Eye and Ear Infirmary, Department of Ophthalmology, Harvard Medical School, 243 Charles Street, Boston, Massachusetts, 02114, USA

^bDepartment of Microbiology and Immunology, University of Oklahoma Health Sciences Center, 975 N.E. 10th, Oklahoma City, Oklahoma, 73104, USA

Abstract

Human adenovirus species D type 37 (HAdV-D37) is an important etiologic agent of epidemic keratoconjunctivitis. Annotation of the whole genome revealed an open reading frame (ORF) in the E3 transcription unit predicted to encode a 31.6 kDa protein. This ORF, also known as CR1- γ , is predicted to be an integral membrane protein containing N-terminal signal sequence, luminal, transmembrane, and cytoplasmic domains. HAdV-D19 (C), another viral pathogen causing epidemic keratoconjunctivitis, contains an ORF 100% identical to its HAdV-D37 homologue but only 66% identical to other HAdV-D homologues. Kinetics of RNA expression and confirmation of splicing to the adenovirus tripartite leader sequence suggest a role for the protein product of CR1- γ in the late stages of the viral replication cycle. Confocal microscopy is consistent with expression in the cytoplasm. Sequence analysis reveals a hypervariable luminal domain and a conserved cytoplasmic domain. The luminal domain is predicted to contain multiple N-glycosylation sites. The cytoplasmic domain contains a putative protein kinase C phosphorylation site and potential YXX ϕ and dileucine (LL) motifs suggesting a potential role in modification of host proteins.

Keywords

Human adenovirus type 37; CR1-gamma; 31.6K; E3 transcription unit; epidemic keratoconjunctivitis

1. Introduction

The E3 transcription unit of human adenoviruses (HAdVs) encodes several proteins that have been shown to modulate the host immune response. Three open reading frames (ORFs), RID- α , RID- β , and 14.7K, are conserved across every species of HAdV. Gene products of these ORFs block apoptosis by down-regulation of FAS, TRAIL 1, and TRAIL

© 2011 Elsevier B.V. All rights reserved.

*Corresponding author. Tel. 617-573-3311; fax: 617-573-4290. james_chodosh@meei.harvard.edu.

Present address: Massachusetts Eye and Ear Infirmary, Howe Laboratory, Department of Ophthalmology, Harvard Medical School, 243 Charles Street, Boston, MA 02114.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

2 receptors and blockade of TNF- α (Benedict et al., 2001; Hilgendorf et al., 2003). Another ORF within the E3 region encodes glycoprotein (gp)19K, known to block presentation of MHC class I molecules on the cell surface (Burgert et al., 1987; Deryckere and Burgert, 1996; Feuerbach et al., 1994), and has homologues in HAdV-B, -C, -D, and -E. While labeled as an “early” transcription region, transcripts from the E3 family are expressed both early and late during viral infection (Bhat and Wold, 1986; Chow and Broker, 1978; Chow et al., 1979; Chow et al., 1980; Chow et al., 1977). However, proteins encoded in this region do not appear to be required for viral replication (Wold and Gooding, 1991).

While several proteins within the adenovirus genome have been extensively studied, many of the genes located within the E3 region remain poorly characterized. The role of those few E3 proteins with known function were determined mainly in HAdV-C. The E3 transcription unit, though, represents an area of major divergence between HAdV-D and other HAdV species (Burgert and Blusch, 2000; Robinson et al., 2008; Robinson et al., 2009b). The number of potential genes within the E3 region also differs among HAdV species. For HAdV-Ds, E3 is predicted to encode 8 ORFs encompassing approximately 5000 bps. Four of those genes have no known function.

HAdV-D19 and HAdV-D37, along with types D8, D53, D54, and D56 are all etiological agents of epidemic keratoconjunctivitis (EKC), a common and highly communicable eye infection. HAdV-D19 and HAdV-D37 are also isolated from the genitourinary tract (de Jong et al., 1981; Phillips et al., 1982; Swenson et al., 1995). Recently our lab has sequenced the complete genome of HAdV-D37, and of a clinical isolate of HAdV-D19 (strain C) causing EKC, termed HAdV-D19 (C) (Robinson et al., 2008; Robinson et al., 2009b). During annotation of these genomes we identified a putative ORF located between the CR1- β (also known as 49K) and RID- α predicted to encode a protein of 31.6 kDa in size. This gene has been previously identified as 31.6K or CR1- γ (Burgert and Blusch, 2000; Davison et al., 2003). Here we describe analysis of the CR1- γ gene from HAdV-D37 and contrast it with homologues from other HAdV-D genomes. Analysis of RNA expression kinetics reveals that the gene is expressed from the tripartite leader sequence, suggesting a function in the late stages of the viral replication cycle. *In silico* analysis suggests an integral membrane protein that is highly glycosylated, similar to other E3 proteins. We also show diversity between homologues of the gene within HAdV-D, suggesting that immune pressure may have driven the evolution of this gene.

2. Materials and Methods

2.1 Cells, virus stock, infection

HAdV-D37 strain GW was obtained from the American Type Culture Collection (ATCC, Manassas, VA). Virus stocks were grown in A549 cells (CCL-185), a human alveolar epithelial cell line. A549 cells were infected with HAdV-D37 at a multiplicity of infection (MOI) of either 1 (for northern blot and immunofluorescence) or 5 (for RT-PCR) in Dulbecco's modified eagle medium (DMEM), supplemented with 2% fetal bovine serum (FBS), penicillin G sulfate, and streptomycin and incubated at 37°C. One hour post infection, cells were washed twice with 1X PBS, and fresh DMEM (supplemented with 2% FBS, penicillin G sulfate, and streptomycin) added. Cultures were allowed to incubate at 37°C until indicated time points.

2.2 RNA Isolation

Total RNA was isolated using TRIZOL (Invitrogen, Carlsbad, CA) following the manufacturers instructions. To remove any genomic DNA contamination, RNA was treated with Turbo DNase (Ambion, Austin, TX). RNA samples were analyzed on a Bio-Rad

Smart-Spec Plus (Bio-Rad, Hercules, CA) spectrophotometer for concentration and purity. Elimination of DNA was confirmed by absence of visible bands for DNase treated RNA used as template (no RT-control).

2.3 Reverse Transcription PCR

RNA (2 μ g), oligo(d)T, RNasin, and Moloney Murine Leukemia Virus reverse transcriptase (M-MLV RT) (Promega, Madison, WI) were used to generate cDNA in a 20 μ l total volume following the manufacturer's recommended protocol. The cDNA product (2 μ l) was amplified by PCR in a total volume of 25 μ l of PCR mix, including 12.5 μ l of 2 $\times$ PCR Master Mix (Promega), 8.5 μ l ddH₂O, and 1 μ l of each primer (10 pmols). The primers are described in Table 1. The reaction mixtures were heated to 94°C for 5 minutes for the initial denaturing step, followed by 30 cycles of 94°C for 30 s, the annealing temperature for 30 s, and 72°C for 30 s. Cycling was followed by final extension at 72°C for 5 min and then kept at 4°C until analysis. PCR products were analyzed by agarose gel electrophoresis in Tris-acetate-EDTA (TAE) buffer. PCR products were visualized after ethidium bromide staining using a Kodak Image Station (Kodak, Medfield, MA). RT-PCR products of interest were gel purified using QIAquick Gel Extraction Kit (Qiagen, Valencia, CA) and sequenced at the Massachusetts General Hospital DNA Sequencing Core Facility, Harvard Medical School.

2.4. Northern Blot

Total RNA (5 μ g) was analyzed using NorthernMax (Ambion). Three volumes of formaldehyde loading dye and ethidium bromide (0.5 μ l) was added to total RNA and denatured at 65°C for 15 minutes. A Millennium™ Marker-Formamide (Ambion) was also prepared by denaturing at 80°C for 10 minutes and used to determine the size of RNA transcripts. A 1% agarose denaturing gel was prepared as described in the manufacturer's protocol. RNA and the marker were loaded to agarose gel and subjected to electrophoresis at 65V for 90 minutes in 1 $\times$ 3-(N-morpholino)propanesulfonic acid (MOPS) gel running buffer. The gel was examined using a UV light to visualize the ribosomal (rRNA) bands for degradation. RNA was transferred to BrightStar®-Plus Positively Charged Nylon Membrane (Ambion) using the manufacturer's protocol.

DNA probes were produced from cDNA products using RT-PCR. Primers used for these reactions are described in Table 1. cDNA products were biotinylated using a North2South Biotin Random Prime DNA Labeling Kit (Pierce Thermo Fisher Scientific, Rockford, IL), following the manufacturer's protocol.

RNA membranes were pre-hybridized in Church's buffer (Sodium Phosphate buffer (0.5M pH 7.2), EDTA, BSA, SDS) with denatured Salmon Sperm DNA (10 minutes at 95°C) (Trevigen, Gaithersburg, MD) at 47°C for 1 hour. Following pre-hybridization, fresh Church's buffer was prepared along with denatured Salmon Sperm DNA and a DNA probe (1/300 dilution) and added to the membranes to hybridize. Hybridization was carried out at 47°C and incubated overnight. To visualize hybridization of biotinylated DNA probes with target RNA transcripts, Chemiluminescence Nucleic Acid Detection Module (Pierce Thermo Fisher Scientific) was used as per the manufacturer's protocol. Film was exposed to the membranes for 30 seconds, 1 minute, or 5 minutes and processed on a Summit QCP X-ray film processor (Summit Industries, Chicago, IL).

2.5. Immunofluorescence

A549 cells were seeded on 2-well chamber slides (Lab-Tek, Rochester, NY) in DMEM supplemented with 2% FBS, penicillin G sulfate, and streptomycin. Infection with HAdV-D37 was carried out as described above. At 24 hours post-infection, cells were washed in phosphate-buffered saline (PBS) and fixed with 3% paraformaldehyde for 20 minutes at

room temperature. Cells were permeabilized with 0.2% saponin in PBS with 5% FBS used to block nonspecific binding, and then treated with 5% FBS in PBS for 30 minutes. Cells were incubated with the primary antibody or a non-specific rabbit polyclonal IgG antibody control (Abcam, Cambridge, MA) diluted in 5% FBS in PBS overnight, washed three times in PBS, and incubated with the secondary antibody (fluorescein-conjugated goat anti-rabbit immunoglobulin G) for 1 hour. To-Pro-3 in PBS was added to the cells and incubated for 10 minutes at room temperature to stain cell nuclei. Slides were mounted and analyzed on a scanning confocal microscope (Leica SP5, Leica Microsystems, Bannockburn, IL)

2.6. Phylogenetic and sequence analysis

Sequences were aligned using the ClustalW (Larkin et al., 2007) option within the software Molecular Evolutionary Genetics Analysis (MEGA) 4.0.2 (<http://www.megasoftware.net/index.html>) (Tamura et al., 2007). Phylogenetic analysis was performed using bootstrap-confirmed neighbor-joining trees (500 replicates) also designed with MEGA 4.0.2. Branches with bootstrap values below 70 (indicative of low confidence) were collapsed. Predicted isoelectric points (pI) and molecular weight (MW) was determined with the ExPASy Proteomics Server from the Swiss Institute of Bioinformatics (SIB) (http://ca.expasy.org/tools/pi_tool.html) (Gasteiger et al., 2003). Simplot analysis was performed using Simplot 3.5.1 software (<http://sray.med.som.jhmi.edu/SCRSoftware/simplot/>) (Lole et al., 1999) with a window size of 100 bps. HAdV-D37 was used as the reference sequence for analysis.

2.7 Protein analysis and modeling

Multi-sequence amino acid alignment was performed using CLC sequence viewer 6 (<http://www.clcbio.com/index.php?id=28>) (CLC bio, Cambridge, MA). Sequenced-based structural motifs were determined using PredictProtein (<http://www.predictprotein.org/>) (Rost et al., 2004) and SignalP 3.0 Server (<http://www.cbs.dtu.dk/services/SignalP/>) (Emanuelsson et al., 2007). Due to the lack of similarity to known protein structures in the protein database, remote homology using the CPHModels 3.0 server (<http://www.cbs.dtu.dk/services/CPHmodels/>) was used to build 3D models of the putative HAdV-D37 CR1- γ protein (Nielsen et al., 2010). This incorporates a scoring matrix including sequence profiles and predicted structural features for modeling. The luminal domain and the transmembrane and cytoplasmic domains were modeled independently using remote homology. Ramachandran Plot analysis was also used to help verify the models (<http://molprobity.biochem.duke.edu/>) (Chen et al., 2010). UCSF Chimera software (<http://www.cgl.ucsf.edu/chimera>) (Pettersen et al., 2004) was used for visualization and analysis.

3. Results

3.1. CR1- γ in HAdV-D19 (C) and HAdV-D37 is identical, but in other HAdV-D is highly divergent

Previously, HAdV-D37 was completely sequenced by our lab (GenBank Accession: DQ900900) (Robinson et al., 2008). During the annotation of this genome, an ORF located in the E3 transcription unit was identified, previously defined as 31.6K (based on the predicted size of the putative protein) (Blusch et al., 2002; Burgert and Blusch, 2000). Homologues of this gene have been predicted in all HAdV-D and HAdV-E genomes. This gene was previously given the name CR1- γ based on a conserved domain (conserved region 1, CR1) previously shown to be present in human cytomegalovirus RL11 family proteins and a subset of other human adenovirus E3 membrane glycoproteins (Davison et al., 2003; Deryckere and Burgert, 1996) (Fig. 1). Sequencing of an HAdV-D19 clinical isolate (HAdV-D19 (C)) (Robinson et al., 2009b), also a known etiological agent of EKC, revealed

a similar ORF with 100% identity to the HAdV-D37 CR1- γ gene but only 66% identity to other HAdV-D homologues (data not shown). Phylogenetic analysis revealed a close similarity between HAdV-D19 (C), HAdV-D22, and HAdV-D37, but divergence with the EKC viruses HAdV-D8, D53, D54, and D56 (Fig. 2). Predicted isoelectric points (pI) and molecular weight (MW) also reveal a highly diverse protein homologue in HAdV-D genomes (Table 2). Predicted pI values range from 6.09 in HAdV-D26 to 9.62 in HAdV-D54, while predicted MW ranges from 28.6 kDa in HAdV-D26 to 32.5 kDa in HAdV-D48.

3.2. CR1- γ is predicted to be a type 1 integral membrane protein with a glycosylated and hypervariable luminal domain

Since many of the proteins located within the E3 transcription unit have been shown or predicted to be glycosylated, membrane bound proteins, PredictProtein analysis was run on the putative primary amino acid structure. Analysis revealed a putative N-terminal signal sequence and suggested the CR1- γ protein to be a type 1 integral membrane protein based on the predicted transmembrane domain located in the C-terminal region of the protein (Fig. 3B). This transmembrane domain separates a luminal domain and a cytoplasmic domain. Analysis of the luminal domain revealed nine possible N-glycosylation sites. The cytoplasmic domain contains a predicted protein kinase C phosphorylation site and potential YXX ϕ and LL motifs. Simplot analysis of the gene and its HAdV-D homologues demonstrated a high degree of divergence. Specifically, the luminal domain appears hypervariable in contrast to the more conserved cytoplasmic domain (Fig. 3A).

3.2. CR1- γ is expressed from the adenovirus major late promoter during later stages of infection

To confirm presence of RNA from this CR1- γ , RT-PCR was performed on total RNA isolated from HAdV-37 infected A549 cells using primers shown in Figure 4A and Table 1. Primers chosen from within the CR1- γ ORF identified a transcript expressed early post-infection (4 hours), similar to another early gene, E1A 13S (Fig. 4B). Since some E3 genes have been shown to be spliced to the tripartite leader sequence (TPL) and expressed at later time points during infection, we designed a forward primer for RT-PCR in the tripartite leader sequence (Fig. 4A). Analysis confirmed that CR1- γ is indeed expressed from the major late promoter, spliced to the tripartite leader sequence, and apparent at 16 hours post-infection analogous to a known adenovirus late gene product, the hexon gene (Fig. 4B). DNA sequencing of the RT-PCR product confirmed that the 5' end of this RNA contained the tripartite leader sequence located upstream of the annotated ATG start site (Fig. 4C).

3.3. Cytoplasmic expression of the CR1- γ protein in A549 cells

To test for possible expression of the CR1- γ protein in HAdV-D37 infected cells, a rabbit affinity-purified polyclonal antibody was designed using keyhole limpet hemocyanin (KLH) conjugated to a peptide (CKAREKSRRIYRPV) within the predicted CR1- γ protein. The peptide chosen for the antibody is located on the putative C-terminal tail of the protein. While western blot analysis was unsuccessful using this antibody, confocal analysis of A549 cells infected with HAdV-D37 first revealed cytoplasmic expression of CR1- γ at 24 hours post-infection (Fig. 5).

3.4. CR1- γ is expressed late and undergoes complex E3 splicing

Since confocal and western blot analyses were not conclusive, we sought to further characterize the RNA transcript from CR1- γ by northern blot. Consistent with our RT-PCR results, northern blot analysis using a probe designed internally for the CR1- γ gene revealed bands as early as 4 hours post-infection (Fig. 6). Two bands were detected at 4 hours and 8 hours post-infection followed by multiple bands at 16–48 hours post-infection. A TPL/CR1-

γ designed probe produced bands detected at 24 and 48 hours post-infection, confirming our results with RT-PCR.

3.5. CR1- γ predicted 3D structure

The 3D structure for the luminal domain and transmembrane and cytoplasmic domains were predicted using remote homology (Fig. 7). These models were analyzed by Ramachandran Plot analysis using MolProbity. For the luminal domain, Ramachandran Plot analysis revealed that among the 188 residues, 159 (84.6%) were in a favored region, 20 (10.6%) were in an allowed region, and 9 (4.8%) were in an outlier region (data not shown). Analysis of the transmembrane and cytoplasmic domains revealed that among the 35 residues in these two regions, 30 (85.7%) were in a favored region, 2 (5.7%) were in an allowed region, and 3 (8.6%) were in an outlier region. Analysis of the conserved residues in HAdV-D37 compared to other EKC viruses identified 142 identical residues (Fig. 7 and Supplementary Fig. 1). Identical amino acids are mainly found in the N-terminal signal sequence, transmembrane and cytoplasmic domains. The luminal domain appears to be highly diverse and contains the most non-conserved residues (Fig. 7). Several of these non-conserved regions can be attributed to three insertions of residues at positions 19–21, 143–153, and 164–177 and a deletion between residues 96 and 97 (Supplemental Fig. 1). Analysis of the insertion at 143–153 reveals a 5 amino acid repeat (TTQPTT).

Discussion

The E3 transcription unit within HAdVs demonstrates not only major interspecies divergence but also surprising intraspecies divergence (Robinson et al., 2009a; Robinson et al., 2008; Robinson et al., 2009b; Robinson et al., 2011b). While E3 protein products are dispensable for viral replication *in vitro*, they have remained a part of the genome throughout the evolution of adenoviruses, suggesting an important function during the virus replication cycle in the living host. Diversity in the E3 region provides reason to believe that these proteins may help define tissue tropisms unique to specific HAdV types (Burgert and Blusch, 2000). We also suggest that HAdVs may have added immunomodulatory functions to successive HAdV species by the addition of E3 genes in order to evade immune pressure by a concurrently evolving host.

While many genes within the E3 region have been characterized, the function of some remains unknown. Herein, we describe the CR1- γ gene in HAdV-D37 and the identical homologue in a clinical isolate of HAdV-D19. Highly variable homologues of this gene are found in all HAdV-D genomes sequenced to date. In other EKC-causing HAdVs (D8, D54, and D56), CR1- γ genes are dissimilar to those of HAdV-D19 (C) and HAdV-D37. Conserved amino acids may play a role in pathogenesis, and/or dissimilar amino acids may determine previously unrecognized differences in virulence among EKC-associated HAdVs. The only homologue of the CR1- γ gene in other HAdV species is found in HAdV-E4 (termed CR1- δ or 30K).

Simplot analysis and close examination of the predicted primary amino acid sequence of CR1- γ reveal several potential regions of interest. Using PredictProtein, the putative amino acid sequence is predicted to contain a N-terminal signal sequence, a luminal domain, a transmembrane domain, and a cytoplasmic domain consisting of 39 residues. Nine N-glycosylation sites were predicted on the luminal domain, consistent with other glycosylated E3 proteins (Blusch et al., 2002; Frietze et al., 2010; Hawkins and Wold, 1995; Li and Wold, 2000; Windheim and Burgert, 2002). A protein kinase C phosphorylation site, located within the cytoplasmic domain, suggests that this protein may function through PKC to modulate other host proteins. Motifs for YXX ϕ and LL were also predicted in this domain suggesting that the protein traffics through the endosomal and/or lysosomal pathways by a

possible interaction with adapter protein (AP) complexes AP-1, AP-2, AP-3, and/or AP-4 [reviewed in (Bonifacino and Traub, 2003)]. These complexes are found in the cytoplasm (AP-2) and *trans*-Golgi network (AP-1, -3, and -4), which suggest a locational target of putative CR1- γ protein. These motifs have been reported for E3 proteins, RID- α and RID- β , which down-regulate apoptosis receptors, and the HAdV-D CR1- β (also known as 49K) protein (Elsing and Burgert, 1998; Hilgendorf et al., 2003; Lichtenstein et al., 2002; Windheim and Burgert, 2002). These data suggest that CR1- γ may be yet another E3 protein that takes advantage of the intracellular trafficking components as a means for immune evasion. Interestingly, Simplot analysis of the gene revealed areas of both diversity and conservation within the putative protein. The luminal domain appears to be highly variable while transmembrane and cytoplasmic domains are conserved. This suggests that the luminal domain is under immune pressure, and may implicate CR1- γ as important to viral tropism or virulence in the host.

RNA transcripts from the CR1- γ gene were detected with RT-PCR analysis both early *and* late during viral infection. Sequencing of the 5' end of RT-PCR transcripts revealed splicing to the tripartite leader sequence suggesting the gene is expressed from the major late promoter. This is consistent with other E3 proteins previously shown to be expressed at both early and late time points, and when late, found to be spliced to the tripartite leader (Blusch et al., 2002; Fietze et al., 2010; Li and Wold, 2000; Windheim and Burgert, 2002). Northern blot analysis using a probe targeted to the interior of the CR1- γ gene confirmed the kinetics of expression, but also revealed multiple bands increasing from 4 to 48 hours post-infection. This data suggests that the E3 transcription unit in HAdV-D37 may undergo complex splicing, similar to that previously described for HAdV-C types (Bhat and Wold, 1986; Chow and Broker, 1978; Chow et al., 1979; Chow et al., 1980; Chow et al., 1977). Attempts to identify the splicing pattern computationally were unsuccessful. It remains to be determined for HAdV-Ds if multiple RNA transcripts are produced via alternative splicing. Protein expression from the CR1- γ gene was suggested by confocal microscopy with an antibody raised against a peptide located within the putative HAdV-D37 CR1- γ protein. Following 24 hours of infection, HAdV-D37 CR1- γ expression was observed in the cytoplasm. Based on the predicted protein structure, this data suggests possible retention in intracellular membranes such as the endoplasmic reticulum or golgi apparatus.

In summary, we show that CR1- γ transcripts are expressed both early and late in viral infection. The protein may be subject to complex post-transcriptional modifications, and be expressed in membrane-bound organelles within the cytoplasm of infected cells. These studies serve as a foundation for further studies of CR1- γ , and its potential role in HAdV pathogenesis.

Highlights

Human adenovirus species D type 37 (HAdV-D37) is an important etiologic agent of epidemic keratoconjunctivitis. Annotation of the whole genome revealed an open reading frame (ORF) also known as CR1- γ , in the E3 transcription unit. Predicted to be an integral membrane protein containing N-terminal signal sequence, luminal, transmembrane, and cytoplasmic domains, the kinetics of RNA expression and confirmation of splicing to the adenovirus tripartite leader sequence suggest a role for the protein product of CR1- γ in the late stages of the viral replication cycle.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Supported by NIH grants EY013124 and P30EY014104, and an unrestricted grant to the Department of Ophthalmology, Harvard Medical School from Research to Prevent Blindness, Inc. The funding sponsors played no role in any aspect of the study.

References

- Benedict CA, Norris PS, Prigozy TI, Bodmer JL, Mahr JA, Garnett CT, Martinon F, Tschopp J, Gooding LR, Ware CF. Three adenovirus E3 proteins cooperate to evade apoptosis by tumor necrosis factor-related apoptosis-inducing ligand receptor-1 and -2. *J Biol Chem*. 2001; 276(5): 3270–3278. [PubMed: 11050095]
- Bhat BM, Wold WS. Genetic analysis of mRNA synthesis in adenovirus region E3 at different stages of productive infection by RNA-processing mutants. *J Virol*. 1986; 60(1):54–63. [PubMed: 3747028]
- Blusch JH, Deryckere F, Windheim M, Ruzsics Z, Arnberg N, Adrian T, Burgert HG. The novel early region 3 protein E3/49K is specifically expressed by adenoviruses of subgenus D: implications for epidemic keratoconjunctivitis and adenovirus evolution. *Virology*. 2002; 296(1):94–106. [PubMed: 12036321]
- Bonifacino JS, Traub LM. Signals for sorting of transmembrane proteins to endosomes and lysosomes. *Annu Rev Biochem*. 2003; 72:395–447. [PubMed: 12651740]
- Burgert HG, Blusch JH. Immunomodulatory functions encoded by the E3 transcription unit of adenoviruses. *Virus Genes*. 2000; 21(1–2):13–25. [PubMed: 11022786]
- Burgert HG, Maryanski JL, Kvist S. "E3/19K" protein of adenovirus type 2 inhibits lysis of cytolytic T lymphocytes by blocking cell-surface expression of histocompatibility class I antigens. *Proc Natl Acad Sci U S A*. 1987; 84(5):1356–1360. [PubMed: 2950523]
- Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*. 2010; 66(Pt 1):12–21. [PubMed: 20057044]
- Chow LT, Broker TR. The spliced structures of adenovirus 2 fiber message and the other late mRNAs. *Cell*. 1978; 15(2):497–510. [PubMed: 719751]
- Chow LT, Broker TR, Lewis JB. Complex splicing patterns of RNAs from the early regions of adenovirus-2. *J Mol Biol*. 1979; 134(2):265–303. [PubMed: 537065]
- Chow LT, Lewis JB, Broker TR. RNA transcription and splicing at early and intermediate times after adenovirus-2 infection. *Cold Spring Harb Symp Quant Biol*. 1980; 44(Pt 1):401–414. [PubMed: 6933037]
- Chow LT, Roberts JM, Lewis JB, Broker TR. A map of cytoplasmic RNA transcripts from lytic adenovirus type 2, determined by electron microscopy of RNA:DNA hybrids. *Cell*. 1977; 11(4): 819–836. [PubMed: 890740]
- Davison AJ, Akter P, Cunningham C, Dolan A, Addison C, Dargan DJ, Hassan-Walker AF, Emery VC, Griffiths PD, Wilkinson GW. Homology between the human cytomegalovirus RL11 gene family and human adenovirus E3 genes. *J Gen Virol*. 2003; 84(Pt 3):657–663. [PubMed: 12604818]
- de Jong JC, Wigand R, Wadell G, Keller D, Muzerie CJ, Wermenbol AG, Schaap GJ. Adenovirus 37: identification and characterization of a medically important new adenovirus type of subgroup D. *J Med Virol*. 1981; 7(2):105–118. [PubMed: 7264611]
- Deryckere F, Burgert HG. Early region 3 of adenovirus type 19 (subgroup D) encodes an HLA-binding protein distinct from that of subgroups B and C. *J Virol*. 1996; 70(5):2832–2841. [PubMed: 8627757]
- Elsing A, Burgert HG. The adenovirus E3/10.4K–14.5K proteins down-modulate the apoptosis receptor Fas/Apo-1 by inducing its internalization. *Proc Natl Acad Sci U S A*. 1998; 95(17): 10072–10077. [PubMed: 9707602]
- Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc*. 2007; 2(4):953–971. [PubMed: 17446895]

- Feuerbach D, Etteldorf S, Ebenau-Jehle C, Abastado JP, Madden D, Burgert HG. Identification of amino acids within the MHC molecule important for the interaction with the adenovirus protein E3/19K. *J Immunol.* 1994; 153(4):1626–1636. [PubMed: 8046235]
- Frietze KM, Campos SK, Kajon AE. Open reading frame E3–10.9K of subspecies B1 human adenoviruses encodes a family of late orthologous proteins that vary in their predicted structural features and subcellular localization. *J Virol.* 2010; 84(21):11310–11322. [PubMed: 20739542]
- Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 2003; 31(13):3784–3788. [PubMed: 12824418]
- Hawkins LK, Wold WS. The E3–20.5K membrane protein of subgroup B human adenoviruses contains O-linked and complex N-linked oligosaccharides. *Virology.* 1995; 210(2):335–344. [PubMed: 7618271]
- Hilgendorf A, Lindberg J, Ruzsics Z, Honing S, Elsing A, Lofqvist M, Engelmann H, Burgert HG. Two distinct transport motifs in the adenovirus E3/10.4–14.5 proteins act in concert to down-modulate apoptosis receptors and the epidermal growth factor receptor. *J Biol Chem.* 2003; 278(51):51872–51884. [PubMed: 14506242]
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal X version 2.0. *Bioinformatics.* 2007; 23(21):2947–2948. [PubMed: 17846036]
- Li Y, Wold WS. Identification and characterization of a 30K protein (Ad4E3–30K) encoded by the E3 region of human adenovirus type 4. *Virology.* 2000; 273(1):127–138. [PubMed: 10891415]
- Lichtenstein DL, Krajcsi P, Esteban DJ, Tollefson AE, Wold WS. Adenovirus RIDbeta subunit contains a tyrosine residue that is critical for RID-mediated receptor internalization and inhibition of Fas- and TRAIL-induced apoptosis. *J Virol.* 2002; 76(22):11329–11342. [PubMed: 12388693]
- Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, Ingersoll R, Sheppard HW, Ray SC. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol.* 1999; 73(1):152–160. [PubMed: 9847317]
- Nielsen M, Lundegaard C, Lund O, Petersen TN. CPHmodels-3.0--remote homology modeling using structure-guided sequence profiles. *Nucleic Acids Res.* 2010; 38(Web Server issue):W576–W581. [PubMed: 20542909]
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem.* 2004; 25(13):1605–1612. [PubMed: 15264254]
- Phillips PA, Harnett GB, Gollow MM. Adenovirus type 19 and a closely related new serotype in genital infection. *Br J Vener Dis.* 1982; 58(2):131–132. [PubMed: 7066666]
- Robinson CM, Rajaiya J, Walsh MP, Seto D, Dyer DW, Jones MS, Chodosh J. Computational analysis of human adenovirus type 22 provides evidence for recombination among species D human adenoviruses in the penton base gene. *J Virol.* 2009a; 83(17):8980–8985. [PubMed: 19553309]
- Robinson CM, Seto D, Jones MS, Dyer DW, Chodosh J. Molecular evolution of human species D adenoviruses. *Infect Genet Evol.* 2011a
- Robinson CM, Shariati F, Gillaspay AF, Dyer DW, Chodosh J. Genomic and bioinformatics analysis of human adenovirus type 37: new insights into corneal tropism. *BMC Genomics.* 2008; 9:213. [PubMed: 18471294]
- Robinson CM, Shariati F, Zaitshik J, Gillaspay AF, Dyer DW, Chodosh J. Human adenovirus type 19: Genomic and bioinformatics analysis of a keratoconjunctivitis isolate. *Virus Res.* 2009b; 139(1): 122–126. [PubMed: 19000724]
- Robinson CM, Singh G, Henquell C, Walsh MP, Peigue-Lafeuille H, Seto D, Jones MS, Dyer DW, Chodosh J. Computational analysis and identification of an emergent human adenovirus pathogen implicated in a respiratory fatality. *Virology.* 2011b; 409(2):141–147. [PubMed: 21056888]
- Rost B, Yachdav G, Liu J. The PredictProtein server. *Nucleic Acids Res.* 2004; 32(Web Server issue):W321–W326. [PubMed: 15215403]

- Swenson PD, Lowens MS, Celum CL, Hierholzer JC. Adenovirus types 2, 8, and 37 associated with genital infections in patients attending a sexually transmitted disease clinic. *J Clin Microbiol.* 1995; 33(10):2728–2731. [PubMed: 8567914]
- Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol.* 2007; 24(8):1596–1599. [PubMed: 17488738]
- Windheim M, Burgert HG. Characterization of E3/49K, a novel, highly glycosylated E3 protein of the epidemic keratoconjunctivitis-causing adenovirus type 19a. *J Virol.* 2002; 76(2):755–766. [PubMed: 11752165]
- Wold WS, Gooding LR. Region E3 of adenovirus: a cassette of genes involved in host immunosurveillance and virus-cell interactions. *Virology.* 1991; 184(1):1–8. [PubMed: 1831308]

Fig 1.

Annotation of CR1- γ in HAdV-D. Comparison of gene homologues of the E3 transcription unit in each species of HAdV. Annotation of CR1- γ and its homologue, CR1- δ , in HAdV-E is shown in purple. Gaps appearing in the E3 transcription map of HAdV-C represent alternative splicing. (Adapted from (Robinson et al., 2011a) with permission.)

Fig 2.
Phylogenetic analysis CR1- γ . Bootstrap neighbor joining tree designed from MEGA 4.0.2. Branches with bootstrap values below 70 were collapsed and values for those branches are not shown. Four Simian adenoviruses (SAdV) have been previously classified within HAdV-E.

A.

B.

MKAFTACVLISLVTLSVA^{IKNQYHVH}NVTRDGYITL^{NVTID}NTTWTRYHLNKWHQICTWSDPSYKCHS^{NGSI} 72
 TIHAF^{NIT}SGQYKAESFTNWFRYYGNHKHEIHIF^{NITV}IEHPTTKAPTTANTATSIKSTTTQPTTRETTQPTTTVSTTTETTQ 156
 TTQLDTTVQ^{NSTV}LVRYLLREESTTEQTEATSSAFSSTA^{NLTS}LAWT^{NETG}VSLMNHQPFSGLD^{IQITFLV}VCGIFILV 241
 VLLYFV^{CCKAREK}^{SR}RPI^{YRPV}IGEPQLQVEGGLRN^{LL}FSFSVW 280

Fig 3.

Sequence and primary amino acid structure analysis. (A) Simplot analysis comparing the HAdV-D37 CR1- γ ORF within HAdV-D and HAdV-E. Each colored line represents a homologous gene. The predicted structural domains of the putative protein are designated above the plot. The predicted signal peptide (SP) is shown in black. The predicted transmembrane domain (TM) is shown in blue. Two red boxes represent the predicted luminal terminal (LD) and cytoplasmic domains (CD). (B) Primary amino acid analysis of HAdV-D37 CR1- γ . Red indicates predicted N-glycosylation sites. Blue represents a predicted transmembrane domain. Green represents a predicted protein C phosphorylation site. Predicted YXX ϕ and LL motifs are represented in orange.

Fig. 4. RT-PCR analysis of RNA expression. (A) Schematic of primer design for RT-PCR and Northern blot analysis. Red arrows represent primers described as TPL/ CR1-γ where the forward and reverse primers are located within the adenovirus tripartite leader and within the CR1-γ gene, respectively. Blue arrows represent primers designated as CR1-γ internal where forward and reverse primers were located exclusively within the CR1-γ gene. (B) RT-PCR analysis of A549 cells infected with HAdV-D37 at a MOI of 5. Total RNA was harvested at 0, 2, 4, 8, 16, 24, and 48 hours post-infection. Primer pairs were as follows: E1A 13S forward and E1A 13S reverse (E1A 13S), CR1-γ forward and CR1-γ reverse (CR1-γ Internal), TPL2 forward and CR1-γ reverse (TPL/CR1-γ), TPL2 forward and hexon reverse

(TPL/Hexon), GAPDH forward and GAPDH reverse (GAPDH). GAPDH was used as a positive control for reverse transcription. E1A 13S and TPL/Hexon were used for positive controls for both early and late viral gene expression, respectively. (C) TPL/CR1- γ PCR products were gel purified and sequenced. The predicted splice acceptor site (AG) is shown in large italics. Red indicates the annotated CR1- γ ATG start site.

Fig. 5.

Immunofluorescence staining of HAdV-D37 CR1- γ . A549 cells were infected with HAdV-D37 at a MOI of 1. At 24 hours post-infection, CR1- γ protein was detected by confocal microscopy with a rabbit affinity-purified antibody raised against a peptide within the annotated CR1- γ protein and an Alexa Fluor 488-conjugated goat anti-rabbit secondary antibody. A non-specific rabbit polyclonal IgG antibody was used as a negative control. To-Pro-3 iodine was used to stain cell nuclei. Images were taken at 63 $\times$.

Fig. 6. Northern blot analysis of RNA expression. A549 cells infected with HAdV-D37 at a MOI of 1. Total RNA from 0, 2, 4, 8, 16, 24, and 48 hours post-infection, was run on a 1% agarose gel.

Fig. 7. Protein modeling and multi-sequence alignment of CR1- γ . Predicted 3D protein model of HAdV-D37 CR1- γ . Identical amino acids among EKC HAdVs are shown in green. Non-conserved amino acids are indicated in red. Similar amino acids in all or at least one other EKC HAdV are represented in yellow and orange, respectively.

Table 1

Primers used for RT-PCR and Northern Blot

Primer	Sequence (5'→3')	Corresponding gene/coding region
E1A 13S forward	TGTAACACAGAGCGCAGGAG	E1A 13S ^a
E1A 13S reverse	GCTGGGACCTACACTTCAGG	E1A 13S ^a
CR1-γ Internal forward	ACGTGGTCAGACCCATCATAAC	E3 CR1-γ ^a
CR1-γ Internal reverse	GGCACTTGAGGTAGCCTCTG	E3 CR1-γ ^a
TPL forward	AGGGCGTACTCCTCGTCATC	Tripartite leader 2 ^a
TPL/CR1-γ reverse	GTAATGCTGCCATTGCTGTG	E3 CR1-γ ^a
Hexon reverse	CCGCGGATGTCAAAGTAAGT	L3 Hexon ^a
GAPDH forward	ATCATCCCTGCCTCTACTGG	Human Glyceraldehyde 3-phosphate dehydrogenase ^b
GAPDH reverse	GTCAGGTCCACCACTGACAC	Human Glyceraldehyde 3-phosphate dehydrogenase ^b

^aSequence derived from HAdV-D37 (GenBank Accession No. DQ900900)^bSequence derived from Human GAPDH (GenBank Accession No. NM 002046.3)

Table 2.Predicted pI and MW of CR1- γ homologues

Homologue	pI	MW
HAdV-8	9.29	29.1K
HAdV-9	9.06	29.2K
HAdV-15	8.48	30.7K
HAdV-17	9.23	30.2K
HAdV-19C	8.5	31.6K
HAdV-19p	6.31	29.7K
HAdV-22	8.26	31.6K
HAdV-26	6.09	28.6K
HAdV-28	6.82	30.3K
HAdV-29	9.06	30.8K
HAdV-36	6.95	31.4K
HAdV-37	8.5	31.6K
HAdV-46	8.93	30.9K
HAdV-48	9.33	32.5K
HAdV-49	8.49	31.7K
HAdV-53	8.67	30.9K
HAdV-54	9.62	29.4K
HAdV-56	8.96	29.8K

Predictions based on ExPASy proteomics server