

RESEARCH ARTICLE

Novel viral splicing events and open reading frames revealed by long-read direct RNA sequencing of adenovirus transcripts

Alexander M. Price¹, Robert T. Steinbock², Richard Lauman^{1,3}, Matthew Charman¹, Katharina E. Hayer⁴, Namrata Kumar¹, Edwin Halko¹, Krystal K. Lum¹, Monica Wei², Angus C. Wilson⁵, Benjamin A. Garcia^{6*}, Daniel P. Depledge^{5,7,8}, Matthew D. Weitzman^{1,9*}

1 Division of Protective Immunity, Department of Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, United States of America, **2** Cell & Molecular Biology Graduate Group, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, **3** Graduate Group in Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, **4** Department of Biomedical and Health Informatics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, United States of America, **5** Department of Microbiology, New York University School of Medicine, New York city, New York, United States of America, **6** Department of Biochemistry and Biophysics, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, United States of America, **7** Institute of Virology, Hannover Medical School, Hannover, Germany, **8** German Center for Infection Research (DZIF), partner site Hannover-Braunschweig, Hannover, Germany, **9** Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, United States of America

* Current address: Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri, United States of America

* weitzmanm@chop.edu



OPEN ACCESS

Citation: Price AM, Steinbock RT, Lauman R, Charman M, Hayer KE, Kumar N, et al. (2022) Novel viral splicing events and open reading frames revealed by long-read direct RNA sequencing of adenovirus transcripts. *PLoS Pathog* 18(9): e1010797. <https://doi.org/10.1371/journal.ppat.1010797>

Editor: Donna M. Neumann, University of Wisconsin-Madison, UNITED STATES

Received: April 12, 2022

Accepted: August 5, 2022

Published: September 12, 2022

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.ppat.1010797>

Copyright: © 2022 Price et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Basecalled fast5 (Nanopore) and fastq (Illumina) datasets generated as part of this study can be downloaded from the

Abstract

Adenovirus is a common human pathogen that relies on host cell processes for transcription and processing of viral RNA and protein production. Although adenoviral promoters, splice junctions, and polyadenylation sites have been characterized using low-throughput biochemical techniques or short read cDNA-based sequencing, these technologies do not fully capture the complexity of the adenoviral transcriptome. By combining Illumina short-read and nanopore long-read direct RNA sequencing approaches, we mapped transcription start sites and RNA cleavage and polyadenylation sites across the adenovirus genome. In addition to confirming the known canonical viral early and late RNA cassettes, our analysis of splice junctions within long RNA reads revealed an additional 35 novel viral transcripts that meet stringent criteria for expression. These RNAs include fourteen new splice junctions which lead to expression of canonical open reading frames (ORFs), six novel ORF-containing transcripts, and 15 transcripts encoding for messages that could alter protein functions through truncation or fusion of canonical ORFs. In addition, we detect RNAs that bypass canonical cleavage sites and generate potential chimeric proteins by linking distinct gene transcription units. Among these chimeric proteins we detected an evolutionarily conserved protein containing the N-terminus of E4orf6 fused to the downstream DBP/E2A ORF. Loss of this novel protein, E4orf6/DBP, was associated with aberrant viral replication center morphology and poor viral spread. Our work highlights how long-read sequencing technologies

European Nucleotide Archive (ENA) under the following study accessions: PRJEB35667 and PRJEB35652. Raw MS files associated with this work have been deposited to the public database the ProteomeXchange Consortium repository via the PRIDE partner repository. The accession number for the MS data is PXD025339 and PXD034464. The newly generated genome and transcriptome annotation can be found at <https://github.com/DepledgeLab/Ad5-annotation> and under the GenBank accession number OP218818. The authors declare that all other data supporting the findings of this study are available within the manuscript and its Supplementary Information files.

Funding: This work was supported through NIH grants R21-AI130618 and R21-AI147163 (ACW), and R21-AI54654, R01-AI145266, and R01-AI121321 (MDW), and R01-AI118891 (MDW and BAG). Additional support came from the NCI T32 Training Grant in Tumor Virology T32-CA115299 (AMP), Individual National Research Service Award F32-AI138432 (AMP), and Pathway to Independence award K99-AI159049 (AMP). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

combined with mass spectrometry can reveal further complexity within viral transcriptomes and resulting proteomes.

Author summary

Adenoviruses are important human pathogens that are also used as powerful tools for gene delivery and as vaccine agents. In addition, studies of the molecular biology of adenovirus have led to seminal discoveries in cell biology, RNA biology, and pathogenesis. While adenoviruses, have been exceptionally well-characterized over the past decades of research, much is still left to discover about their biology. By applying long-read RNA sequencing we have comprehensively reannotated the transcriptome of the Ad5 serotype. In doing so, we discovered an additional 35 abundant viral transcripts that potentially encode for 20 novel viral open reading frames. We demonstrate that one of these novel proteins joins two disparate genetic units into a chimeric protein that is evolutionarily conserved among adenoviruses. Discovery of this protein, E4orf6/DBP, recontextualizes prior studies on the function of the critical viral effector E4orf6. Furthermore, E4orf6/DBP is shown to modulate the morphology of membrane-less viral replication centers in the nucleus and function in cell-to-cell spread of the virus. Together, we showcase how modern techniques in RNA sequencing and mass spectrometry can reveal additional complexity in a well-studied model viral pathogen.

Introduction

Adenoviruses (AdV) are common viral pathogens across multiple species with distinct tissue tropisms including gut, eye, and lung [1]. Among the human adenoviruses, serotypes 2 (Ad2) and 5 (Ad5) from subgroup C are the most prevalent, and cause benign to severe respiratory infections [2]. These two serotypes are highly homologous, sharing 94.7% nucleotide homology between their genomes and 69.2–100% amino acid identity amongst conserved open reading frames (ORFs) [3,4]. AdVs readily infect most transformed human cell lines, and for many decades have served as valuable tools contributing to many seminal discoveries in molecular biology [5]. RNA splicing was discovered by the analysis of adenovirus encoded RNAs [6,7], as well as other important findings in messenger RNA capping and polyadenylation [8,9]. It is now understood that essentially all AdV mRNAs are capped, spliced, polyadenylated, and exported from the nucleus using host cell machinery [10].

AdV are capable of infecting non-dividing cells and reprogramming cellular processes for productive viral infection. This rewiring involves a highly regulated cascade of viral gene expression over various kinetic classes [5]. The first viral gene to be expressed after infection is E1A, a multi-functional transcription factor that activates downstream viral transcription, liberates E2F from RB proteins, and alters host transcriptional responses to the virus [11–14]. While all E1A molecules have identical 5' and 3' nucleotide sequences, splicing of differently sized internal introns allows for the production of discrete proteins that lack specific functional domains conserved across serotypes [15]. Early after infection, E1A is expressed mainly as large and small isoforms, but later in infection alternative splicing leads to the production of a 9 Svedberg E1A isoform (E1A-9s) as well as low abundance doubly-spliced E1A-11s and E1A-10s. The second viral gene to be activated after infection is E1B, predominantly consisting of two spliced isoforms producing 19-kilodalton and 55-kilodalton proteins, with two less

abundant isoforms potentially encoding polypeptides of 156 and 93 residues [16]. While E1B-19K acts to block cellular apoptosis [17], E1B-55K is another multifunctional protein that can cooperate with E1A to alter cellular gene expression downstream of p53, as well as form the targeting component of a viral ubiquitin ligase [18–23]. The remaining early transcription units are all transcriptionally activated by E1A and encode products of related function. The E2 region on the reverse strand of the AdV genome has both an early and a late promoter, as well as two distinct polyadenylation sites, leading to upstream E2A and downstream E2B transcripts [24]. E2A encodes for the viral DNA-binding protein (DBP), while alternative splicing to E2B encodes for the protein-priming terminal protein (pTP), as well as the AdV DNA polymerase (AdPol) [25–27]. The E3 region encoded on the top strand also has two polyadenylation sites leading to E3A and E3B transcription units. The E3 gene products are primarily involved in modulating the host innate immune system [28–30]. Like E1A, the E4 region on the reverse strand has identical 5' and 3' regions, and encodes up to six different ORFs by removal of a first intron of varying length. E4 region transcripts encode for multifunctional proteins that are involved in the regulation of transcription, splicing, and translation of viral RNAs, as well as antagonizing intrinsic cellular defenses [31–33]. Additionally, AdV encodes two Pol III-derived virus associated (VA) RNAs involved in the inactivation of Protein Kinase acting on RNA (PKR) [34,35]. Ultimately, the concerted efforts of the AdV early proteins lead to a cellular state that allows for the replication and amplification of the viral DNA genome [36,37].

Prior to viral DNA replication, the AdV Major Late Promoter (MLP) is thought to be largely silent with small amounts of RNA being made that terminate at the first downstream (L1) polyadenylation site [38]. At this time, so-called intermediate genes pIX and IVa2 begin to be expressed from promoters within the E1B cassette and antisense to the MLP. Both pIX and IVa2 co-terminate at polyadenylation sites within the early genes that they overlap with (E1B and E2B, respectively) and are involved in late gene transcription and packaging [39,40]. Only after viral DNA replication has occurred does the MLP fully activate, supporting the hypothesis that active replication *in cis* is a prerequisite for full viral late gene expression [41–43]. The Major Late Transcriptional Unit (MLTU) begins with a series of three constitutive exons spliced together to form the tripartite leader, before downstream splicing to late cassettes defined by one of five alternative polyadenylation sites (termed L1-L5) [38]. Splicing within the tripartite leader to the so-called “i” exon leads to a putative ORF upstream of subsequent late gene splicing events and destabilizes these RNA molecules [44,45]. An additional intermediate promoter has been reported within the L4 region that allows for the early expression of L4-22K and L4-33K proteins important for the splicing of other late genes [46,47]. The MLTU encodes for primarily structural capsid components or proteins involved with packaging of new virions, and their expression ultimately leads to the death of the host cell. More recently, a novel late gene, UXP, was discovered on the reverse strand of the genome [48,49]. The UXP promoter is located between E4 and E2, and splices downstream to the exons within the E2A region to continue translation of an ORF in an alternate reading frame to that of DBP. This exciting finding suggests that our knowledge of AdV transcripts is incomplete, especially within the complex MLTU region.

The complete Ad5 genome was sequenced in 1991 using Sanger sequencing of viral genome fragments inserted into plasmid DNA and amplified in bacteria [3]. This genome sequence was then annotated in 2003 based on homology to similar serotypes of AdV [4]. As such, the current reference annotation for Ad5 available on the National Center for Biotechnology Information ([AC_000008](https://doi.org/10.1093/ncbi/taaa000008)) is incomplete, and lacks critical information such as transcription start sites (TSS), cleavage and polyadenylation sites (CPAS), and the resulting 5' and 3' untranslated regions (UTR) that the aforementioned information dictates. In recent years,

new technologies have allowed for high-throughput investigation of gene expression utilizing various techniques. The effect of AdV infection on host gene expression has been shown for Ad5 by microarray analysis [50,51], as well as for Ad2 by Illumina-based short-read sequencing [52,53]. Analyses of both single-end and paired-end short-read RNA-seq data from cells infected with Ad2 revealed both temporal viral gene expression and high-depth splicing information, and identified both previously confirmed and novel RNA splice site junctions [54]. In addition, temporal analysis of Ad5 viral gene expression was performed using digital PCR to determine expression kinetics of a subset of known viral genes [55]. The late RNA tripartite leader splicing was also analyzed by short-read sequencing across a number of human AdV serotypes [44].

While the quality and depth of current short-read sequencing technologies are high, the complex nature of many viral transcriptomes precludes the unambiguous mapping of these short reads to any one particular RNA isoform due to extreme gene density and overlapping transcriptional units [56,57]. In this regard, the ability of long-read RNA sequencing to map full-length transcripts has the potential to revolutionize detection of novel isoforms and multiply spliced RNA at the single-molecule level [58–60]. Two recent studies have leveraged the power of Oxford Nanopore Technologies to characterize an exceptional diversity of adenoviral splicing [61,62]. Donovan-Banfield *et al.*, performed direct RNA sequencing over a time course of infection with Ad5 in MRC5 fibroblast cells and characterized over 11,000 different splicing patterns [61]. Their study used an “ORF-centric” approach to define low level splice sites leading to altered production of canonical genes, as well as multiple upstream and alternative open reading frames. In contrast, Westergren-Jakobssen *et al.*, used direct cDNA based sequencing on a time course of Ad2-infected IMR90 lung fibroblast cells [62]. This study uncovered over 900 potential RNA transcripts, including upstream transcription from the inverted terminal repeats, and compared novel ORFs to existing Ad2 proteomics data. While the diversity and complexity of potential adenovirus splicing is immense, it remains difficult to determine functional relevance of lowly expressed alternative transcript isoforms.

In this study, we have re-annotated the Ad5 genome and transcriptome using a combination of short-read and long-read RNA sequencing technologies. The high read depth and accuracy of base-calling achieved by Illumina-based short-read sequencing allowed for both the detection of single nucleotide polymorphisms within transcriptionally active regions of the viral DNA genome, as well as error-correction of the inherently noisier base-calling of Nanopore-based long-read direct RNA sequencing (dRNA-seq). Detection of full-length RNA transcripts and the assignment of TSS and CPAS transcriptome-wide was enabled by dRNA-seq. Furthermore, by combining highly accurate splice site junctions from short-read sequencing and full-length isoform context from long-read sequencing, we were able to reevaluate the splicing complexity of AdV transcriptional units. Using this integrated approach, we have discovered 35 additional viral polyadenylated RNAs for a total of 90 unique mRNAs produced by Ad5. Of these novel isoforms, 14 RNAs encode for a canonical ORF with changes in upstream or downstream splicing. The remaining 21 encode new ORFs or alter existing ORFs by internal truncations or in-frame fusion of genes from separate transcription units. In particular, we have focused on one novel transcript, a fusion of the N-terminus of E4orf6 to the downstream DBP ORF which generates an abundant and evolutionarily conserved chimeric protein that we called E4orf6/DBP. We show that E4orf6/DBP regulates the morphology of viral replication centers and ultimately impacts viral spread. Taken together, our data reveal exceptional transcriptional complexity of AdV and highlight the necessity of revisiting transcriptome annotations following the emergence of appropriate new technologies.

Results

RNA-seq reveals high-confidence SNPs within the Ad5 genome

Illumina-based RNA sequencing (RNA-seq) relies on the fractionation of RNA molecules before reverse transcription into complementary DNA, and therefore loses information such as RNA modifications and the context of distal splice junctions within full-length molecules. However, the accuracy of each individual base call is very high [63]. Using bcftools, a common variant-calling algorithm designed to assess allele-specific variation within RNA-seq, we were able to detect single nucleotide polymorphisms (SNPs) within the RNA transcriptome that likely emerge from mutation within the DNA genome [64,65]. While RNA modifications such as inosine can be read as SNPs during the process of reverse transcription, these events did not approach the 99% read depth stringency we required among our three biological replicates to call a conserved variant [66]. While this technique will only capture SNPs within the actively transcribed region of the genome, nearly every nucleotide of the gene-dense AdV genome is transcribed at a sufficient level for this strategy to provide meaningful data.

In total, we discovered 24 SNPs and no insertions or deletions in the Ad5 genome when compared to the original annotation (Fig 1). Of these deviations, exactly half (12) are not predicted to change amino acid coding capacity, with two SNPs occurring within untranslated regions of viral RNA and the other ten leading to synonymous amino acid codons within all reading frames annotated to be protein producing. The remaining 12 mutations are predicted to lead to coding sequence variations at the amino acid level, with all examples being missense mutations and no evidence of premature stop codons. Importantly, none of the mutations discovered generated novel RNA splice sites. These data demonstrate the ability to call mutations within the DNA genomes of viruses using solely high-depth RNA sequencing data. While these SNPs could have arisen during cell culture passage solely within our lab, whole-plasmid sequencing of a common Ad5 mutagenesis platform (pTG3602, kind gift of Patrick Hearing) revealed exactly the same 24 SNPs with no additions or Indels. Furthermore, detecting only 24 SNPs out of 35,938 nucleotides highlights the overall genomic stability of AdV.

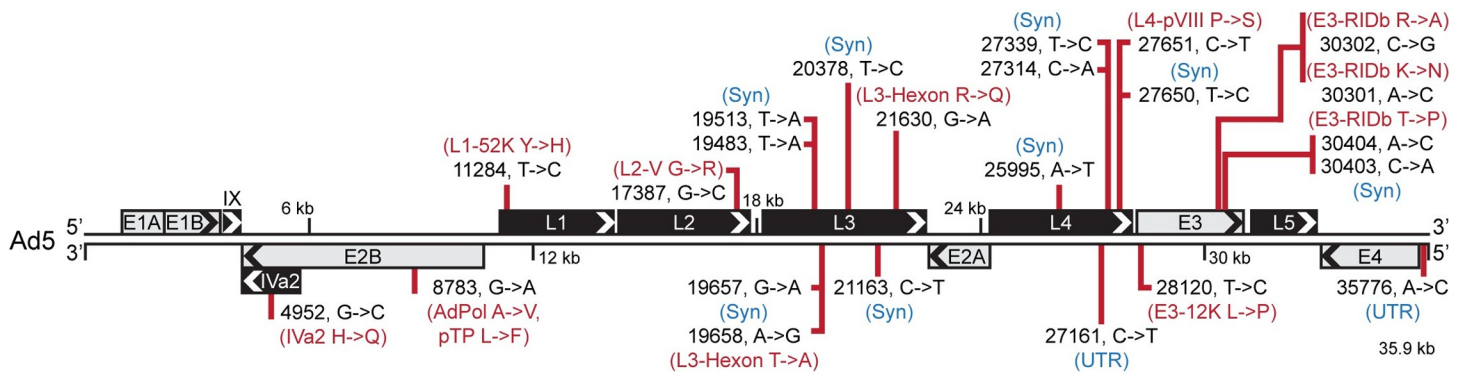


Fig 1. RNA-seq reveals high-confidence SNPs within the Ad5 genome. The 35,938 base pair linear genome of Ad5 is displayed in the traditional left to right format. Major transcriptional units are shown as boxes above or below the genome with arrowheads denoting the orientation of the open reading frames (ORFs) encoded within. Grey boxes denote early gene transcriptional units while black boxes denote late genes. Bcftools was used to analyze short-read RNA seq data to predict single nucleotide polymorphisms (SNPs) and insertions/deletions (InDels) that approach 100% of the RNA reads when compared to the reference Ad5 genome (AC_000008). In total, 24 such SNPs were discovered and their positions within the genome is highlighted by red vertical lines. For each SNP, the nucleotide position as well as the top strand reference base and corrected base are shown in black text (nucleotide position, reference base -> corrected base). If indicated SNPs fell within untranslated regions (UTR), or did not change the encoded amino acid of any annotated reading frame potentially impacted by the SNP, these were marked with blue text denoting either UTR or Syn (synonymous mutation), respectively. For any SNP that led to an amino acid change within an annotated ORF, these ORFs as well as the identity of the reference amino acid and corrected amino acid are highlighted in red.

<https://doi.org/10.1371/journal.ppat.1010797.g001>

Combined short-read and long-read sequencing showcases adenovirus transcriptome complexity

To compare short-read Illumina sequencing and long-read nanopore sequencing directly, A549 cells were infected with Ad5 for 24 hours and total RNA was harvested in biological triplicate. Three samples were prepared into standard strand-specific Illumina RNA-seq libraries targeting the polyadenylated mRNA fraction. The same RNA samples were then pooled and poly(A) purified before submitting to direct RNA sequencing (dRNA-seq) on an Oxford Nanopore Technologies MinION MkIb platform [67]. Resulting sequence reads were aligned to the Ad5 reference genome using either GSNAP for short-reads [68], or MiniMap2 for long-reads [69]. Overall sequencing depth for both forward and reverse reads are shown in Fig 2. While Illumina sequencing provided on average three times the read depth when compared to dRNA sequencing, the overall coverage plots were similar.

dRNA-seq is performed in the 3' → 5' direction and thus allows precise mapping of the 3' ends of transcripts at which poly(A) tails are added (cleavage and polyadenylation site, CPAS) [67]. Where the quality of input RNA is high, a variable proportion of sequence reads extend all the way to their transcription start site (TSS). By collapsing sequence reads to their 5' and 3' ends, we were able to implement a peak-calling approach to predict TSS and CPAS [58], and map their positions along both the forward and reverse strand of the viral genome (Fig 2). In addition, ContextMap2 [70] was used to mine Illumina RNA-seq data for short read sequences containing poly(A) stretches that could be aligned against the viral genome for an orthogonal method of CPAS detection (Fig 2). Mapping the TSS on the forward strand revealed the locations of the promoters for E1A, E1B, pIX, MLP, and E3, while the reverse strand revealed the promoters for E4, UXP, E2-early, E2-late, and IVa2. While we were able to reliably detect the L4 promoter TSS at 12 hpi, this was not clearly detected above noise at 24 hpi [46]. When mapping CPAS loci, we saw great concordance between the dRNA-seq and ContextMap2 performed on short-read sequences. On the forward strand we were able to detect previously mapped CPAS events at the E1A, E1B/pIX, E3A, E3B, and individual L1 through L5 sites. On the reverse strand we detected CPAS at the E4, UXP/DBP, and E2B/IVa2 locations. Exact nucleotide locations of both TSS and CPAS can be found in S1 Table. In addition, we also detected TSS and CPAS around the RNA pol III-derived VA RNA I (Fig 2). While pol III transcripts are generally not polyadenylated, and thus would not be captured by our nanopore sequencing approach, it was previously reported that low levels of polyadenylation can occur on these transcripts [71]. Given the high abundance of AdV VA RNAs (up to 10⁸ copies per cell during late infection), it remains likely that low level VA RNA polyadenylation events are occurring [35].

To generate accurate splicing maps of AdV transcripts, we combined the sensitivity of short-read sequencing to identify RNA junctions and then placed them in the context of full-length RNA isoforms using dRNA sequencing. Due to the spurious nature of low-level AdV splicing events [54], we set abundance thresholds for the highly abundant viral late transcripts of 500 reads for short-read junctions, and at least ten events detected in the long-read sequencing when collapsed by FLAIR [72]. Using this method, we readily detected other recently discovered viral isoforms, such as multiple splice sites preceding the pVII ORF [54], the so-called X, Y, and Z leaders embedded in E3 and preceding L5-Fiber [54,73], and the newly described UXP [48,49]. Using full-length RNAs, we were able to detect novel splice sites producing canonical ORFs that only differ in UTRs for E1A, L4-100K, L4-33K, L4-pVIII, DBP, and E4orf6/7. In addition, we discovered canonical ORF isoforms embedded within transcripts generated from non-canonical promoters, such as Fiber driven by the E3 promoter, multiple E3 proteins driven by the Major Late Promoter, and DBP driven by the E4 promoter. Within

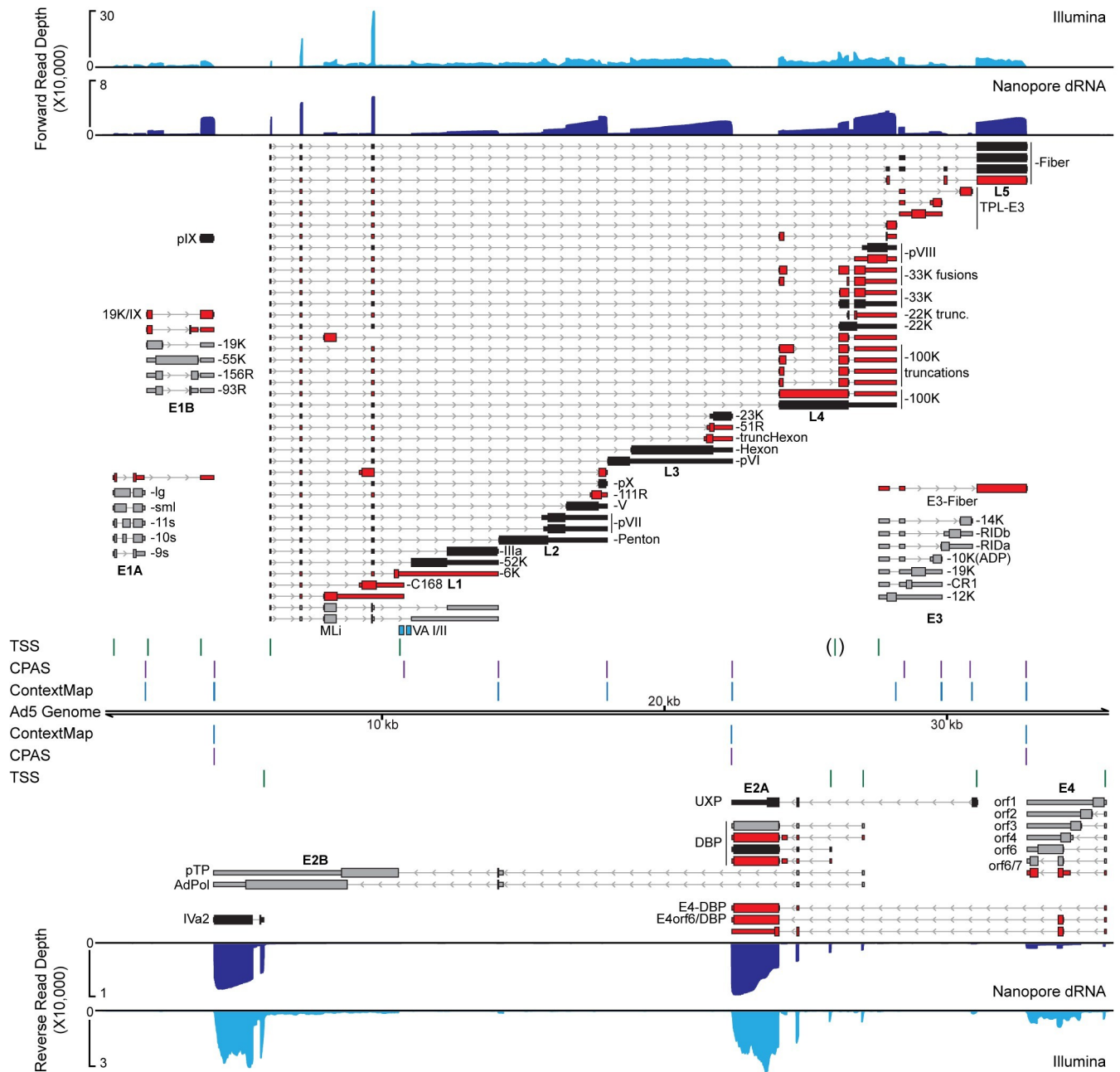


Fig 2. Combined short-read and long-read sequencing showcases adenovirus transcriptome complexity. A549 cells were infected with Ad5 for 24 hours before RNA was extracted and subjected to both short-read and long-read sequencing. Sequence coverage provided by short-read stranded RNA-seq (Illumina, light blue), as well as nanopore long-read direct RNA-seq (Nanopore dRNA, dark blue), is shown along the Ad5 genome. For both tracks, reads aligning to the forward strand are plotted above the genome, while reads aligning to the reverse strand are shown below the genome. For dRNA-seq datasets, reads can be reduced to their 5' and 3' ends and peak-calling applied to predict individual transcription start sites (TSS, green vertical lines) or cleavage and polyadenylation sites (CPAS, magenta vertical lines), respectively. The TSS for the L4 promoter was only detected at 12 hpi and is thus displayed in brackets. Similarly, the ContextMap algorithm can predict, albeit at lower sensitivity, CPAS sites from poly(A) containing fragments within Illumina RNA-seq data (ContextMap, light blue vertical lines). Individual RNA transcripts are shown above and below the genome, with thin bars denoting 5' and 3' untranslated regions (UTR), thick bars denoting open reading frames (ORFs), and thin lines with arrowheads denoting both introns and orientation of transcription. Previously characterized early genes are denoted in grey, while previously characterized late genes are denoted in black. RNA isoforms discovered in this study are highlighted in red. Names of transcriptional units are shown under each cluster of transcripts, while the name of the protein derived from the respective ORF is listed after each transcript to the right or left. The position of Pol III-derived noncoding RNAs virus associated VA-I and VA-II are highlighted in teal boxes.

<https://doi.org/10.1371/journal.ppat.1010797.g002>

transcriptional units, we discovered the presence of internal splice sites leading to in-frame truncations of existing ORFs, such as E1B-19K, L3-Hexon, L4-22K, and four distinct isoforms of truncated L4-100K. We also discovered splicing events predicted to lead to in-frame fusion events within transcriptional units, such as fusions between N-terminal fragments of L4-100K and L4-33K or L4-pVIII or the X-Z-Fiber ORF. Furthermore, gene fusion events were observed that join disparate transcriptional units, such as an N-terminal fragment of E1B-19K and pIX (19K/IX), the internal major late exon to L4-100K (MLi/100K), and E4orf6 to DBP (E4orf6/DBP). Lastly, we detected multiple transcripts that could encode novel ORFs with a minimum of 50 amino acids. Two ORFs, C168 and L1-6K, have been predicted in the related serotype Ad2 but there was only evidence for production of C168. Novel splice acceptors within L2 and L3 lead to RNAs containing 111 residue and 51 residue ORFs (L2-111R and L3-51R, respectively). Both of these splice sites are conserved in Ad2 [54,62]. Finally, a transcription unit crossing RNA joining the E4orf6/7 splice donor to the major two exons within E2A leads to a potential 133 residue ORF (E4-Unk) that would be translated in an alternate frame from both DBP and UXP. The complete splicing map of Ad5 is included as [S1 Fig](#). Overall, we discovered 35 new isoforms for a total of 90 expressed RNA isoforms during Ad5 infection. When compared to recently published adenovirus long-read sequencing studies [61,62], we find concordance for 33/35 transcripts from the matching Ad5 strain and 28/35 transcripts from the related Ad2 strain ([S2 Table](#)).

Detection of novel Ad5 protein species utilizing mass spectrometry

Our laboratory recently generated a high depth whole cell proteome (WCP) using fractionation and label-free quantification during a time-course of Ad5 infection [74]. In this study, canonical viral proteins were detected and quantified comprehensively, so we reasoned that our existing WCP contained enough depth to provide evidence for the existence of novel viral ORFs predicted by RNA sequence. While there are a total of 20 novel ORFs predicted by this study, 15 of them are fusions between otherwise canonical AdV ORFs. Therefore, for these 15 chimeric transcripts only a single unique tryptic peptide covering the junction region defines the unique protein from the shared peptides of other canonical AdV viral proteins. In addition, Parallel Reaction Monitoring Mass Spectrometry (PRM-MS) was used to detect specific peptides of known mass and charge to validate many of the lower abundance peptide junctions identified using the Data Dependent Acquisition-based fractionated dataset. In total, our mass spectrometry analysis found high confidence evidence for 12/20 novel ORFs (60%) and is summarized in [Table 1](#). Thus, it is likely that many more of the novel AdV ORFs exist, but were simply not detected due to unfavorable mass and charge characteristics coupled with fusion proteins only possessing a single informative junctional peptide. While these novel proteins are uniformly low in abundance compared to the canonical viral proteins, they might affect the function of the canonical proteins they are composed of in dominant negative ways by swapping of key domains. In addition, these novel peptides might be loaded into MHC molecules for signaling to the adaptive immune system [75].

Direct RNA Sequencing unambiguously distinguishes early and late transcription

We next determined if we could provide unambiguous detection of viral transcripts over a time-course of infection that captured early and late viral kinetics. By aligning long reads to the fully re-annotated viral transcriptome (as opposed to the viral genome), and only counting the reads that could be unambiguously assigned to a single transcript, we were able to detect all of the canonical and newly discovered transcripts ([Fig 3](#)). At 12 hours post-infection (hpi)

Table 1. Unique peptides of novel AdV ORFs detected by mass spectrometry. Specific detected viral peptides are listed along with their amino acid number in brackets. PSM, peptide spectral match within whole cell proteome. PRM Detection, parallel reaction monitoring run performed. PF/TF, peptide fragments/total possible peptide fragments detected by PRM.

Protein Name	Peptide	PSM	q-Value	PRM Detection	PF/TF	
E1B-19K/IX	[57–81] SCAAAAAMSTNSFDGSIVSSYLTR	Nd		Yes	20/25	
MLi	[1–23] MRADREELDLPPPIGGVAIDVVK	15	2.37E-05			
	[3–23] ADREELDLPPPIGGVAIDVVK	187	2.37E-05			
	[3–31] ADREELDLPPPIGGVAIDVVKVEVPATGR	24	2.37E-05			
	[6–23] EELDLPPPIGGVAIDVVK	35	8.26E-05			
	[24–31] VEVATGR	74	0.000167147			
	[32–39] TLVLAQVK	46	0.000139108			
	[40–63] TCAVLAAVHGLYLHEVDLTTAHK	110	2.37E-05			
	[64–75] EAEWEFEPLAWR	12	2.37E-05			
	[107–113] AQSPDVR	47	0.000433765			
	[116–136] RSELDNIAQMGAVHGLELPR	71	2.37E-05			
	[116–137] RSELDNIAQMGAVHGLELPRR	3	2.37E-05			
	[117–136] SELDDNIAQMGAVHGLELPR	53	2.37E-05			
	C168	[120–134] DLSESASTGSENLSR	2	0.000783189		
	L1-6K	[1–16] MYLDIQVMPAAVVEAR	1	0.00836917		
L2-111R	[77–93] SLWFLQIWPSAASVSR	4	8.26E-05			
	[94–111] CRDSEEECTVGGAWPATA	6	2.37E-05	No		
L3-51R	[38–47] EASNINNSCR	2	0.00530179			
L4-100Ktr3	[60–81] SVPTEDKKQDQDNAEPYQQQPR	8	2.37E-05			
	[67–81] KQDQDNAEPYQQQPR	10	2.37E-05			
	[68–81] QDQDNAEPYQQQPR	8	2.37E-05	No		
L4-100Ktr4	[155–172] LNFYPVFAVPEPYQQQPR	20	2.37E-05	Yes	13/18	
MLi-100K	[137–145] RQPYQQQPR	16	8.26E-05			
L4-100K-33K	[93–98] GPLLPR	13	0.00496589	Yes	3/6	
E4-Unk	[94–106] SSAKPPPSADAVR	3	0.00343938			
E4orf6/DBP	[39–60] ATILEDHPLLPECNTLTMHNEK	Nd		Yes	17/22	

<https://doi.org/10.1371/journal.ppat.1010797.t001>

the majority of viral transcripts detected were early RNAs, particularly E1A-large and E1A-small, E1B-19K and E1B-55K, early promoter DBP, E3-12K, E3-19K, and E4orf3 (Fig 3A). However, at this time point we still detected low-level viral late transcripts that progressed beyond the L1 polyadenylation site, corroborating recent observations [55]. At 24 hpi, however, viral gene expression shifted to be dominated by late gene expression, as well as early transcripts derived from the E1B and late promoter DBP locus (Fig 3B). At late times post-infection we also saw the E3-Fiber transcript, as well as the tripartite leader containing E3 locus. The transcripts spanning disparate transcriptional units, i.e., 19K/IX (14-fold), E3-Fiber (15-fold), and E4orf6/DBP (6-fold), increase substantially at 24 hpi, implicating these messages as novel late transcripts, with expression as abundant as the recently described late UXP transcript [48,49]. Furthermore, while all permutations of the X, Y, and Z leaders preceding Fiber were previously detected by short-read sequencing, these could not be phased to full-length transcript isoforms [54,73]. Our full-length RNA data indicate that all Fiber transcripts can be detected, but MLP-Fiber and Y-Fiber are the most abundant, followed by XY-Fiber, and then all other isoforms. While the previous lack of detection of some of these novel transcripts can be explained by low overall abundance (e.g., L2-111R, L4-100K/VIII), many of the L4-100K truncations and L4-33K fusions are expressed at levels higher than that of the *bona fide* late transcript UXP. These data demonstrate that the newly discovered viral transcripts can be

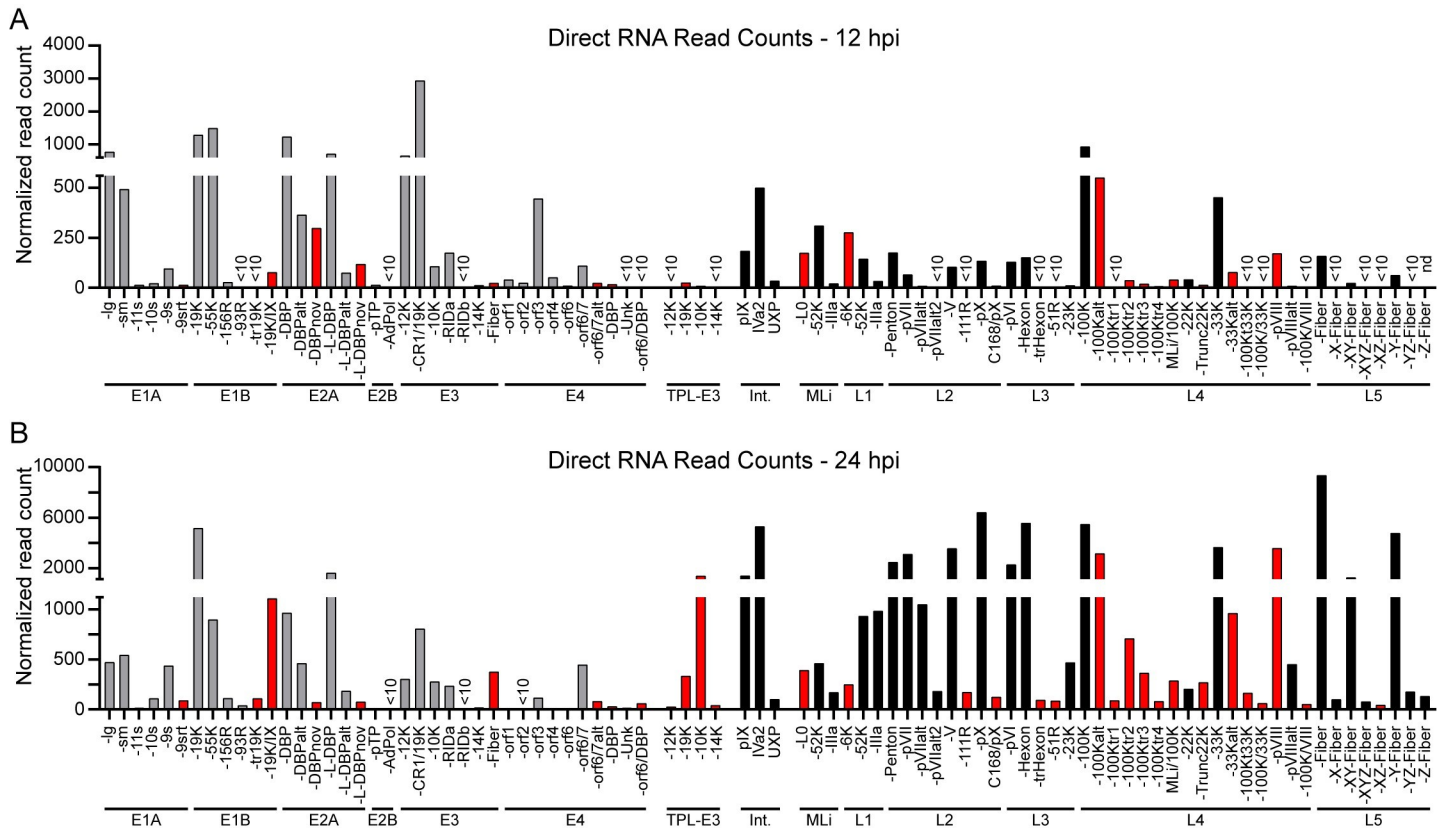


Fig 3. Direct RNA Sequencing (dRNA-seq) unambiguously distinguishes early and late transcription. (A) dRNA-seq was performed on polyadenylated RNA from Ad5-infected A549 cells extracted at 12 hours post-infection (hpi). Sequence reads were aligned to the re-annotated transcriptome and filtered to retain only unambiguous primary alignments. Normalized read count indicates the number of RNAs for a particular transcript once normalized to the total number of mappable reads (human plus adenovirus) for the entire sequencing reaction. For all panels, grey bars indicate early genes, black bars indicate late genes, and red bars indicate novel isoforms discovered in this study. Undetectable transcripts (nd) or those with fewer than 10 counts of a particular isoform detected (<10) are indicated. (B) Same as in Panel A, but with RNA harvested at 24 hpi.

<https://doi.org/10.1371/journal.ppat.1010797.g003>

reproducibly detected over a time-course of infection with Ad5, as well as display differential expression based on the stage of infection.

The novel fusion transcript between E4orf6 and DBP is expressed, translated, and conserved

We were particularly intrigued by the discovery of transcripts that start within the E4 locus and proceed downstream to the E2A locus, since this event unites viral gene cassettes previously associated with host antiviral antagonism and viral DNA replication, respectively [33]. To be able to splice to downstream E2A acceptors, nascent pre-mRNA must bypass the cleavage and polyadenylation site within the E4 transcriptional unit (Fig 4A). Reverse transcription followed by PCR from the E4orf6 region to the downstream DBP exon showed two products of expected size (E4orf6/DBP and E4-Unk) capable of bypassing the E4 polyadenylation site (Fig 4B). To ascertain whether any chimeric E4orf6 protein is produced, we leveraged the well-characterized RSA3 antibody that recognizes the N-terminus of E4orf6 [76,77]. Upon immunoblot of A549 cells infected with WT Ad5 for 40 hours the RSA3 antibody detected both the 34 kilodalton full-length E4orf6, as well as E4orf6/7 which results from a splicing and frame-shifting event (Fig 4C). Intriguingly, the antibody also detected a

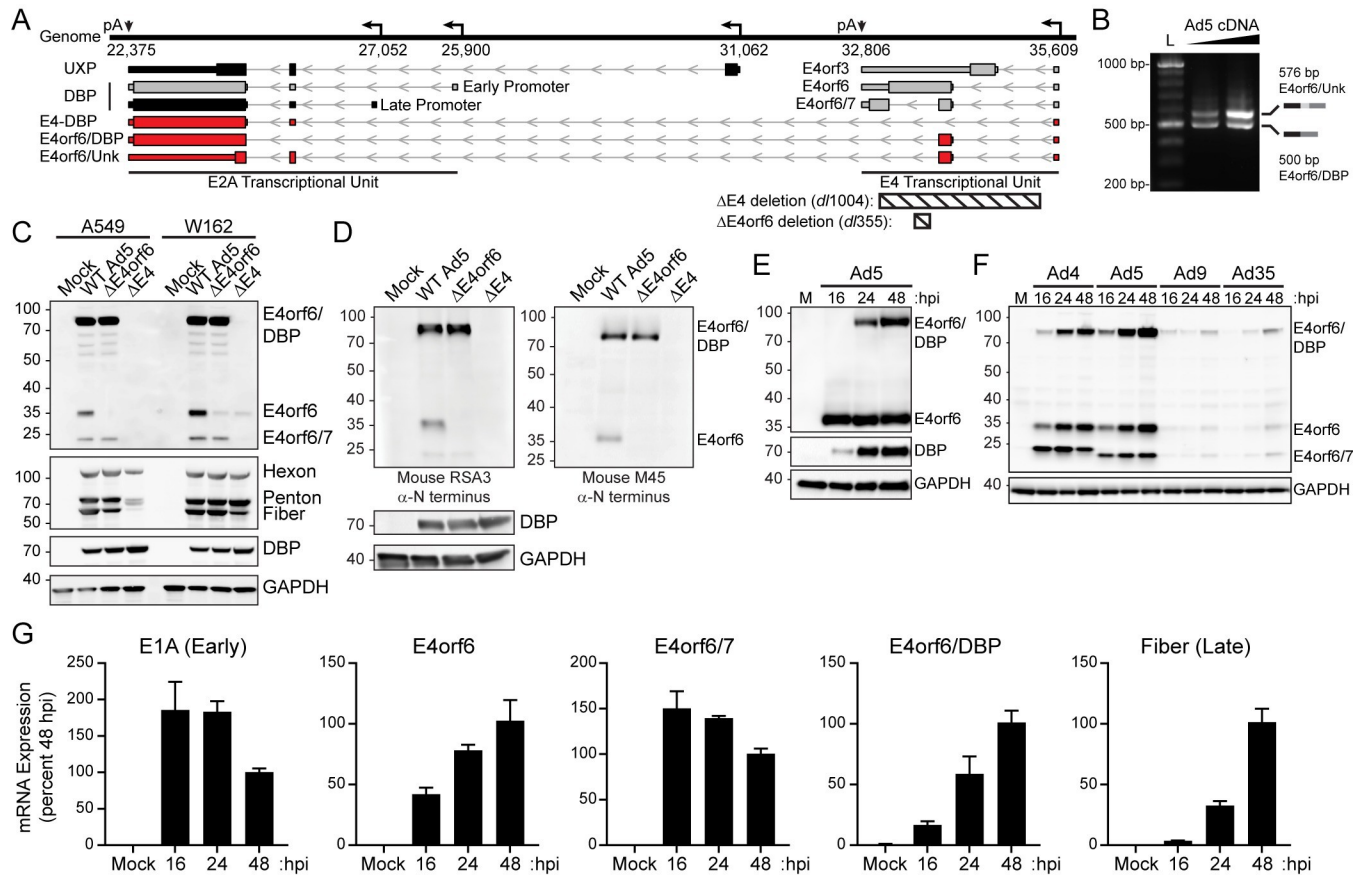


Fig 4. Novel fusion transcript between E4orf6 and DBP is expressed, translated, and conserved. (A) Enlarged transcriptome map of Ad5 E4 and E2A transcriptional units. Promoter transcription start sites are indicated with left-facing arrows, and cleavage and polyadenylation sites (pA) labeled with downward facing arrows. Novel E4-derived transcripts that terminate in E2A are highlighted in red. Ad5 mutant viruses *dl1004* and *dl355* contain deletions (indicated by hashed boxes) that remove most of the E4 region and splice donors ($\Delta E4$) or a 14-base deletion inside E4orf6 that only abrogates E4orf6 expression ($\Delta E4orf6$). (B) Reverse-transcriptase PCR on cDNA derived from Ad5-infection of A549 cells reveals characteristic bands of both E4orf6/Unk and E4orf6/DBP. L denotes DNA ladder and triangle indicates increasing cDNA concentration. (C) A549 or W162 cells were uninfected (mock) or infected with WT Ad5, $\Delta E4orf6$, or $\Delta E4$ viruses for 40 hours and proteins detected by immunoblot analysis. When blotting with antisera raised against the N-terminus of E4orf6, a prominent band is detected at the predicted size of E4orf6/DBP. This band is absent during $\Delta E4$ infection and not observed in W162 cells where only the E4 region is provided *in trans*. Kilodalton size markers are shown to the left of each blot. (D) Infections and immunoblot analysis were performed as described in panel C. Two independently derived anti-N-terminal E4orf6 antibodies (RSA3 and M45) detect E4orf6/DBP. (E) Proteins expressed over a time-course of Ad5 infection in A549 were detected by immunoblot analysis at indicated hpi. (F) A549 cells were infected with adenoviruses from four different serotypes in a time-course. All tested adenovirus serotypes express a protein corresponding to E4orf6/DBP. (G) Quantitative reverse-transcriptase PCR was performed to demonstrate mRNA accumulation of Ad5 E1A (a representative early transcript), Fiber (a representative late transcript), and three E4orf6 containing transcript isoforms. Transcripts were normalized to expression level at 48 hpi and internal HPRT1 housekeeping gene.

<https://doi.org/10.1371/journal.ppat.1010797.g004>

robust band around 80 kilodaltons which corresponds to the predicted molecular weight of the N-terminus of E4orf6 fused to the entire open reading frame of the 72 kilodalton DBP. Leveraging the power of AdV genetic knockouts, we showed that a $\Delta E4orf6$ virus (*dl355* [78]) lacks expression of full-length E4orf6 due to a 14 bp deletion leading to premature translation termination, while retaining the RNA splice donor that allows for expression of both E4orf6/7 as well as the putative E4orf6/DBP. Furthermore, infection with a $\Delta E4$ virus (*dl1004* [31]) that abrogates all splice donors and acceptors within the E4 region led to loss of the bands corresponding to E4orf6, E4orf6/7, and E4orf6/DBP. The $\Delta E4$ virus is highly defective for viral replication, however most functions can be restored by infecting Vero W162 cells that contain an integrated copy of the E4 region and therefore provide expression of E4 RNAs *in trans* upon infection [79]. Importantly, in human cells splicing of pre-

mRNA must occur *in cis*, and it is very rare to observe *trans*-splicing. Infection of W162 cells with $\Delta E4orf6$ or $\Delta E4$ mutant viruses restored low-level expression of both E4orf6 and E4orf6/7, but $\Delta E4$ virus infection did not show expression of E4orf6/DBP (Fig 4C). Finally, to rule out that the detected band was a cross-reactive species, an independently derived anti-N-terminal E4orf6 antibody (M45) was used to recapitulate the finding with the previously described viral mutants (Fig 4D). When reviewing the literature for additional examples of uncropped immunoblots utilizing N-terminal E4orf6 antibodies, we noticed that multiple other groups also detected a band corresponding to the molecular weight of E4orf6/DBP [77,80]. Thus, the genetic and biochemical evidence indicates that E4orf6/DBP is produced during Ad5 infection.

We were surprised to see such a prominent band for E4orf6/DBP protein when compared to the equivalently expressed E4orf6 mRNA or more highly expressed E4orf6/7 mRNA. Since our transcriptomic analysis (Fig 3) predicted that E4orf6/DBP might possess late gene kinetics, we explored E4orf6/DBP expression over a time-course of infection. Upon infection of A549 cells with WT Ad5 we saw that E4orf6 had reached maximal expression as early as 16 hpi, whereas E4orf6/DBP was first detected at 24 hpi and reached maximal expression at 48 hpi (Fig 4E). Similarly, when A549 cells are infected with diverse human AdV serotypes, the E4orf6/DBP protein was detected and increased in expression in a time dependent manner (Fig 4F). Finally, when quantitative reverse transcription PCR was performed to assess mRNA expression characteristics of E4orf6, E4orf6/7, and E4orf6/DBP, the E4orf6/DBP transcript had kinetics more similar to a known late gene (Fiber) than a representative early mRNA (E1A) (Fig 4G). Thus, E4orf6/DBP is conserved among diverse AdV serotypes, and expressed as a true late gene during infection.

E4orf6, but not E4orf6/DBP, generates an E3 ubiquitin ligase complex with E1B55K

E4orf6 is most well characterized to bridge the viral targeting receptor E1B55K to a cellular E3 ubiquitin ligase composed of Elongin B, Elongin C, and Cullin5 [21–23]. The region necessary for binding E1B55K has been mapped to the N-terminus of E4orf6, and therefore we considered whether fusion of this 58 amino acid stretch to the disordered N-terminus of DBP could regulate ubiquitin ligase activity (Fig 5A). We generated N-terminal flag-tagged versions of both E4orf6 and E4orf6/DBP in mammalian expression constructs. HEK293 cells contain an integrated copy of AdV E1A and E1B such that when E4orf6 is transfected a viable E1B55K/E4orf6 ubiquitin ligase is reconstituted [81]. We verified that expression of E4orf6 alone was sufficient to degrade known ligase targets Rad50 and Mre11, and that addition of the N-terminal flag tag to E4orf6 did not hinder this activity (Fig 5B). Expression of flag-E4orf6/DBP led to detection of an RSA3-responsive band at the expected size, but had no effect on stability of Rad50 or Mre11. Furthermore, co-expression of E4orf6 with E4orf6/DBP did not hinder the ability of E4orf6 to form a functional ubiquitin ligase.

Previous studies using monoclonal antibodies against E4orf6 have reported E4orf6 to be present at viral replication centers (VRCs) marked by DBP by indirect immunofluorescence assay [82–84]. In fact, E4orf6 localization to DBP was thought to bring a functional viral ubiquitin ligase to sites of viral DNA replication and RNA transcription [85–87]. Now that we know a major fraction of the E4orf6 signal detected at late times of infection by RSA3 or M45 immunoblot is actually E4orf6/DBP, we revisited this subcellular localization determined by immunofluorescence. In uninfected HEK293 cells, flag-tagged E4orf6 and E4orf6/DBP were both diffusely nucleoplasmic (Fig 5C). During infection, DBP-marked replication centers recruit the cellular deubiquitinase USP7 [88]. While flag-E4orf6/DBP

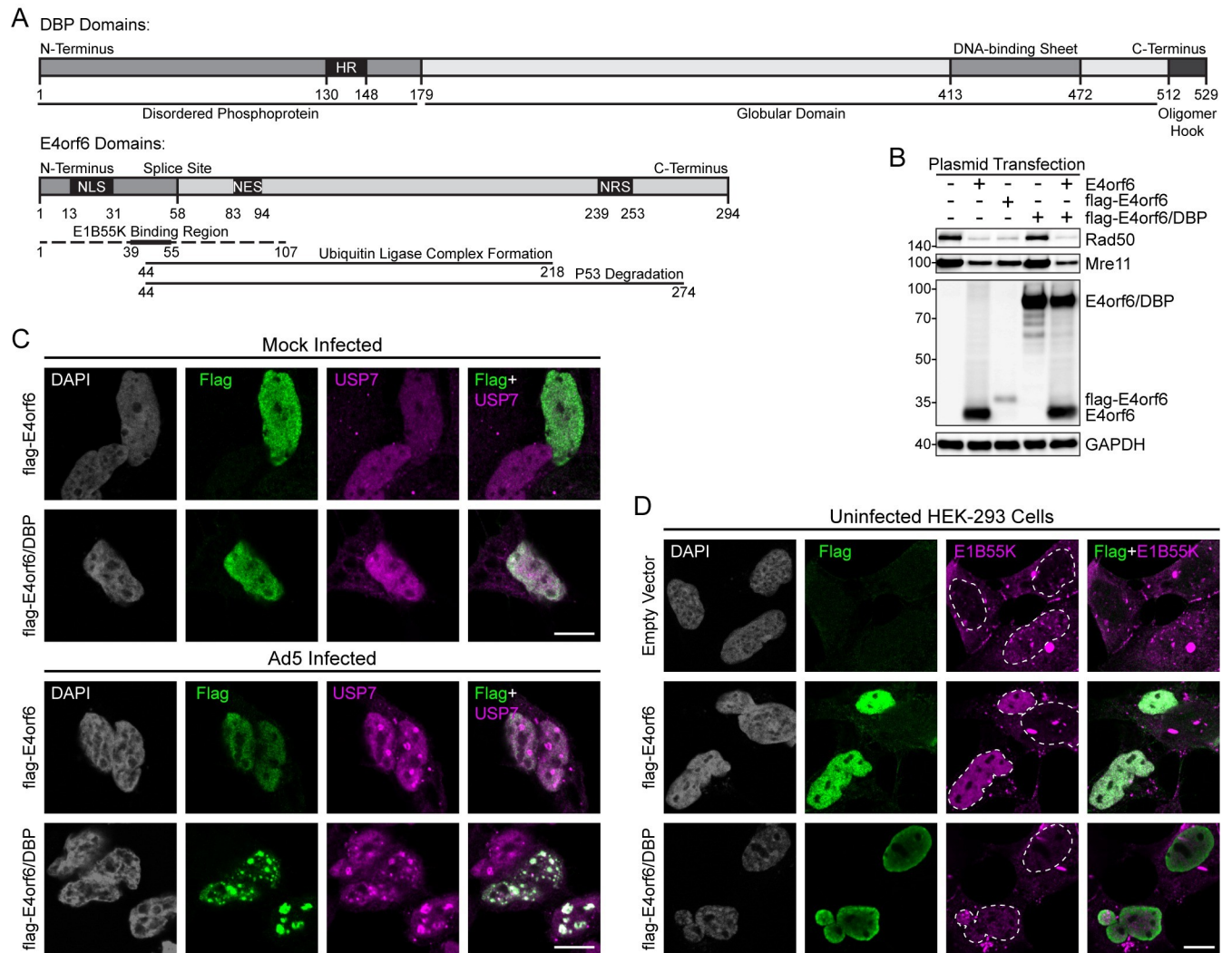


Fig 5. The E4orf6 protein generates an E3 ubiquitin ligase complex with E1B55K but this is not observed for the E4orf6/DBP fusion. (A) Domain map of DBP and E4orf6 open reading frames. HR: host range, NLS: nuclear localization sequence, NES: nuclear export sequence, NRS: nuclear retention sequence. (B) HEK293 cells were transfected with E4orf6, flag-E4orf6, or flag-E4orf6/DBP and subjected to immunoblot analysis. E4orf6, but not E4orf6/DBP, induces degradation of Mre11 and Rad50 through a complex with E1B55K (C) HEK293 cells were transfected with flag-E4orf6 or flag-E4orf6/DBP and then subsequently mock-infected or Ad5-infected for 24 hours. Immunofluorescence was performed for flag tag (green) or cellular USP7 (magenta). Both proteins are nuclear but only E4orf6/DBP localizes to viral replication centers marked by USP7. (D) HEK293 cells were transfected with flag-E4orf6 or flag-E4orf6/DBP. Immunofluorescence was performed for flag tag (green) or integrated E1B55K (magenta). Only E4orf6 induces relocation of E1B55K from cytoplasmic aggresomes to the nucleus. Dashed white lines outline the nuclear periphery. White scale bar denotes 10 μ m.

<https://doi.org/10.1371/journal.ppat.1010797.g005>

was recruited to USP7-marked VRCs during infection, flag-E4orf6 remained markedly diffuse throughout the nucleus. Another known consequence of the interaction of E4orf6 with E1B55K is the dispersal of E1B55K “aggresomes” present in uninfected HEK293 cells [89]. We confirmed that expression of flag-E4orf6 leads to a relocation of E1B55K into a diffuse nuclear pattern within transfected cells, however E4orf6/DBP was did not reorganize E1B55K (Fig 5D). Thus, E4orf6/DBP has different cellular localizations compared to E4orf6, and does not bind to E1B55K or participate in ubiquitin ligase activities. Furthermore, our findings recontextualize previous knowledge about the localization of E4orf6 during infection.

Loss of E4orf6/DBP has minimal impact on viral genome replication or protein expression

Both E4orf6 and DBP are critical for AdV replication, and therefore any strategy to delete fragments of E4orf6/DBP would have substantial off-target effects. Instead, we opted to create a silent mutation in E4orf6 that abrogates the ability of the cellular U1 snRNP from recognizing the RNA splice donor necessary for creation of both E4orf6/7 and E4orf6/DBP [90] (Fig 6A). While this strategy involved creating a double knockout, prior work has demonstrated that E4orf6/7 is dispensable for viral replication in cell culture due to redundancy with AdV E1A [78,91]. Importantly, prior research studying E4orf6/7 deleted viruses has employed frame-shifting deletions within the downstream E4orf6/7 exon, thus the E4orf6 splice donor (and therefore E4orf6/DBP expression) remained intact. To confirm loss of E4orf6/DBP expression, we performed time-course infections in A549 cells using WT Ad5, the E4orf6/7;E4orf6/DBP double KO (Δ SS), and the previously used *dl355* Δ E4orf6 viruses (Fig 6B). Loss of E4orf6 resulted in the previously reported decrease of AdV late protein expression (Hexon, Penton, and Fiber) at 24 hpi, and a partial loss of degradation of the ubiquitin ligase substrate Rad50. In contrast, the Δ SS virus showed a complete loss of both E4orf6/7 and E4orf6/DBP, was fully competent for degrading Rad50, and showed no defect in accumulation of viral early proteins (E4orf6 and DBP) or viral late proteins (Hexon, Penton and Fiber). To complement specific loss of E4orf6/DBP we devised a retroviral system to generate cell lines which express adenoviral proteins under the control of the endogenous viral E4 promoter (S2A Fig). This system can be used to express either flag-E4orf6 or flag-E4orf6/DBP only in cells infected by AdV and expressing the viral E1A transcription factor. Using this orthogonal system, we corroborated our finding that E4orf6/DBP co-localized with VRCs, whereas E4orf6 remained broadly nucleoplasmic upon infection (S2B and S2C Fig). When these A549:E4orf6/DBP cells were infected with WT Ad5 or the Δ SS virus we again observed no defect in the expression of viral early or late proteins (Fig 6C). Retroviral expression of flag-E4orf6/DBP was expressed with similar kinetics and abundance to endogenous E4orf6/DBP.

Besides the formation of viral replication centers, DBP is a single-stranded DNA binding protein essential for replication of AdV DNA genomes [27]. We next asked whether E4orf6/DBP might modulate either of these functions of DBP. To ascertain whether the loss of E4orf6/DBP affected genome replication, we infected A549 cells with WT Ad5 or the Δ SS virus, and isolated genomic DNA to perform quantitative PCR over a time-course of infection (Fig 6D). While there was a slight defect (<2-fold) of viral genome accumulation with the Δ SS virus at 48 hpi, there was no significant decrease in DNA replication at early times post-infection. We next utilized the Δ SS virus to assay for localization of endogenous E4orf6 proteins and viral replication centers. Performing indirect immunofluorescence with the monoclonal anti-E4orf6 antibody (RSA3) during WT Ad5 infection, we confirmed previous reports from the literature that the E4orf6 signal co-localized with viral replication centers (Fig 6E). Of note, infection with the *dl355* virus entirely lacking E4orf6 still resulted in RSA3 staining at VRCs (Fig 6E). Within our 24 hpi snapshot we observed multiple cells that show the earliest signs of infection (i.e., small viral replication centers marked by USP7; closed arrow) indicating that these cells had not progressed to the late stage of infection, yet early expression of E4orf6 at VRCs was not detected. These data support a model where VRC localized E4orf6 signal was likely E4orf6/DBP expressed with late kinetics, and not E4orf6. In contrast, the Δ SS virus that expresses wildtype-levels of E4orf6, but not E4orf6/DBP, only had diffuse nuclear staining of RSA3 signal and lacked the intense staining at viral replication centers. In addition, infection with *dl356* (which lacks E4orf6/7 [78]) and *dl1002* (which lacks both E4orf6 and 6/7, but retains E4orf6/DBP due to the mutation being downstream of the splice donor [31]) both

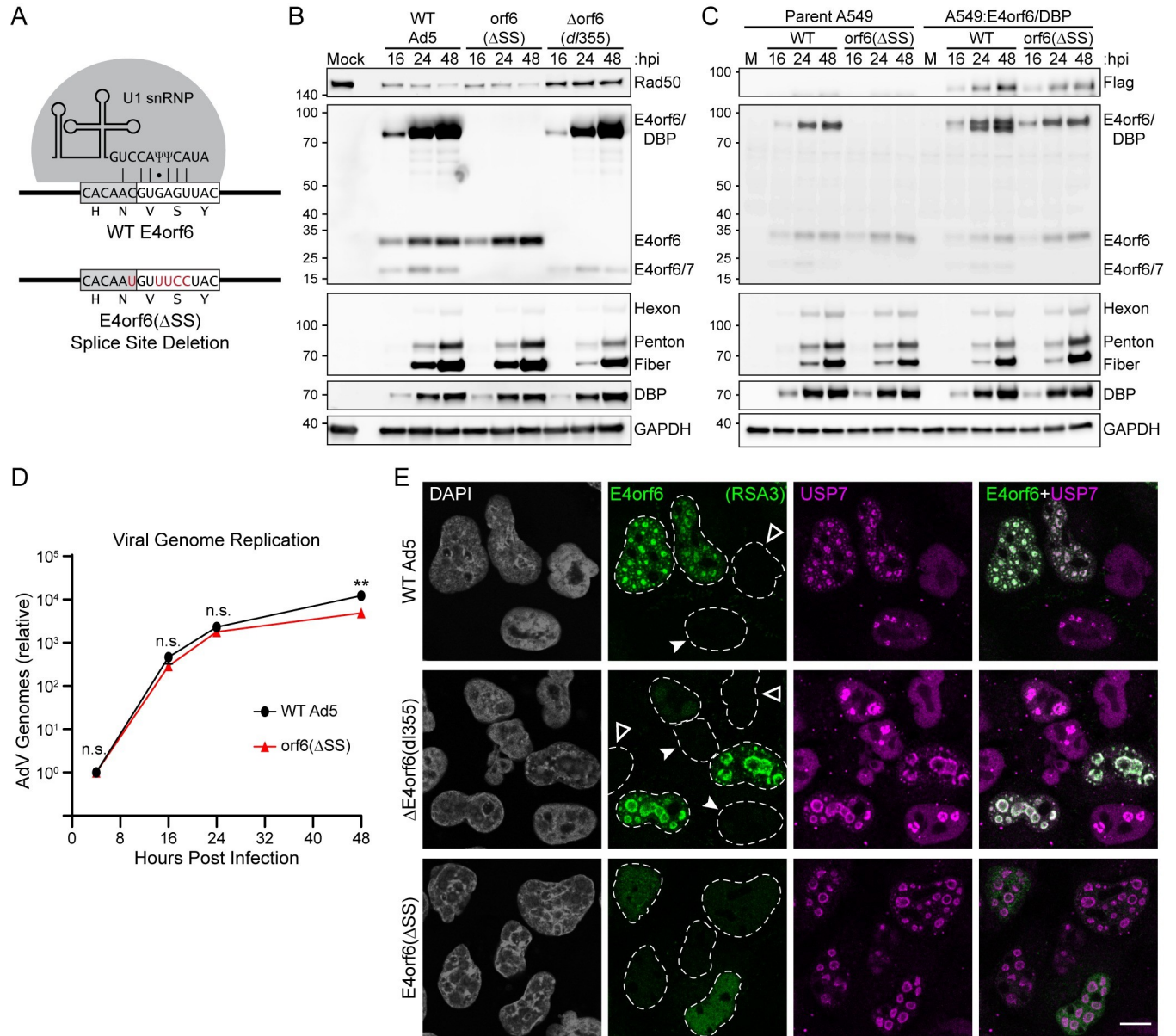


Fig 6. Loss of E4orf6/DBP has minimal impact on viral genome replication or protein expression. (A) E4orf6 splice donor is recognized by base-pairing to the cellular U1 snRNP. Silent mutations shown in red abrogate downstream splicing and were used to create the E4orf6/7, E4orf6/DBP double knockout virus (E4orf6 Δ SS). (B) A549 cells were infected with WT Ad5, E4orf6 Δ SS, or Δ E4orf6 (*dI355*) and harvested over a time-course of infection (hpi, hours post-infection). Immunoblot analysis was performed with antibodies to detect the indicated viral and cellular proteins. Kilodalton size markers are shown to the left of each blot. (C) Immunoblot analysis of a time-course infection was performed as in panel B. Cell lines were parent A549 or A549 transduced to express flag-E4orf6/DBP under control of the adenoviral E4 promoter. (D) A549 cells were infected with WT Ad5 or E4orf6 Δ SS in biological triplicate. Adenoviral genome copy number was determined by qPCR and normalized to the amount of input genomes at 4 hpi. Significance was determined by unpaired, two-tailed t-test (n.s., not significant; **, p-value<0.01). (E) A549 cells were infected with WT Ad5, Δ E4orf6, or E4orf6 Δ SS for 24 hours before immunofluorescence was performed. Cells were stained with antibodies against E4orf6 N terminal domain (RSA3, green) or cellular USP7 as a marker of viral replication centers (magenta). Dashed white lines outline the nuclear periphery. Open white arrowheads denote uninfected cells with diffuse USP7 and no viral staining, while closed arrowheads denote infected cells that lack E4orf6 staining but show USP7 at viral replication centers. White scale bar denotes 10 μ m.

<https://doi.org/10.1371/journal.ppat.1010797.g006>

retain prominent staining of RSA3 at VRCs (S3 Fig). These data corroborate our overexpression studies and confirm that endogenous E4orf6 does not localize to viral replication centers. Furthermore, while loss of E4orf6/DBP had only minimal effect on viral genome replication or protein expression, we noticed a striking change in the appearance of VRCs.

Loss of E4orf6/DBP leads to altered viral replication center morphology and small plaque phenotypes

Nuclear-replicating DNA viruses often compartmentalize viral DNA replication and other processes into membraneless organelles termed viral replication centers (VRCs) [37,92]. Early work characterized these sites as small, spherical bodies that possess liquid-like characteristics of being able to fuse and split apart. However, recent studies have expanded upon the standard model of VRCs and demonstrate dynamic morphological changes that occur upon the onset of late gene expression. While early replication centers are round or crescent shaped [93], these VRCs appear to break apart and become more diffuse before ultimately coalescing in ring-like structures within the center of the nucleus [94] (Fig 7A). This biphasic model of VRC morphology correlates with changes in the spatial location and rate of genome synthesis, and ultimately leads to the accumulation of viral DNA in virus-induced post-replicative (ViPR)

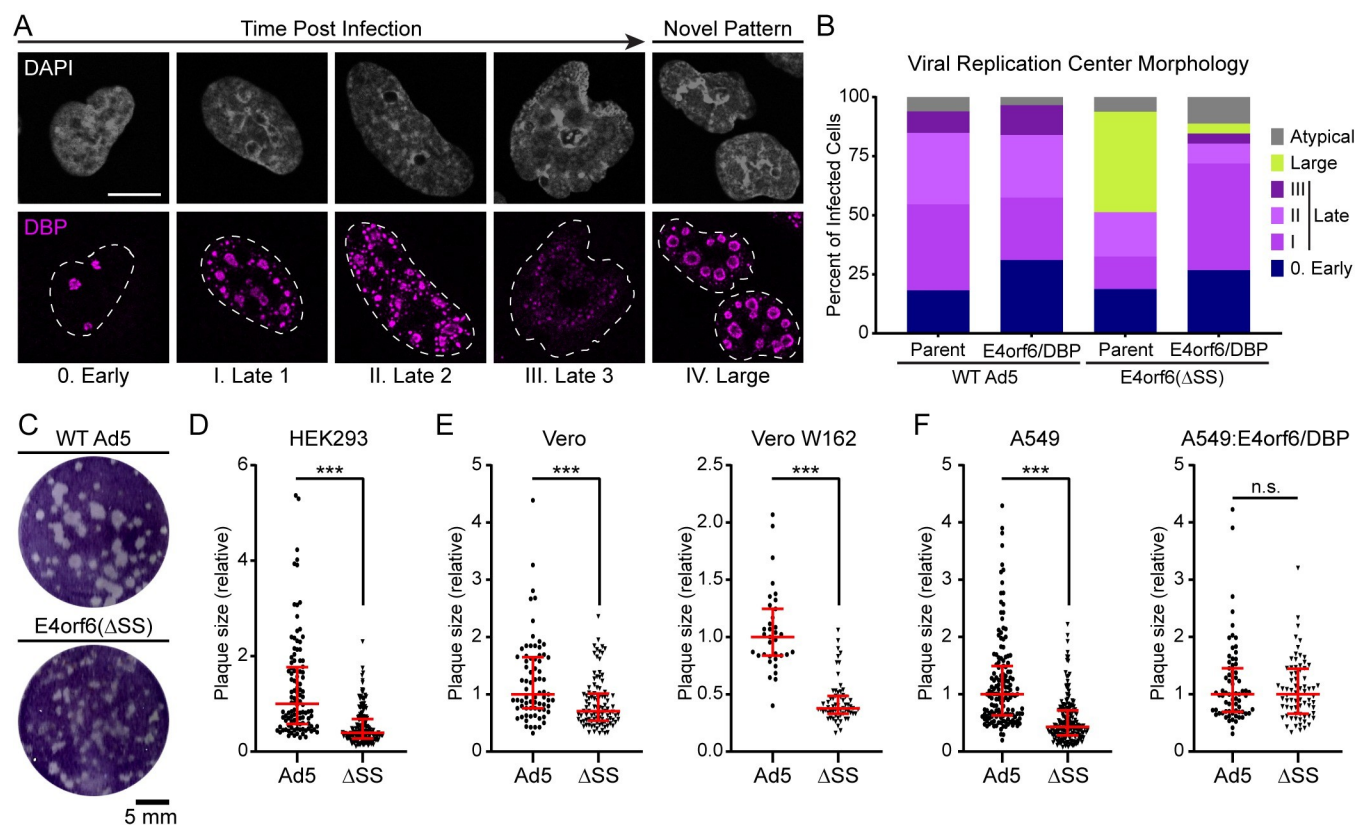


Fig 7. Loss of E4orf6/DBP leads to altered viral replication center morphology and small plaque phenotypes. (A) Wildtype adenovirus infection leads to temporal progression of viral replication center (VRC) morphology progressing from early stage (0. Early) through three distinct stages of late replication centers (Late 1–3). A novel, large replication center morphology was seen with infection of E4orf6 Δ SS virus. Representative immunofluorescence images for each stage are shown by staining A549 cells with VRC marker DBP (magenta) and DAPI for DNA (grey). (B) Parent A549 (Parent) or A549 cells expressing E4p-flag-E4orf6/DBP (E4orf6/DBP) were infected with WT Ad5 or E4orf6 Δ SS virus for 24 hours. Viral replication centers were stained for DBP, and morphology was scored in a blinded manner using the key provided in panel A. The percentage of infected cells demonstrating each replication center morphology is shown as bar chart. Data are representative of three independent experiments. (C) HEK293 cells were infected with limiting dilution of WT Ad5 or E4orf6 Δ SS for six days to allow the formation of plaques. Plaque formation was negative stained with crystal violet and imaged. Scale bar shows 5 mm. (D) WT Ad5 or E4orf6 Δ SS (Δ SS) plaque size in HEK293 cells was quantified with ImageJ and normalized to the median plaque size in WT infection. Individual plaques are plotted as a single point. (E) Plaque size was quantified in Vero cells or Vero-W162 cells. W162 cells express the adenovirus E4 region *in trans* and complement the loss of E4orf6/7 in the Δ SS virus. (F) Plaque size was quantified in A549 cells or A549:E4orf6/DBP cells. The A549:E4orf6/DBP cells express E4orf6/DBP under control of the E4 promoter *in trans*, and thus complement the loss of E4orf6/DBP in the Δ SS virus. For all plaque assays, red error bars denote median and interquartile range. Statistical significance was performed using non-parametric Mann-Whitney t-test (n.s., not significant; ***, p-value < 0.001).

<https://doi.org/10.1371/journal.ppat.1010797.g007>

bodies [95,96]. While the progression through the distinct VRC morphologies is correlated with time post-infection, in any given snapshot of infection multiple classes of VRCs can be detected due to the asynchronous nature of viral replication on a cell-by-cell basis. Strikingly, when we visualized viral infection by staining with DBP (Fig 7A) or USP7 (Fig 6E) during infection with the Δ SS virus, we noticed a large fraction of cells possessed a novel VRC morphology characterized by exceptionally large, round or donut-shaped bodies. This novel VRC morphology was not seen with the loss of only E4orf6/7 (S3 Fig). To quantify VRC morphology we performed phenotype analysis by confocal microscopy for A549 cells infected by WT Ad5 or Δ SS virus for 24 hours within three independent replicates. All infected cells were categorized in a blinded fashion into early replication centers (0. Early), one of three distinct late-stage VRC morphologies (I., early VRCs with numerous small foci; II., more complete breakup of VRCs into a multitude of medium bodies; or III., ring-like patterns with ViPR bodies), the novel phenotype (IV., large), or a sixth category of minority variants that defied typical classification (Atypical). At this time point greater than 60% of WT-infected cells had progressed to the late stage of VRC morphology (Fig 7B). While a similar fraction of Δ SS-infected cells were infected and reached the early stage of VRC morphology, greater than 40% of cells now possessed the novel large VRC phenotype. When WT or Δ SS virus were used to infect A549: E4orf6/DBP cells that complement E4orf6/DBP under control of the endogenous viral promoter, the amount of Δ SS-infected cells displaying the novel VRC phenotype was dramatically reduced and instead displayed normal late-stage VRC morphologies (Fig 7B). This implies that while E4orf6/DBP is only a small fraction of the total pool of DBP protein isoforms, it has an outsized impact on the evolution of VRC morphology within infected cells.

Despite the striking change in VRC morphology, the Δ SS virus has minimal effects on genome accumulation or viral protein expression during a single cycle of infection (Fig 6). We next asked what effect the loss of E4orf6/DBP might have on viral replication during multiple cycles of infection. One measure of viral spread over multiple cycles of infection is the ability to form infectious plaques in a monolayer of cells. We performed serial dilutions of WT Ad5 and Δ SS virus on HEK293 and allowed plaques to form underneath agarose for exactly five days (Fig 7C). Despite the similar number of total plaques, the plaques formed by WT Ad5-infection were much larger than those formed by Δ SS-infection in the same amount of time, as could be quantified by ImageJ software (Fig 7D). Importantly, a virus that only lacks E4orf6/7 formed plaques that were not significantly different than WT infection of HEK293 cells (S4 Fig). When we performed the same assay in Vero cells, or Vero W162 cells that can complement the lack of E4orf6/7 but not E4orf6/DBP during Δ SS-infection, we again quantified a significant reduction in plaque size when comparing the WT Ad5 to Δ SS virus (Fig 7E). Finally, while a significant disparity in plaque size could once again be quantified in parental A549 cells, plaques formed on A549:E4orf6/DBP cells showed no significant difference in plaque size (Fig 7F). Therefore, these data suggest that E4orf6/DBP contributes to the cell-to-cell spread or cell killing during a multi-cycle viral infection.

Discussion

DNA viruses encode large amounts of information in compact genomes through alternative splicing, overlapping transcripts, and transcription from both strands of the genome. Here we integrate short-read cDNA sequencing and long-read direct RNA sequencing to re-annotate both the Ad5 DNA genome and RNA transcriptome. Using high quality and high depth short-read sequencing, we were able to detect SNPs within the transcribed regions of the genome approaching 100% penetrance, indicating that these sites were likely present in the genome and not due to RNA editing or modifications. We recapitulated the known TSS and CPAS

sites throughout the Ad5 genome, and annotated novel splicing events within the viral transcriptome. Of these 35 novel RNAs, 15 are likely to encode for altered ORFs including multiple fusion transcripts that span transcriptional units previously thought to be distinct. Of these, we focused on one transcript, E4orf6/DBP, which we show produces a novel conserved protein with late gene kinetics. Study of E4orf6/DBP recontextualized our knowledge of the critical AdV effector protein E4orf6, as well as revealed a unique role for this chimeric viral protein in regulating viral replication center morphology and cell-to-cell spread. Overall, we have provided a more comprehensive annotation of a complex viral transcriptome that highlights multiple new gene products for future study.

Using RNA-seq data to call SNPs in viral DNA genomes represents a compelling approach to identify SNPs, since high quality short-read sequencing data sets already exist for many DNA viruses [75,97–100]. While half of the SNPs we called were synonymous or in non-coding regions, missense mutations have the potential to change the coding sequence of protein amino acids in meaningful ways. In addition, while SNPs are often tolerated during alignment of RNA-seq data, annotation of the correct primary amino acid sequence is critical for downstream analysis of mass spectrometry data [101]. These SNPs were validated by DNA-sequencing of a common Ad5 plasmid used in viral mutagenesis by multiple laboratories, highlighting a need to revisit existing reference genomes for common DNA viruses. Of note, Westergren-Jakobssen *et al.*, also discovered SNPs by directly sequencing the DNA of Ad2 [62], however none of these SNPs overlapped with the nucleotide polymorphisms we detected in the related Ad5.

Previous studies of AdV transcription detected numerous splice sites beyond those employed by known isoforms. However, the constraints of short-read sequencing precluded proper assembly of these sites into full-length transcripts [44,54,73]. Furthermore, targeted expression analysis over a time-course of infection was limited to already known transcripts [55]. In the past two years, two groups have independently used direct RNA sequencing on the Oxford Nanopore Technology platform to profile adenoviral RNA expression of two related serotypes, Ad2 and Ad5 [61,62]. While both studies highlighted an exceptional diversity of viral transcripts produced during infection, many novel transcripts were expressed at very low levels or only seen very late after infection. By combining high-depth short-read sequencing, direct RNA long-read sequencing, and stringent cutoffs, we can now report high confidence novel transcripts as well as show regulated expression over a time-course of infection. Nearly all of the high confidence novel transcripts we report on were also detected by both prior studies, including the E4orf6/DBP transcript we have focused on (S2 Table). We have also added ORF predictions to previously detected splice sites, such as L2-111R, X-Z-Fiber, and the pVIII ORF derived from splicing directly from the tripartite leader to the L4-33K splice acceptor. While this last site was previously predicted to lead to the expression of a small 42 amino acid ORF [54], we propose that this transcript instead primarily encodes for pVIII with a small upstream ORF, as it is over five times as abundant as the canonical pVIII spliced RNA.

Of the novel transcripts we have so far detected, all of them display delayed late kinetics during infection. Of particular interest is the transcript encoding for a fusion event between the E4 transcriptional unit and the E2 transcriptional unit. This transcript, E4orf6/DBP, would have to skip the canonical E4 cleavage and polyadenylation site for the pre-mRNA to progress downstream to DBP for splicing. The three transcripts displaying this pattern, including E4-promoter driven DBP and frameshifted E4-Unk, are all much more abundant during the late phase of infection even though canonical E4 transcripts are expressed early. Other transcripts displaying readthrough of polyadenylation include an E1A-9s mRNA, E3-driven Fiber, as well as nearly all canonical late transcripts. In fact, Fiber driven by the major late promoter must bypass seven actionable polyadenylation sites before revealing the 3' splice acceptor. It will be interesting to see the extent to which differential polyadenylation is regulated during

the life cycle of the virus, as has been previously reported for herpes simplex virus [58,99,102]. Alternatively, AdV late transcripts might employ recursive splicing to lessen the need for long introns [103].

Of the 20 novel Ad5 protein isoforms we detected within our reannotated transcriptome, we have been able to provide mass spectrometry confirmation for 12 candidates (60%). For the remaining eight proteins absence of evidence is not evidence of absence, since mass spectrometry requires peptide fragments of ideal mass and charge for detection. By exploring just one of these novel proteins, we have been able to provide genetic and biochemical evidence for the expression, evolutionary conservation, and function of E4orf6/DBP. While it appears that E4orf6/DBP regulates both the morphology of viral replication centers and cell-to-cell spread, the molecular mechanism of these functions remains to be determined. The DBP protein is known to bind single-stranded DNA and oligomerize to form chains upon replicating DNA [27]. While loss of E4orf6/DBP did not affect viral DNA accumulation, it is tempting to speculate that addition of the E4orf6 fragment to the N-terminus of DBP might affect the localization of viral DNA replication or the ability of DBP to form oligomers. Prior to this study, the most recently discovered protein of Ad5, UXP, was shown to be generated from an mRNA that is co-terminal with DBP, translated in an alternate reading frame to DBP, localized to DBP-marked replication centers, and led to altered VRC morphology in its absence [48,49]. The complete protein complement of viral replication centers is unknown, and it is likely that E4orf6/DBP might alter viral or cellular protein components that localize to these compartments. These interactions might be modulated by affecting the N-terminus of DBP which contains substantial intrinsic disorder and is post-translationally modified by phosphorylation, both of which are known to mediate protein-protein interactions. Finally, recent evidence has provided support for a model where AdV particles are packaged in proximity to VRCs [104]. Changes in AdV particle production could result from modification of these VRCs or access to the viral DNA, and these alterations might impact cell-to-cell spread.

A complication arising from the study of viral pathogens is that otherwise simple knockouts or mutations might alter the coding potential of nearby or anti-sense coding units [105]. In our characterization of E4orf6/DBP we have discovered that the previously understood localization of E4orf6 to VRCs was incorrect due to the presence of a previously undiscovered protein being recognized by the same monoclonal antibody. Importantly, future research should identify whether the known functions of existing viral ORFs can be explained, at least in part, by the presence of these novel isoforms. We have provided an example of how modern tools in RNA sequencing and mass spectrometry can offer a more complete reference and framework for undertaking research on a given complex pathogen. In doing so, we have added new depth to a common virus that has been exceptionally well-characterized for many decades.

Methods

Cell culture

All cell lines were obtained from American Type Culture Collection (ATCC) and cultured at 37°C and 5% CO₂. A549 cells (ATCC CCL-185) were maintained in Ham's F-12K medium (Gibco, 21127-022) supplemented with 10% v/v FBS (VWR, 89510-186) and 1% v/v Pen/Strep (100 U/ml of penicillin, 100 µg/ml of streptomycin, Gibco, 15140-122). HEK293 cells (ATCC CRL-1573), Vero cells (ATCC CCL-81), and W162 cells (Kind gift of G. Ketner) were grown in DMEM (Corning, 10-013-CV) supplemented with 10% v/v FBS and 1% v/v Pen/Strep. All cell lines tested negative for mycoplasma infection and were routinely tested afterwards using the LookOut Mycoplasma PCR Detection Kit (Sigma-Aldrich).

Viral infections

Adenovirus serotype 5 (Ad5), Ad4, Ad9, and Ad35 were originally purchased from ATCC. Ad5 $\Delta E4orf6$ (*dl355*), $\Delta E4orf6/7$ (*dl356*), $\Delta E4orf6,6/7$ (*dl1002*), $E4in6,6/7$ and $\Delta E4$ (*dl1004*) were previously described and obtained from D. Ornelles [31,78,106]. Construction of the $E4orf6$ splice site deleted virus (ΔSS) was performed as previously described [107,108], using Ad5 cloning plasmids that were a kind gift from P. Hearing. In brief, we used site-directed mutagenesis to silently mutate the $E4orf6$ splice donor in the plasmid pBS-E4. Subsequently, a linear HindIII/KpnI fragment containing the ΔSS mutation was electroporated into BJ5183 *E. coli* cells (Fisher Scientific, 50-125-019) with the ClaI linearized pTG3602*inOrf3* Ad5 plasmid. Overnight colonies were screened for insertion of the splice-site deletion, expanded in DH5 α *E. coli*, and maxi-prepped. Insertion of the splice-site deletion repair of the $E4orf3$ insert were verified with Sanger Sequencing (GeneWiz), and integrity of the viral plasmid was compared by restriction-fragment mapping to wildtype Ad5 (pTG3602). All viruses were expanded on HEK293 cells, purified using two sequential rounds of ultracentrifugation in CsCl gradients, and stored in 40% v/v glycerol at -20°C (short term) or -80°C (long term). Viral stock titer was determined on HEK293 cells by plaque assay, and all subsequent infections were performed at a multiplicity of infection (MOI) of 10 PFU/cell. Cells were infected at 80–90% confluent monolayers by incubation with diluted virus in a minimal volume of low serum (2%) F-12K for two hours. After infection viral inoculum was removed by vacuum and full serum growth media was replaced for the duration of the experiment.

Immunoblotting and antibodies

For immunoblotting analysis protein samples were prepared by directly lysing cells in lithium dodecyl sulfate (LDS) loading buffer (NuPage) supplemented with 1% beta-mercaptoethanol (BME) and boiled at 95°C for 10 min. Equal volumes of protein lysate were separated by SDS-PAGE in MOPS buffer (Invitrogen) before being transferred onto a nitrocellulose membrane (Millipore) at 35 V for 90 minutes in 20% methanol solution. Membranes were stained with Ponceau to confirm equal loading and blocked in 5% w/v non-fat milk in TBST supplemented with 0.05% w/v sodium azide. Membranes were incubated with primary antibodies in milk overnight, washed for three times in TBST, incubated with HRP-conjugated secondary (Jackson Laboratories) for 1 h and washed an additional three times in TBST. Proteins were visualized with Pierce ECL or Pico Western Blotting Substrate (Thermo Scientific) and detected using a Syngene G-Box. Images were processed and assembled in Adobe Photoshop and Illustrator CC 2022. Antibodies against adenovirus proteins include rabbit polyclonal against adenovirus Hexon, Penton, and Fiber (Gift from J. Wilson, WB 1:10,000), mouse anti-DBP (Gift from A. Levine, Clone: B6-8, WB 1:1000, IF 1:400), and two mouse anti- $E4orf6$ (Gift from P. Hearing, Clone: RSA3 or M45, WB 1:400, IF 1:400). Antibodies against cellular proteins include RAD50 (GeneTex GTX70228, WB 1:1000), MRE11 (Novus NB100–142, WB 1:1000), USP7 (Bethyl Laboratories A300–033A, IF 1:500), and GAPDH (GeneTex GTX100118, WB 1:30,000).

Indirect immunofluorescence assay

Cells were grown on glass coverslips in 24-well plates, mock-infected or infected with virus for the appropriate time, washed twice with PBS and then fixed in 4% w/v Paraformaldehyde for 15 minutes. Cells were permeabilized with 0.5% v/v Triton-X in PBS for 10 mins, and blocked in 3% w/v BSA in PBS (+ 0.05% w/v sodium azide) for one hour. Primary antibody dilutions were added to coverslips in 3% w/v BSA in PBS (+ 0.05% w/v sodium azide) for 1 h, washed with 3% BSA in PBS three times, followed by secondary antibodies and 4,6-diamidino-

2-phenylindole (DAPI) for one hour. Secondary antibodies were used at 1:500 dilution and conjugated to Alexa-Fluor 488 (Invitrogen A-11001 or A-11008), or Alexa-Fluor 568 (Invitrogen A-11004 or A-11011) in anti-mouse or anti-rabbit. Coverslips were mounted onto glass slides using ProLong Gold Antifade Reagent (Cell Signaling Technologies). Immunofluorescence was visualized using a Zeiss LSM 710 Confocal microscope (Cell and Developmental Microscopy Core at UPenn) and ZEN 2011 software. Images were processed in FIJI (v1.53c) and assembled in Adobe Photoshop and Illustrator CC 2022.

RNA Isolation, DNA Isolation, and Quantitative PCR

Total RNA was isolated from cells by either TRIzol extraction (Thermo Fisher) or RNeasy Micro kit (Qiagen), following manufacturer protocols. RNA was treated with RNase-free DNase I (Qiagen), either on-column or after ethanol precipitation. To test quality, RNA was converted to complementary DNA (cDNA) using 1 µg of input RNA in the High-Capacity RNA-to-cDNA kit (Thermo Fisher). Total DNA was isolated using the PureLink Genomic DNA kit (Invitrogen). Quantitative PCR was performed using the standard protocol for SYBR Green reagents (Thermo Fisher) in a QuantStudio 7 Flex Real-Time PCR System (Applied Biosystems). Primers used included E1A (Fp: GTACCGGAGGTGATCGATCTTA, Rp: TCAGGCTCAGGTTTCAGACA), E4orf6 (Fp: TTTACTGGAAATATGACTACGTC, Rp: AAGACCTCGCACGTAAC), E4orf6/7 (Fp: GTAGGGATCGTCTACCTCCTTT, Rp: AGAAGTCCACGCGTTGTG), E4orf6/DBP (Fp: GTAGGGATCGTCTACCTCCTTT, Rp: CTGGCCATTTCTTCTCGTTG), Fiber (Fp: GAAAGGCGTCTAACCAGTCA, Rp: AAAGGCACAGTTGGAGGAC), DBP (Fp: GCCATTGCGCCCAAGAAGAA, Rp: CTGTCCACGATTACCTCTGGTGAT), HPRT1 (Fp: TGACACTGGCAAAACAATGCA, Rp: GGTCTTTTCACCAGCAAGCT), and Tubulin (Fp: CCAGATGCCAAGTGACAA GAC, Rp: GAGTGAGTGACAAGAGAAGCC).

Plaque assay

Viral particles were serially diluted in DMEM supplemented with 2% v/v FBS and 1% v/v Pen/Strep to infect confluent monolayers of HEK293 cells, Vero cells, or A549 cells seeded in 6-well plates. After incubation for 2 h at 37°C, the infection media was removed, and cells were overlaid with DMEM containing 0.45% w/v SeaPlaque agarose (Lonza) in addition to 2% v/v FBS and 1% v/v Pen/Strep. Between 5 to 6 days post-infection plaques were fixed with 10% (wt/vol) trichloroacetic acid was added to each well for 30 min at room temperature. Plaques were stained using 1% w/v crystal violet in 50% v/v ethanol. Plaque assays were scanned and resulting images were processed in FIJI (v1.53c) to determine plaque area. Plaque size in arbitrary units were normalized to the median plaque size from wildtype Ad5 infection in a given cell line.

Proteomics data analysis

Biological triplicate cells were lysed with 8M urea and sonicated in a Diagenode Biorupter bath sonicator at 4°C on high 30 seconds on and 30 seconds off for 5 minutes. The cells were centrifuged at 10,000xg for 10 minutes at 4°C and the lysates transferred to a new tube. The lysate was then reduced with 5mM dithiothreitol (DTT) for 1 hour at RT and alkylated using 10mM iodoacetamide (IAM) for 45 minutes in the dark and 15 minutes in the light, followed by trypsin at a 1:50 ratio overnight at RT. The following day 200 µg of peptides for each time point and bio-replicate were cleaned using Hamilton spin tip columns. Samples were separated on a standard linear 80-minute gradient from 5% to 60%, using standard proteomics buffers of 0.1% formic acid in aqueous and 0.1% formic acid in 80% acetonitrile (ACN) using a Thermo

Dionex 3000 LC. Samples were quantified using a Thermo Fusion MS instrument and batch randomized to account for instrument variation. The PRM method was designed with the MS1 having a window of 350–1200m/z, resolution of 60K, AGC target of 100% and MIT of 50ms, the tSIM MS² scan having an AGC target of 200% and MIT of 54ms, resolution of 30K and HCD fragmentation of 28%.

Thermo raw files were imported into Skyline using 2 missed cleavages with a minimum length of 5 and a maximum length of 30 amino acids. Fragments required 4 or more product ions within 5 minutes of the predicted RT window and a mass tolerance of 10ppm or less to be considered for evaluation. Peptides were manually evaluated with the requirement of 5 or more overlapping fragments to be considered a “real” identification in comparison to mock.

Data from *Dybas et al.* [74] was downloaded from the PRIDE database and searched using both Sequest and Byonic sequentially within the Thermo Proteome Discoverer 2.4 platform [109,110]. Using both the new Adenovirus fasta with the previously used human Uniprot fasta database searched at 10ppm MS¹ precursor tolerance and 0.02 Da fragment search tolerance for Sequest and 10ppm fragment search tolerance for Byonic. Both search programs used a missed cleavage of 2 and fixed modifications for carbamidomethylation for cysteines and variable modifications for methionine oxidation and N-terminal acetylation. Each search had an independently applied 1% FDR for identified spectra followed by a protein level 1% FDR. Manual validation of peptides was performed with Proteome Discoverer software and visualized through the Coon lab interactive peptide spectral annotator (IPSA) [111].

Illumina sequencing and mapping

Total RNA from three biological replicates of A549 cells infected with Ad5 for 24 hours were sent to Genewiz for preparation into strand-specific RNA-Seq libraries. Libraries were then run spread over three lanes of an Illumina HiSeq 2500 using a 150bp paired-end protocol. Raw reads were mapped to the GRCh37/hg19 genome assembly and the Ad5 genome using the RNA-seq aligner GSNAP [68] (version 2019-09-12). The algorithm was given known human gene models provided by GENCODE (release_27_hg19) to achieve higher mapping accuracy. We used R package ggplot2 for visualization. Downstream analysis and visualization was done using deepTools2 [112]. Splice junctions were extracted using regtools [113] and visualized in Integrative Genomics Viewer [114].

Variant calling and whole-plasmid sequencing

Illumina RNA-seq reads were aligned to the Ad5 genome obtained from NCBI (https://www.ncbi.nlm.nih.gov/nucore/AC_000008) using GSNAP [68]. To identify variants such as single nucleotide polymorphisms (SNPs) and insertions/deletions (InDels), we combined mpileup and call from the bcftools (v1.9) package [64,65]. Here we used the following flags “--redo-BAQ --min-BQ 30 --per-sample-mF” and “--multiallelic-caller --variants-only” respectively. Finally, we only considered variants if they were called significantly in all 3 replicates. We only observed SNPs but no InDels. To confirm these results, we sent the WT Ad5 plasmid (pTG3602) for whole-plasmid sequencing to PlasmidSaurus (Eugene, Oregon).

Direct RNA sequencing on nanopore arrays

Direct RNA sequencing libraries were generated from 800–900 ng of poly(A) RNA, isolated using the Dynabeads mRNA Purification Kit (Invitrogen, 61006). Isolated poly(A) RNA was subsequently spiked with 0.3 µl of a synthetic Enolase 2 (ENO2) calibration RNA (Oxford Nanopore Technologies Ltd.) and prepared for sequencing using standard protocol steps previously described [58,67]. Sequencing was carried out on a MinION Mk1b with R9.4.1 (rev D)

flow cells (Oxford Nanopore Technologies Ltd.) for 20 hours and generated 550,000–770,000 sequence reads per dataset. Raw fast5 datasets were basecalled using Guppy v6.0.1 -c rna_r9.4.1_70bps_hac.cfg -r --calib_detect --trim_strategy rna --reverse_sequence true and subsequently aligned against the adenovirus Ad5 reference genome (AC_000008.1) using MiniMap2.24 (-ax splice -k14 -uf --secondary = no), a splice aware aligner [69]. Resulting SAM files were parsed using SAMtools v1.9 [115].

Defining TSS and CPAS

Transcription start sites (TSS) as well as RNA cleavage and polyadenylation sites were identified as follows. Sorted BAM files containing sequence reads aligned to the Ad5 genome were parsed to BED12 files using BEDtools [116], separated by strand, truncated to their 5' and 3' termini, and output as BED6 files. Peak regions denoting TSS and CPAS were identified using the HOMER [117] findpeaks module (-o auto -style tss) using a --localSize of 100 and 500 and --size of 15 and 50 for TSS and CPAS, respectively. TSS peaks were compared against Illumina annotated splice sites to identify and remove peak artefacts derived from local alignment errors around splice junctions. Only TSS and CPAS peaks found in both 12 hpi and 24 hpi datasets were used for downstream analysis, except the TSS for the L4 promoter which was only detected at 12 hpi. To predict CPAS sites on the viral genome, we also used the RNA-seq aligner ContextMap2 (version 2.7.9) [70] which has poly(A) read mapping implemented on our short-read data. To run this tool, we used the following optional flags “-aligner_name bowtie-polyA-strandspecific”. Due to the previously reported errors when using ContextMap2 at very high read depth, we chose to randomly subsample 10 million, 20 million and 30 million and run the tool on each of the subsets. We only report poly(A) sites if they were called in all three replicates and in at least two of the subsample groups.

Splice junction correction and sequence read collapsing

Illumina-assisted correction of splice junctions in direct RNA-Seq data was performed using FLAIR v1.3 [72] in a stranded manner. Briefly, Illumina reads aligning to the Ad5 genome were split according to orientation and mapping strand [-f83 & -f163 (forward) and -f99 & -f147 (reverse)] and used to produce strand-specific junction files that were filtered to remove junctions supported by less than 100 Illumina reads. Direct RNA-Seq reads were similarly aligned to the Ad5 genome and separated according to orientation [-F4095 (forward) and -f16 (reverse)] prior to correction using the FLAIR correct module (default parameters). Resulting BED12 files were parsed to extend the termini of each individual sequence read to the nearest TSS and CPAS with BlockStarts and BlockSizes (BED12 cols 11 & 12) corrected to reflect this. BED12 files were subsequently collapsed by identifying all reads sharing the same BlockStarts and BlockSizes and reducing these to a single representative. Resulting data were visualized along with the raw read data using IGV [114] and low abundance isoforms (supported by less than 500 junctional reads or 10 full-length reads from Illumina or nanopore data, respectively) removed prior to producing the final annotation.

Isoform counting

Using our new Ad5 annotation, we generated a transcriptome database by parsing our GFF3 file to a BED12 file using the *gff3ToGenePred* and *genePredtoBED* functions within UCSCutils (<https://github.com/itsvenu/UCSC-Utills-Download>) and subsequently extracting a fasta sequence for each transcript isoform using the *getfasta* function within BEDtools [116]. Direct RNA-Seq reads were then aligned against the transcriptome database using parameters

optimized for transcriptome-level alignment (minimap2 -ax map-ont -p 0.99). Isoform counts were generated by filtering only for primary alignments (SAM flag 0) with a mapping quality (MapQ) > 0.

Supporting information

S1 Table. Nucleotide locations of TSS and CPAS. Nucleotide locations for transcript start sites (TSS) and cleavage and polyadenylation sites (CPAS) within re-annotated transcriptome. (XLSX)

S2 Table. Comparison between novel transcripts identified in this study and recent long-read sequencing studies of adenovirus. Table comparing the presence or absence of the 35 novel transcripts published in this study versus Donovan-Banfield et al. 2020 (Ad5 in MRC5 fibroblasts) and Westergren Jakobsson et al 2021 (Ad2 in IMR90 fibroblasts). (XLSX)

S1 Fig. Expanded adenoviral transcriptome. Expanded transcriptome showing all high confidence transcripts. Color scheme is the same as Fig 2. All transcripts are labeled as denoted in Fig 3. (TIF)

S2 Fig. E4orf6 and E4orf6/DBP have different localization patterns. (A) The retroviral plasmid pLNE4 was designed to allow the expression of transgenes under the control of the Ad5 E4 promoter. This has the effect of only expressing transgenes after viral infection and E1A expression, thus allowing for the regulated expression of potentially toxic viral proteins. (B) A549 cells were transduced with a pLNE4 vector expressing flag-tagged E4orf6. Uninfected cells (open arrowhead) show no expression of flag-tagged proteins. DAPI staining marks nuclei. (C) Same as in panel B, but A549 cells were transduced with a pLNE4 vector expressing flag-tagged E4orf6/DBP. While E4orf6 is broadly nucleoplasmic, E4orf6/DBP strongly localizes to viral replication centers marked by the cellular protein USP7 (magenta). Immunofluorescence was performed 24 hours post-infection. (TIF)

S3 Fig. N-terminal E4orf6 localization of diverse E4 mutant viruses. (A) A549 cells were infected with the indicated mutant adenoviruses for 24 hours and proteins were resolved by immunoblot. RSA3 antibody shows various banding patterns for E4orf6/DBP, E4orf6, and E4orf6/7. * denotes E4orf6 frameshift product that is produced downstream of the common E4orf6 splice donor in *dl1002*. (B) A549 cells were infected with the viruses from A for 24 hours before immunofluorescence was performed. Cells were stained with antibodies against E4orf6 N terminal domain (RSA3, green) or cellular USP7 as a marker of viral replication centers (magenta). White scale bar denotes 10 μ m. (TIF)

S4 Fig. Loss of E4orf6/7 does not result in a small plaque phenotype. (A) HEK293 cells were infected with limiting dilution of WT Ad5 or Δ E4orf6/7 (*dl356*) for six days to allow the formation of plaques. Plaque formation was negative stained with crystal violet and imaged. Scale bar shows 5 mm. (B) WT Ad5 or Δ E4orf6/7 plaque size in HEK293 cells was quantified with ImageJ and normalized to the median plaque size in WT infection. Individual plaques are plotted as a single point. Red error bars denote median and interquartile range. Statistical significance was performed using non-parametric Mann-Whitney t-test (n.s., not significant). (TIF)

Acknowledgments

We thank members of the Weitzman, Depledge, and Mohr/Wilson Labs for insightful discussions and input. We are grateful to A. Berk, P. Hearing, G. Ketner, A. Levine, D. Ornelles, and J. Wilson for generous gifts of viruses, antibodies, and reagents. We thank the Upenn Cell and Developmental Biology Microscopy Core for imaging assistance.

Author Contributions

Conceptualization: Alexander M. Price, Matthew D. Weitzman.

Data curation: Alexander M. Price, Katharina E. Hayer, Daniel P. Depledge.

Formal analysis: Alexander M. Price, Robert T. Steinbock, Richard Lauman, Matthew Charman, Katharina E. Hayer, Daniel P. Depledge.

Funding acquisition: Alexander M. Price, Angus C. Wilson, Benjamin A. Garcia, Matthew D. Weitzman.

Investigation: Alexander M. Price, Robert T. Steinbock, Richard Lauman, Matthew Charman, Namrata Kumar, Edwin Halko, Krystal K. Lum, Monica Wei, Daniel P. Depledge.

Methodology: Alexander M. Price.

Visualization: Alexander M. Price.

Writing – original draft: Alexander M. Price, Matthew D. Weitzman.

Writing – review & editing: Alexander M. Price, Robert T. Steinbock, Richard Lauman, Angus C. Wilson, Daniel P. Depledge, Matthew D. Weitzman.

References

1. Berk AJ. Adenoviridae. 6th ed. In: Knipe David M., Howley Peter M., editors. *Fields Virology*. 6th ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2013. pp. 1704–1731.
2. Khanal S, Ghimire P, Dharmoon AS. The Repertoire of Adenovirus in Human Disease: The Innocuous to the Deadly. *Biomedicine*. 2018;6. <https://doi.org/10.3390/biomedicine6010030> PMID: 29518985
3. Chroboczek J, Bieber F, Jacrot B. The sequence of the genome of adenovirus type 5 and its comparison with the genome of adenovirus type 2. *Virology*. 1992; 186: 280–285. [https://doi.org/10.1016/0042-6822\(92\)90082-z](https://doi.org/10.1016/0042-6822(92)90082-z) PMID: 1727603
4. Davison AJ, Benko M, Harrach B. Genetic content and evolution of adenoviruses. *J Gen Virol*. 2003; 84: 2895–2908. <https://doi.org/10.1099/vir.0.19497-0> PMID: 14573794
5. Berk AJ. Recent lessons in gene expression, cell cycle control, and cell biology from adenovirus. *Oncogene*. 2005; 24: 7673–85. <https://doi.org/10.1038/sj.onc.1209040> PMID: 16299528
6. Chow LT, Gelinas RE, Broker TR, Roberts RJ. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell*. 1977; 12: 1–8. [https://doi.org/10.1016/0092-8674\(77\)90180-5](https://doi.org/10.1016/0092-8674(77)90180-5) PMID: 902310
7. Berget SM, Moore C, Sharp PA. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *PNAS*. 1977; 74: 3171–3175. <https://doi.org/10.1073/pnas.74.8.3171> PMID: 269380
8. Sommer S, Salditt-Georgieff M, Bachenheimer S, Darnell JE, Furuichi Y, Morgan M, et al. The methylation of adenovirus-specific nuclear and cytoplasmic RNA. *Nucleic Acids Res*. 1976; 3: 749–65. <https://doi.org/10.1093/nar/3.3.749> PMID: 1272797
9. Philipson L, Wall R, Glickman G, Darnell JE. Addition of polyadenylate sequences to virus-specific RNA during adenovirus replication. *Proc Natl Acad Sci U S A*. 1971; 68: 2806–9. <https://doi.org/10.1073/pnas.68.11.2806> PMID: 5315962
10. Berk AJ. Discovery of RNA splicing and genes in pieces. *Proc Natl Acad Sci USA*. 2016; 113: 801–805. <https://doi.org/10.1073/pnas.1525084113> PMID: 26787897
11. Montell C, Fisher EF, Caruthers MH, Berk AJ. Resolving the functions of overlapping viral genes by site-specific mutagenesis at a mRNA splice site. *Nature*. 1982; 295: 380–384. <https://doi.org/10.1038/295380a0> PMID: 7057903

12. Winberg G, Shenk T. Dissection of overlapping functions within the adenovirus type 5 E1A gene. *EMBO J*. 1984; 3: 1907–1912. <https://doi.org/10.1002/j.1460-2075.1984.tb02066.x> PMID: 6479152
13. Fonseca GJ, Thillainadesan G, Yousef AF, Ablack JN, Mossman KL, Torchia J, et al. Adenovirus evasion of interferon-mediated innate immunity by direct antagonism of a cellular histone posttranslational modification. *Cell Host Microbe*. 2012; 11: 597–606. <https://doi.org/10.1016/j.chom.2012.05.005> PMID: 22704620
14. Zemke NR, Berk AJ. The Adenovirus E1A C Terminus Suppresses a Delayed Antiviral Response and Modulates RAS Signaling. *Cell Host Microbe*. 2017; 22: 789–800.e5. <https://doi.org/10.1016/j.chom.2017.11.008> PMID: 29241042
15. Pelka P, Ablack JNG, Fonseca GJ, Yousef AF, Mymryk JS. Intrinsic Structural Disorder in Adenovirus E1A: a Viral Molecular Hub Linking Multiple Diverse Processes. *J Virol*. 2008; 82: 7252–7263. <https://doi.org/10.1128/JVI.00104-08> PMID: 18385237
16. Blackford AN, Grand RJA. Adenovirus E1B 55-Kilodalton Protein: Multiple Roles in Viral Infection and Cell Transformation. *J Virol*. 2009; 83: 4000–4012. <https://doi.org/10.1128/JVI.02417-08> PMID: 19211739
17. Han J, Sabbatini P, Perez D, Rao L, Modha D, White E. The E1B 19K protein blocks apoptosis by interacting with and inhibiting the p53-inducible and death-promoting Bax protein. *Genes Dev*. 1996; 10: 461–477. <https://doi.org/10.1101/gad.10.4.461> PMID: 8600029
18. Yew PR, Berk AJ. Inhibition of p53 transactivation required for transformation by adenovirus early 1B protein. *Nature*. 1992; 357: 82–85. <https://doi.org/10.1038/357082a0> PMID: 1533443
19. Bridge E, Ketner G. Interaction of adenoviral E4 and E1b products in late gene expression. *Virology*. 1990; 174: 345–53. [https://doi.org/10.1016/0042-6822\(90\)90088-9](https://doi.org/10.1016/0042-6822(90)90088-9) PMID: 2137659
20. Cathomen T, Weitzman MD. A functional complex of adenovirus proteins E1B-55kDa and E4orf6 is necessary to modulate the expression level of p53 but not its transcriptional activity. *J Virol*. 2000; 74: 11407–12. <https://doi.org/10.1128/jvi.74.23.11407-11412.2000> PMID: 11070042
21. Harada JN, Shevchenko A, Shevchenko A, Pallas DC, Berk AJ. Analysis of the adenovirus E1B-55K-anchored proteome reveals its link to ubiquitination machinery. *J Virol*. 2002; 76: 9194–206. <https://doi.org/10.1128/jvi.76.18.9194-9206.2002> PMID: 12186903
22. Querido E, Blanchette P, Yan Q, Kamura T, Morrison M, Boivin D, et al. Degradation of p53 by adenovirus E4orf6 and E1B55K proteins occurs via a novel mechanism involving a Cullin-containing complex. *Genes Dev*. 2001; 15: 3104–17. <https://doi.org/10.1101/gad.926401> PMID: 11731475
23. Hidalgo P, Ip WH, Dobner T, Gonzalez RA. The biology of the adenovirus E1B 55K protein. *FEBS Letters*. 2019; 593: 3504–3517. <https://doi.org/10.1002/1873-3468.13694> PMID: 31769868
24. Winnacker EL. Adenovirus DNA: structure and function of a novel replicon. *Cell*. 1978; 14: 761–773. [https://doi.org/10.1016/0092-8674\(78\)90332-x](https://doi.org/10.1016/0092-8674(78)90332-x) PMID: 357012
25. Webster A, Leith IR, Nicholson J, Hounsell J, Hay RT. Role of preterminal protein processing in adenovirus replication. *J Virol*. 1997; 71: 6381–6389. <https://doi.org/10.1128/JVI.71.9.6381-6389.1997> PMID: 9261355
26. Brenkman AB, Breure EC, van der Vliet PC. Molecular architecture of adenovirus DNA polymerase and location of the protein primer. *J Virol*. 2002; 76: 8200–8207. <https://doi.org/10.1128/jvi.76.16.8200-8207.2002> PMID: 12134025
27. de Jong RN, van der Vliet PC, Brenkman AB. Adenovirus DNA replication: protein priming, jumping back and the role of the DNA binding protein DBP. *Curr Top Microbiol Immunol*. 2003; 272: 187–211. https://doi.org/10.1007/978-3-662-05597-7_7 PMID: 12747551
28. Robinson CM, Rajaiya J, Zhou X, Singh G, Dyer DW, Chodosh J. The E3 CR1-gamma gene in human adenoviruses associated with epidemic keratoconjunctivitis. *Virus Res*. 2011; 160: 120–127. <https://doi.org/10.1016/j.virusres.2011.05.022> PMID: 21683743
29. Singh G, Robinson CM, Dehghan S, Jones MS, Dyer DW, Seto D, et al. Homologous recombination in E3 genes of human adenovirus species D. *J Virol*. 2013; 87: 12481–12488. <https://doi.org/10.1128/JVI.01927-13> PMID: 24027303
30. Wold WSM, Tollefson AE, Hermiston TW. E3 Transcription Unit of Adenovirus. In: Doerfler W, Böhm P, editors. *The Molecular Repertoire of Adenoviruses I: Virion Structure and Infection*. Berlin, Heidelberg: Springer; 1995. pp. 237–274. https://doi.org/10.1007/978-3-642-79496-4_13 PMID: 7555057
31. Bridge E, Ketner G. Redundant control of adenovirus late gene expression by early region 4. *J Virol*. 1989; 63: 631–8. <https://doi.org/10.1128/JVI.63.2.631-638.1989> PMID: 2911117
32. Weitzman MD. Functions of the adenovirus E4 proteins and their impact on viral vectors. *Front Biosci*. 2005; 10: 1106–17. <https://doi.org/10.2741/1604> PMID: 15769610
33. Weitzman MD, Ornelles DA. Inactivating intracellular antiviral responses during adenovirus infection. *Oncogene*. 2005; 24: 7686–96. <https://doi.org/10.1038/sj.onc.1209063> PMID: 16299529

34. Weinmann R, Raskas HJ, Roeder RG. Role of DNA-dependent RNA polymerases II and III in transcription of the adenovirus genome late in productive infection. *Proc Natl Acad Sci USA*. 1974; 71: 3426–3439. <https://doi.org/10.1073/pnas.71.9.3426> PMID: 4530313
35. Vachon VK, Conn GL. Adenovirus VA RNA: An essential pro-viral non-coding RNA. *Virus Res*. 2016; 212: 39–52. <https://doi.org/10.1016/j.virusres.2015.06.018> PMID: 26116898
36. Hoeben RC, Uil TG. Adenovirus DNA Replication. *Cold Spring Harb Perspect Biol*. 2013;5. <https://doi.org/10.1101/cshperspect.a013003> PMID: 23388625
37. Charman M, Herrmann C, Weitzman MD. Viral and Cellular Interactions During Adenovirus DNA Replication. *FEBS Letters*. 2019 [cited 5 Dec 2019]. <https://doi.org/10.1002/1873-3468.13695> PMID: 31764999
38. Shaw AR, Ziff EB. Transcripts from the adenovirus-2 major late promoter yield a single early family of 3' coterminal mRNAs and five late families. *Cell*. 1980; 22: 905–916. [https://doi.org/10.1016/0092-8674\(80\)90568-1](https://doi.org/10.1016/0092-8674(80)90568-1) PMID: 7460018
39. Parks RJ. Adenovirus protein IX: a new look at an old protein. *Mol Ther*. 2005; 11: 19–25. <https://doi.org/10.1016/j.ymthe.2004.09.018> PMID: 15585402
40. Zhang W, Imperiale MJ. Requirement of the adenovirus IVa2 protein for virus assembly. *J Virol*. 2003; 77: 3586–3594. <https://doi.org/10.1128/jvi.77.6.3586-3594.2003> PMID: 12610134
41. Carter TH, Ginsberg HS. Viral transcription in KB cells infected by temperature-sensitive “early” mutants of adenovirus type 5. *J Virol*. 1976; 18: 156–166. <https://doi.org/10.1128/JVI.18.1.156-166.1976> PMID: 1255869
42. Crossland LD, Raskas HJ. Identification of adenovirus genes that require template replication for expression. *J Virol*. 1983; 46: 737–748. <https://doi.org/10.1128/JVI.46.3.737-748.1983> PMID: 6854739
43. Thomas GP, Mathews MB. DNA replication and the early to late transition in adenovirus infection. *Cell*. 1980; 22: 523–533. [https://doi.org/10.1016/0092-8674\(80\)90362-1](https://doi.org/10.1016/0092-8674(80)90362-1) PMID: 7448872
44. Ramke M, Lee JY, Dyer DW, Seto D, Rajaiya J, Chodosh J. The 5'UTR in human adenoviruses: leader diversity in late gene expression. *Sci Rep*. 2017; 7: 618. <https://doi.org/10.1038/s41598-017-00747-y> PMID: 28377580
45. Soloway PD, Shenk T. The adenovirus type 5 i-leader open reading frame functions in cis to reduce the half-life of L1 mRNAs. *J Virol*. 1990; 64: 551–558. <https://doi.org/10.1128/JVI.64.2.551-558.1990> PMID: 2296076
46. Morris SJ, Scott GE, Leppard KN. Adenovirus Late-Phase Infection Is Controlled by a Novel L4 Promoter. *J Virol*. 2010; 84: 7096–7104. <https://doi.org/10.1128/JVI.00107-10> PMID: 20444889
47. Biasiotto R, Akusjärvi G. Regulation of Human Adenovirus Alternative RNA Splicing by the Adenoviral L4-33K and L4-22K Proteins. *Int J Mol Sci*. 2015; 16: 2893–2912. <https://doi.org/10.3390/ijms16022893> PMID: 25636034
48. Tollefson AE, Ying B, Doronin K, Sidor PD, Wold WSM. Identification of a New Human Adenovirus Protein Encoded by a Novel Late I-Strand Transcription Unit. *J Virol*. 2007; 81: 12918–12926. <https://doi.org/10.1128/JVI.01531-07> PMID: 17881437
49. Ying B, Tollefson AE, Wold WSM. Identification of a Previously Unrecognized Promoter That Drives Expression of the UXP Transcription Unit in the Human Adenovirus Type 5 Genome. *J Virol*. 2010; 84: 11470–11478. <https://doi.org/10.1128/JVI.01338-10> PMID: 20739525
50. Miller DL, Myers CL, Rickards B, Collier HA, Flint SJ. Adenovirus type 5 exerts genome-wide control over cellular programs governing proliferation, quiescence, and survival. *Genome Biol*. 2007; 8: R58. <https://doi.org/10.1186/gb-2007-8-4-r58> PMID: 17430596
51. Miller DL, Rickards B, Mashiba M, Huang W, Flint SJ. The adenoviral E1B 55-kilodalton protein controls expression of immune response genes but not p53-dependent transcription. *J Virol*. 2009; 83: 3591–3603. <https://doi.org/10.1128/JVI.02269-08> PMID: 19211769
52. Zhao H, Dahlö M, Isaksson A, Syvänen A-C, Pettersson U. The transcriptome of the adenovirus infected cell. *Virology*. 2012; 424: 115–128. <https://doi.org/10.1016/j.virol.2011.12.006> PMID: 22236370
53. Zhao H, Chen M, Valdés A, Lind SB, Pettersson U. Transcriptomic and proteomic analyses reveal new insights into the regulation of immune pathways during adenovirus type 2 infection. *BMC Microbiol*. 2019; 19: 15. <https://doi.org/10.1186/s12866-018-1375-5> PMID: 30642258
54. Zhao H, Chen M, Pettersson U. A new look at adenovirus splicing. *Virology*. 2014; 456–457: 329–341. <https://doi.org/10.1016/j.virol.2014.04.006> PMID: 24889252
55. Crisostomo L, Soriano AM, Mendez M, Graves D, Pelka P. Temporal dynamics of adenovirus 5 gene expression in normal human cells. *PLOS ONE*. 2019; 14: e0211192. <https://doi.org/10.1371/journal.pone.0211192> PMID: 30677073

56. Brandes N, Linial M. Gene overlapping and size constraints in the viral world. *Biology Direct*. 2016; 11: 26. <https://doi.org/10.1186/s13062-016-0128-3> PMID: 27209091
57. O'Grady T, Wang X, Höner zu Bentrup K, Baddoo M, Concha M, Flemington EK. Global transcript structure resolution of high gene density genomes through multi-platform data integration. *Nucleic Acids Res*. 2016; 44: e145–e145. <https://doi.org/10.1093/nar/gkw629> PMID: 27407110
58. Depledge DP, Srinivas KP, Sadaoka T, Bready D, Mori Y, Placantonakis DG, et al. Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. *Nat Commun*. 2019; 10: 1–13. <https://doi.org/10.1038/s41467-019-08734-9> PMID: 30765700
59. Tombácz D, Balázs Z, Csabai Z, Moldován N, Szűcs A, Sharon D, et al. Characterization of the Dynamic Transcriptome of a Herpesvirus with Long-read Single Molecule Real-Time Sequencing. *Sci Rep*. 2017; 7: 43751. <https://doi.org/10.1038/srep43751> PMID: 28256586
60. Viehweger A, Krautwurst S, Lamkiewicz K, Madhugiri R, Ziebuhr J, Hölzer M, et al. Direct RNA nanopore sequencing of full-length coronavirus genomes provides novel insights into structural variants and enables modification analysis. *Genome Res*. 2019; 29: 1545–1554. <https://doi.org/10.1101/gr.247064.118> PMID: 31439691
61. Donovan-Banfield I, Turnell AS, Hiscox JA, Leppard KN, Matthews DA. Deep splicing plasticity of the human adenovirus type 5 transcriptome drives virus evolution. *Commun Biol*. 2020; 3: 1–14. <https://doi.org/10.1038/s42003-020-0849-9> PMID: 32170151
62. Westergren Jakobsson A, Segerman B, Wallerman O, Bergström Lind S, Zhao H, Rubin C-J, et al. The Human Adenovirus 2 Transcriptome: an Amazing Complexity of Alternatively Spliced mRNAs. Parrish CR, editor. *J Virol*. 2021; 95: e01869–20. <https://doi.org/10.1128/JVI.01869-20> PMID: 33239457
63. Weirather JL, de Cesare M, Wang Y, Piazza P, Sebastiano V, Wang X-J, et al. Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Res*. 2017; 6: 100. <https://doi.org/10.12688/f1000research.10571.2> PMID: 28868132
64. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011; 27: 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330> PMID: 21653522
65. Narasimhan V, Danecek P, Scally A, Xue Y, Tyler-Smith C, Durbin R. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics*. 2016; 32: 1749–1751. <https://doi.org/10.1093/bioinformatics/btw044> PMID: 26826718
66. Li X, Xiong X, Yi C. Epitranscriptome sequencing technologies: decoding RNA modifications. *Nat Methods*. 2017; 14: 23–31. <https://doi.org/10.1038/nmeth.4110> PMID: 28032622
67. Garalde DR, Snell EA, Jachimowicz D, Sipos B, Lloyd JH, Bruce M, et al. Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods*. 2018; 15: 201–206. <https://doi.org/10.1038/nmeth.4577> PMID: 29334379
68. Wu TD, Reeder J, Lawrence M, Becker G, Brauer MJ. GMAP and GSNAP for Genomic Sequence Alignment: Enhancements to Speed, Accuracy, and Functionality. *Methods Mol Biol*. 2016; 1418: 283–334. https://doi.org/10.1007/978-1-4939-3578-9_15 PMID: 27008021
69. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018; 34: 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191> PMID: 29750242
70. Bonfert T, Friedel CC. Prediction of Poly(A) Sites by Poly(A) Read Mapping. *PLoS ONE*. 2017; 12: e0170914. <https://doi.org/10.1371/journal.pone.0170914> PMID: 28135292
71. Borodulina OR, Kramerov DA. Transcripts synthesized by RNA polymerase III can be polyadenylated in an AAUAAA-dependent manner. *RNA*. 2008; 14: 1865–1873. <https://doi.org/10.1261/ra.1006608> PMID: 18658125
72. Tang AD, Soulette CM, Baren MJ van, Hart K, Hrabeta-Robinson E, Wu CJ, et al. Full-length transcript characterization of SF3B1 mutation in chronic lymphocytic leukemia reveals downregulation of retained introns. *bioRxiv*. 2018; 410183. <https://doi.org/10.1101/410183>
73. Hidalgo P, Anzués L, Hernández-Mendoza A, Guerrero A, Wood CD, Valdés M, et al. Morphological, Biochemical, and Functional Study of Viral Replication Compartments Isolated from Adenovirus-Infected Cells. *J Virol*. 2016; 90: 3411–3427. <https://doi.org/10.1128/JVI.00033-16> PMID: 26764008
74. Dybas JM, Lum KK, Kulej K, Reyes ED, Lauman R, Charman M, et al. Adenovirus Remodeling of the Host Proteome and Host Factors Associated with Viral Genomes. *mSystems*. 2021; e0046821. <https://doi.org/10.1128/mSystems.00468-21> PMID: 34463575
75. Stern-Ginossar N, Weisburd B, Michalski A, Le VTK, Hein MY, Huang S-X, et al. Decoding human cytomegalovirus. *Science*. 2012; 338: 1088–1093. <https://doi.org/10.1126/science.1227919> PMID: 23180859

76. Marton MJ, Baim SB, Ornelles DA, Shenk T. The adenovirus E4 17-kilodalton protein complexes with the cellular transcription factor E2F, altering its DNA-binding properties and stimulating E1A-independent accumulation of E2 mRNA. *J Virol.* 1990; 64: 2345–2359. <https://doi.org/10.1128/JVI.64.5.2345-2359.1990> PMID: 2139141
77. Boivin D, Morrison MR, Marcellus RC, Querido E, Branton PE. Analysis of synthesis, stability, phosphorylation, and interacting polypeptides of the 34-kilodalton product of open reading frame 6 of the early region 4 protein of human adenovirus type 5. *J Virol.* 1999; 73: 1245–1253. <https://doi.org/10.1128/JVI.73.2.1245-1253.1999> PMID: 9882328
78. Halbert DN, Cutt JR, Shenk T. Adenovirus early region 4 encodes functions required for efficient DNA replication, late gene expression, and host cell shutoff. *J Virol.* 1985; 56: 250–7. <https://doi.org/10.1128/JVI.56.1.250-257.1985> PMID: 4032537
79. Weinberg DH, Ketner G. A cell line that supports the growth of a defective early region 4 deletion mutant of human adenovirus type 2. *Proc Natl Acad Sci USA.* 1983; 80: 5383–5386. <https://doi.org/10.1073/pnas.80.17.5383> PMID: 6310575
80. Querido E, Marcellus RC, Lai A, Charbonneau R, Teodoro JG, Ketner G, et al. Regulation of p53 levels by the E1B 55-kilodalton protein and E4orf6 in adenovirus-infected cells. *J Virol.* 1997; 71: 3788–98. <https://doi.org/10.1128/JVI.71.5.3788-3798.1997> PMID: 9094654
81. Graham FL, Smiley J, Russell WC, Nairn R. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J Gen Virol.* 1977; 36: 59–74. <https://doi.org/10.1099/0022-1317-36-1-59> PMID: 886304
82. Ornelles DA, Shenk T. Localization of the adenovirus early region 1B 55-kilodalton protein during lytic infection: association with nuclear viral inclusions requires the early region 4 34-kilodalton protein. *J Virol.* 1991; 65: 424–429. <https://doi.org/10.1128/JVI.65.1.424-429.1991> PMID: 1824641
83. Gonzalez RA, Flint SJ. Effects of mutations in the adenoviral E1B 55-kilodalton protein coding sequence on viral late mRNA metabolism. *J Virol.* 2002; 76: 4507–4519. <https://doi.org/10.1128/jvi.76.9.4507-4519.2002> PMID: 11932416
84. Marshall LJ, Moore AC, Ohki M, Kitabayashi I, Patterson D, Ornelles DA. RUNX1 Permits E4orf6-Directed Nuclear Localization of the Adenovirus E1B-55K Protein and Associates with Centers of Viral DNA and RNA Synthesis. *J Virol.* 2008; 82: 6395–6408. <https://doi.org/10.1128/JVI.00043-08> PMID: 18417565
85. Babiss LE, Ginsberg HS, Darnell JE. Adenovirus E1B proteins are required for accumulation of late viral mRNA and for effects on cellular mRNA translation and transport. *Mol Cell Biol.* 1985; 5: 2552–8. <https://doi.org/10.1128/mcb.5.10.2552-2558.1985> PMID: 2942759
86. Pilder S, Moore M, Logan J, Shenk T. The adenovirus E1B-55K transforming polypeptide modulates transport or cytoplasmic stabilization of viral and host cell mRNAs. *Mol Cell Biol.* 1986; 6: 470–6. <https://doi.org/10.1128/mcb.6.2.470-476.1986> PMID: 2946932
87. Gonzalez R, Huang W, Finnen R, Bragg C, Flint SJ. Adenovirus E1B 55-kilodalton protein is required for both regulation of mRNA export and efficient entry into the late phase of infection in normal human fibroblasts. *J Virol.* 2006; 80: 964–974. <https://doi.org/10.1128/JVI.80.2.964-974.2006> PMID: 16378998
88. Ching W, Koyuncu E, Singh S, Arbelo-Roman C, Hartl B, Kremmer E, et al. A Ubiquitin-specific Protease Possesses a Decisive Role for Adenovirus Replication and Oncogene-mediated Transformation. *PLoS Pathog.* 2013; 9: e1003273. <https://doi.org/10.1371/journal.ppat.1003273> PMID: 23555268
89. Goodrum FD, Shenk T, Ornelles DA. Adenovirus early region 4 34-kilodalton protein directs the nuclear localization of the early region 1B 55-kilodalton protein in primate cells. *J Virol.* 1996; 70: 6323–6335. <https://doi.org/10.1128/JVI.70.9.6323-6335.1996> PMID: 8709260
90. Querido E, Chu-Pham-Dang H, Branton PE. Identification and elimination of an aberrant splice product from cDNAs encoding the human adenovirus type 5 E4orf6 protein. *Virology.* 2000; 275: 263–266. <https://doi.org/10.1006/viro.2000.0516> PMID: 10998326
91. O'Connor RJ, Hearing P. The E4-6/7 Protein Functionally Compensates for the Loss of E1A Expression in Adenovirus Infection. *J Virol.* 2000; 74: 5819–5824. <https://doi.org/10.1128/jvi.74.13.5819-5824.2000> PMID: 10846061
92. Pied N, Wodrich H. Imaging the adenovirus infection cycle. *FEBS Letters.* 2019; 593: 3419–3448. <https://doi.org/10.1002/1873-3468.13690> PMID: 31758703
93. Pombo A, Ferreira J, Bridge E, Carmo-Fonseca M. Adenovirus replication and transcription sites are spatially separated in the nucleus of infected cells. *EMBO J.* 1994; 13: 5075–85. <https://doi.org/10.1002/j.1460-2075.1994.tb06837.x> PMID: 7957073
94. Komatsu T, Robinson DR, Hisaoka M, Ueshima S, Okuwaki M, Nagata K, et al. Tracking adenovirus genomes identifies morphologically distinct late DNA replication compartments. *Traffic.* 2016; 17: 1168–1180. <https://doi.org/10.1111/tra.12429> PMID: 27492875

95. Komatsu T, Quentin-Froignant C, Carlon-Andres I, Lagadec F, Rayne F, Ragues J, et al. In Vivo Labeling of Adenovirus DNA Identifies Chromatin Anchoring and Biphasic Genome Replication. *J Virol*. 2018; 92: e00795–18. <https://doi.org/10.1128/JVI.00795-18> PMID: 29997215
96. Genoveso MJ, Hisaoka M, Komatsu T, Wodrich H, Nagata K, Okuwaki M. Formation of adenovirus DNA replication compartments and viral DNA accumulation sites by host chromatin regulatory proteins including NPM1. *FEBS J*. 2020; 287: 205–217. <https://doi.org/10.1111/febs.15027> PMID: 31365788
97. Arvey A, Tempera I, Tsai K, Chen H-S, Tikhmyanova N, Klichinsky M, et al. An Atlas of the Epstein-Barr Virus Transcriptome and Epigenome Reveals Host-Virus Regulatory Interactions. *Cell Host Microbe*. 2012; 12: 233–245. <https://doi.org/10.1016/j.chom.2012.06.008> PMID: 22901543
98. Arias C, Weisburd B, Stern-Ginossar N, Mercier A, Madrid AS, Bellare P, et al. KSHV 2.0: A Comprehensive Annotation of the Kaposi's Sarcoma-Associated Herpesvirus Genome Using Next-Generation Sequencing Reveals Novel Genomic and Functional Features. *PLoS Pathog*. 2014; 10: e1003847. <https://doi.org/10.1371/journal.ppat.1003847> PMID: 24453964
99. Rutkowski AJ, Erhard F, L'Hernault A, Bonfert T, Schilhabel M, Crump C, et al. Widespread disruption of host transcription termination in HSV-1 infection. *Nat Commun*. 2015; 6: 7126. <https://doi.org/10.1038/ncomms8126> PMID: 25989971
100. Garren SB, Kondaveeti Y, Duff MO, Carmichael GG. Global Analysis of Mouse Polyomavirus Infection Reveals Dynamic Regulation of Viral and Host Gene Expression and Promiscuous Viral RNA Editing. *PLoS Pathog*. 2015; 11: e1005166. <https://doi.org/10.1371/journal.ppat.1005166> PMID: 26407100
101. Evans VC, Barker G, Heesom KJ, Fan J, Bessant C, Matthews DA. De novo derivation of proteomes from transcriptomes for transcript and protein identification. *Nat Methods*. 2012; 9: 1207–1211. <https://doi.org/10.1038/nmeth.2227> PMID: 23142869
102. Hennig T, Michalski M, Rutkowski AJ, Djakovic L, Whisnant AW, Friedl MS, et al. HSV-1-induced disruption of transcription termination resembles a cellular stress response but selectively increases chromatin accessibility downstream of genes. *PLoS Pathog*. 2018; 14: e1006954. <https://doi.org/10.1371/journal.ppat.1006954> PMID: 29579120
103. Gehring NH, Roignant J-Y. Anything but Ordinary—Emerging Splicing Mechanisms in Eukaryotic Gene Regulation. *Trends Genet*. 2021; 37: 355–372. <https://doi.org/10.1016/j.tig.2020.10.008> PMID: 33203572
104. Condezo GN, Martín CS. Localization of adenovirus morphogenesis players, together with visualization of assembly intermediates and failed products, favor a model where assembly and packaging occur concurrently at the periphery of the replication center. *PLoS Pathog*. 2017; 13: e1006320. <https://doi.org/10.1371/journal.ppat.1006320> PMID: 28448571
105. Depledge DP, Mohr I, Wilson AC. Going the Distance: Optimizing RNA-Seq Strategies for Transcriptomic Analysis of Complex Viral Genomes. *J Virol*. 2019;93. <https://doi.org/10.1128/JVI.01342-18> PMID: 30305358
106. Huang MM, Hearing P. Adenovirus early region 4 encodes two gene products with redundant effects in lytic infection. *J Virol*. 1989; 63: 2605–2615. <https://doi.org/10.1128/JVI.63.6.2605-2615.1989> PMID: 2724411
107. Chartier C, Degryse E, Gantzer M, Dieterle A, Pavirani A, Mehtali M. Efficient generation of recombinant adenovirus vectors by homologous recombination in *Escherichia coli*. *J Virol*. 1996; 70: 4805–4810. <https://doi.org/10.1128/JVI.70.7.4805-4810.1996> PMID: 8676512
108. Evans JD, Hearing P. Distinct roles of the Adenovirus E4 ORF3 protein in viral DNA replication and inhibition of genome concatenation. *J Virol*. 2003; 77: 5295–5304. <https://doi.org/10.1128/jvi.77.9.5295-5304.2003> PMID: 12692231
109. Anderson DC, Li W, Payan DG, Noble WS. A New Algorithm for the Evaluation of Shotgun Peptide Sequencing in Proteomics: Support Vector Machine Classification of Peptide MS/MS Spectra and SEQUEST Scores. *J Proteome Res*. 2003; 2: 137–146. <https://doi.org/10.1021/pr0255654> PMID: 12716127
110. Bern M, Kil YJ, Becker C. Byonic: Advanced Peptide and Protein Identification Software. *Curr Protoc Bioinformatics*. 2012; 40: 13.20.1–13.20.14. <https://doi.org/10.1002/0471250953.bi1320s40> PMID: 23255153
111. Brademan DR, Riley NM, Kwiecien NW, Coon JJ. Interactive Peptide Spectral Annotator: A Versatile Web-based Tool for Proteomic Applications. *Mol Cell Proteomics*. 2019; 18: S193–S201. <https://doi.org/10.1074/mcp.TIR118.001209> PMID: 31088857
112. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. 2016; 44: W160–165. <https://doi.org/10.1093/nar/gkw257> PMID: 27079975

113. Feng Y-Y, Ramu A, Cotto KC, Skidmore ZL, Kunisaki J, Conrad DF, et al. RegTools: Integrated analysis of genomic and transcriptomic data for discovery of splicing variants in cancer. *bioRxiv*. 2018; 436634. <https://doi.org/10.1101/436634>
114. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative Genomics Viewer. *Nat Biotechnol*. 2011; 29: 24–26. <https://doi.org/10.1038/nbt.1754> PMID: 21221095
115. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25: 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352> PMID: 19505943
116. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010; 26: 841–842. <https://doi.org/10.1093/bioinformatics/btq033> PMID: 20110278
117. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010; 38: 576–589. <https://doi.org/10.1016/j.molcel.2010.05.004> PMID: 20513432