Human adenovirus (HAdV) infections cause disease world-wide. Whole genome sequencing has now distinguished 90 distinct genotypes in 7 species (A-G). Over half of these 90 HAdVs fall within species D, with essentially all of the HAdV-D whole genome sequences generated in the last decade. Herein, we describe recent new findings made possible by mining of this expanded genome database, and propose future directions to elucidate new functional elements and new functions for previously known viral components.

Keywords: adenovirus, genome, evolution, transcription factor, interactome

INTRODUCTION

Human adenovirus (HAdV) infections represent a significant source of morbidity and mortality, world-wide and at all ages, through highly transmittable infections at mucosal sites, including the eye, and urinary, respiratory, and gastrointestinal tracts (Horwitz, 1996). HAdV causes fatal acute respiratory distress syndrome in healthy adults and is especially lethal in infants and the immune compromised (Banthumkosol, 1998; Ryu et al., 2003; Wallot et al., 2006; Engelmann et al., 2016; Tan et al., 2016; Zhang et al., 2016). No FDA-approved therapy for acute HAdV infection is available. At resolution of acute infection, persistence may develop within nasopharyngeal lymphoid tissue (Neumann et al., 1987; Garnett et al., 2002, 2009; Zhang et al., 2010; Assadian et al., 2016), as yet uncharacterized cells in the gastrointestinal tract (Roy et al., 2009), and possibly the ocular surface (Kaye et al., 2005), permitting evolution of new HAdVs through homologous recombination between two or more HAdVs infecting the same cell(s) (Lee et al., 2005, 2018; Echavarria et al., 2006; McCarthy et al., 2009; Seto et al., 2010).

HAdVs are divided phylogenetically into seven species (A-G), with a total of 90 recognized genotypes with whole genome sequences in GenBank, including the original 51 “serotypes”—determined by serum neutralization—which now all have been fully sequenced (Table 1) (Robinson et al., 2013a). Human adenovirus species D (HAdV-D) is the largest and most rapidly growing among all HAdV species, and contains viruses associated with epidemic keratoconjunctivitis (EKC), a severe, hyperacute ocular surface infection (Butt and Chodosh, 2006). A collaboration funded by the American Recovery and Reinvestment Act of 2009 came to fruition with the complete whole genome sequencing and analysis of all previously unsequenced HAdV-D serotypes (Robinson et al., 2013a), leading to a new understanding of adenovirus ontogeny (Jones et al., 2007; Robinson et al., 2008, 2009a,b; Robinson et al., 2011b,c; Robinson et al., 2013a,b; Walsh et al., 2009, 2010a,b; Arnold et al., 2010; Torres et al., 2010; Dehghan et al., 2011, 2013a,b; Walsh et al., 2011; Liu et al., 2011; Seto et al., 2011, 2013; Singh et al., 2012, 2013; Zhou et al., 2012)—including those...
HAdV-Ds associated with EKC (Robinson et al., 2008, 2009b, 2011b; Walsh et al., 2009; Zhou et al., 2012)—and ultimately to a new typing system for HAdV based on genomics (Seto et al., 2011).

Recent published work demonstrates how genome “mining,” in-depth analyses of the growing HAdV genome database, can bring about new realizations and add critical new information to prior ones. The trimeric fiber protein on adenoviruses mediates viral entry through interaction of the distal most “knob” structure on the fiber with host cell receptors. In a phylogenetic analysis of HAdV-D fiber genes, HAdV-D types associated with EKC were recently shown to form a unique clade (Ismail et al., 2016). By proteotyping, a new in silico methodology described in detail below, EKC virus-associated fiber knobs were uniquely shared, and signature amino acid positions distinguished EKC from non-EKC types. Remarkably, human corneal epithelial cell tropism could be predicted by the presence of a lysine or alanine at residue 240, and this amino acid residue in EKC viruses showed evidence for positive selection. These data added to the prior observation by Huang and coworkers that artificial mutation to a lysine at residue 240 in a non-EKC virus could confer infection of Chang cells, a conjunctiva derived continuous cell line (Huang et al., 1999). However, because Chang cells came later known to be contaminated by HeLa cells, the importance of residue 240 to ocular tropism was until this new observation, in some doubt.

Another recently published effort provided further evidence of the importance and potential for HAdV genome mining. Late adenoviral gene expression is initiated by the adenovirus potential new leader sequence embedded within the E3 region, for eukaryotic initiation factor 4F or cap binding protein complex (Logan and Shenk, 1984; Zhang et al., 1989; Logan and Shenk, 1984). The HAdV tripartite leader is a 200-nucleotide 5’ noncoding region that circumvents the requirement for eukaryotic initiation factor 4F or cap binding protein complex (Ziff and Evans, 1978; Akusjärvi and Pettersson, 1979; Dolph et al., 1988; Zhang et al., 1989), and permits translation of HAdV mRNAs at late times in infection when cap-dependent translation is blocked due to shut down of host cellular cap-dependent mRNA translation. HAdV 5’ untranslated regions (5’UTRs) are critical for cap-independent initiation, and impact mRNA localization and stability. The HAdV tripartite leader (TPL), composed of three introns (TPL 1-3), drives translation of HAdV late mRNA. The annotation of 72 HAdV genotypes for the HAdV TPL and another previously described leader, the i-leader, let to identification of newly identified polycistronic mRNAs for RID-α and RID-β within the E3 transcription unit, and a potential new open reading frame (ORF) within the i-leader sequence, with termination of this potential protein in TPL3 (Ramke et al., 2017). In addition, the authors also identified a potential new leader sequence embedded within the E3 region, tentatively named the j-leader (Figure 1).

**STRUCTURE AND INFECTION**

The HAdV is non-enveloped, icosahedral in shape, and contains a double stranded DNA genome of ∼36,000 base pairs (bp).
Ismail et al. Adenoviromics

FIGURE 1 | Putative "j"-leader located within the CR1-α E3 gene. (A) Schematic for the location of a newly detected leader ("j"-leader) embedded within the E3 CR1-α gene, experimentally determined to be spliced to some, but not all mRNAs of the E3 genes. (B) Gel photomicrograph of mRNA transcripts amplified with forward primer from TPL1 and reverse primers from CR1-γ, CR1-β, and RID-α. Primers were chosen to elicit similarly sized bands to facilitate subsequent sequencing. (C) Nucleotide sequence of the PCR product for CR1-β. The putative j-leader sequence and splice sites are shown in yellow and green, respectively. Note an additional 4 nucleotide 5′ UTR (AACC) prior to the CR1-β start site (red). The 5′ UTR in (C) prior to the splice site for the j-leader is from TPL3. Adapted from Ramke et al. (2017) with permission.

with ~1 open reading frame (ORF) for every 1000 nucleotides. Viral DNA is associated with four (interior) core proteins including Mu, VII, V, and terminal protein. The histone-like protein (p) VII protects viral DNA from cellular DNA damage responses (Lischwe and Sung, 1977; Karen and Hearing, 2011; Avgousti et al., 2017). The outer protein coat (capsid) of the virus consists of 240 hexon capsomers and 12 penton capsomers, along with several minor capsid proteins. The latter include pVI, pIIIa, pVIII, and pIX and are important to capsid stability. Each penton capsomer contains a ring of five penton base proteins which bind and support the trimeric fiber protein with its distal fiber knob. During viral infection, the fiber knob binds to one of several host cell receptors (Nemerow, 2000; Goosney and Nemerow, 2003; Nemerow et al., 2009). The penton base protein contains two hypervariable loops. The interaction between fiber knob and a host cell receptor brings about secondary contact between the hypervariable loop 2 (HVL2) arginine-glycine-aspartic acid (RGD) motif in each penton base protein (five per penton base capsomer) with host cell integrins αvβ3, αvβ5, and αvβ1, that in turn induce endocytosis of the virus (Li et al., 1998a,b; Li et al., 2000). HAdV structural proteins can serve multiple functions. For example, the minor capsid structural protein VI (pVI) plays a critical role in at least three distinct aspects of the viral “life” cycle: endosomal escape during cell entry, nuclear assembly during viral replication, and stability of the intact, infectious virus outside the host (Wodrich et al., 2003; Wietthoff et al., 2005; Moyer et al., 2011, 2016). These findings suggest that, as with pVI, other HAdV structural proteins may have multiple functions yet to be elucidated.

GENOMICS AND EVOLUTION

The relatively large genome database for HAdV-D (over 50 unique viruses with available whole genome sequences) (Tables 1, 2) has permitted detailed analyses of genome relationships within this clinically important adenovirus species. HAdV-D genomes are highly conserved (>90%). However, whole genome analyses of HAdV-D have revealed specific loci of genetic hypervariability in the hexon, penton base, fiber, and E3 CR1α, β, and γ genes (Figure 2), dictating nonsynonymous amino acid changes in corresponding proteins (Figure 3). GC content confers genome stability and resistance to recombination (Gruss et al., 1991), and the genomes of HAdV-D have among the highest GC content among HAdV species (~56%). The hypervariable regions in HAdV-D were found to be sharply reduced in GC nucleotide content relative to the rest of the genome (Robinson et al., 2013a). Mutations in HAdV are relatively infrequent, with genome stability now documented in some types across decades (Hofmayer et al., 2009; Mahadevan et al., 2010; Seto et al., 2010; Dehghan et al., 2013b; Robinson et al., 2013a; Alkhalaf et al., 2015). However, those regions of the genome shown to be hypervariable and relatively low in GC content are the very same also shown to undergo homologous recombination (Robinson et al., 2009a, 2011b; Walsh et al., 2009; Zhou et al., 2012; Singh et al., 2013), driving the evolution of new genotypes.

Adenoviruses recombine specifically during viral replication (Williams et al., 1975; Meinschad and Winnacker, 1980; Munz et al., 1983), and do so by both homologous and heterologous mechanisms (Young et al., 1984; Epstein and Young, 1991; Crawford-Miksza and Schnurr, 1996). However, the evidence
TABLE 2 | Species and molecular types of human adenovirus (HAdV) genotypes 52–90.

| HAdV type | #Name | GenBank accession no. | Genome length | Year published |
|-----------|-------|----------------------|---------------|----------------|
| HAdV-G52  | P52H52F52/2003/USA | DQ923122.2 | 34250 | 2007 |
| HAdV-D53  | P37H22F8/2005/DEU | FJ196625 | 34909 | 2009 |
| HAdV-D54  | P54H54F8/2000/JPN | AB333801 | 34920 | 2009 |
| HAdV-B55  | P14H11F14/2006/CHN | FJ643676 | 34755 | 2009 |
| HAdV-D56  | P56H15F9/2008/FRA | HM770721 | 35066 | 2010 |
| HAdV-C57  | P1H57F6/2001/RUS | HQ003817 | 35818 | 2011 |
| HAdV-D58  | P58H68F29/1996/ARG | HQ883276 | 35217 | 2011 |
| HAdV-D59  | P64H25F6/2007/USA | JF799911 | 35072 | 2012 |
| HAdV-D60  | P60H20F60/2009/CAN | HQ007053 | 35050 | 2013 |
| HAdV-A61  | P31H31F31/2004/JPN | JF964962 | 33776 | 2011 |
| HAdV-D62  | P62H62F62/1993/GBR | JN162671 | 35127 | 2014 |
| HAdV-D63  | P30H30F29/1959/JPN | JN837866 | 35188 | 2012 |
| HAdV-D64  | P22H19F37/1953/USA | EF121005 | 35231 | 2006 |
| HAdV-D65  | P58H10F9/2004/BGD | AP012285 | 35172 | 2012 |
| HAdV-B66  | P6H7F3/1987/ARG | JN960678 | 35080 | 2012 |
| HAdV-D67  | P67H9F67/2005/BGD | AP012302 | 35075 | 2013 |
| HAdV-B68  | P1H63F16/2004/ARG | JN960678 | 35538 | 2011 |
| HAdV-D69  | P53H15F69/1955/SAU | JN226748 | 35124 | 2013 |
| HAdV-D70  | P7H70F29/2014/DEU | KF641339 | 35186 | 2015 |
| HAdV-D71  | P9H20F71/1987/DEU | KF642027 | 35192 | 2013 |
| HAdV-D72  | P72H63F72/1995/DEU | KF68335 | 34553 | 2013 |
| HAdV-D73  | P6H7H5/2015/DEU | KF68335 | 34553 | 2013 |
| HAdV-D74  | P70H7F4F1/2015/DEU | KF68335 | 34553 | 2013 |
| HAdV-D75  | P75H2F6/2015/DEU | KF68335 | 34553 | 2013 |
| HAdV-B76  | P2H31F16/DEU | KF633445 | 35538 | 2013 |
| HAdV-B77  | P3H34F7/1985/DEU | KF633445 | 35538 | 2013 |
| HAdV-B78  | P3H31F11/2015/DEU | KF633445 | 35538 | 2013 |
| HAdV-D79  | P1H34F11/2015/JPN | LC177352 | 34779 | 2016 |
| HAdV-D80  | P19,29H28F22/2014/DEU | TBA | 34909 | unpublished |
| HAdV-D81  | P65H49F60/2012/JPN | AB765926.1 | 35198 | 2014 |
| HAdV-D82  | P56H15F37/2011/JPN | LC066535.1 | 35122 | 2015 |
| HAdV-D83  | P5H9F15/2010/PAR | KX927426.1 | 35207 | 2016 |
| HAdV-D84  | P3H17F4/2011/PAN | MF416150 | 35257 | 2017 |
| HAdV-D85  | P3H17F8/2015/JPN | LC314153 | 35203 | 2018 |
| HAdV-B86  | P9H25F25/1978/SWE | TBA | 35147 | unpublished |
| HAdV-D87  | P9H15F25/1967/USA | MF476841 | 35159 | unpublished |
| HAdV-D88  | P8H15F9/1963/USA | MF476841 | 35115 | unpublished |
| HAdV-D89  | P9H2F2/2015/DEU | TBA | 35998 | unpublished |
| HAdV-D90  | P3H7F67/2017/BGD | TBA | 34207* | unpublished |

*Name indicates molecular type (P, penton base; H, hexon; F, fiber)/year of isolation/country of isolation. TBA: GenBank accession number, to be assigned. Metagenomics project, missing the inverted terminal repeat sequences.

for homologous recombination as the major mechanism driving HAdV-D evolution is unassailable (Robinson et al., 2013a; Singh et al., 2013). Specifically, recombination occurs in the two penton base hypervariable regions (these code for two hypervariable loops (HVLs) on the penton base protein, separated from one another by ~125 conserved amino acids), seven hexon hypervariable regions (these are closely adjacent in the hexon gene and determine two adjacent HVLs on the hexon protein), fiber (fiber gene and protein are entirely hypervariable), and E3 CR1α, β, and γ (each also entirely hypervariable). For homologous recombination between two HAdVs to occur, at least two virus types with high nucleotide sequence homology at corresponding locations in both genomes must co-infect the same cell, and viral DNA replication should be ongoing. Co-infection by two or more HAdVs has been well documented (Lee et al., 2005; Echavarria et al., 2006; Vora et al., 2006; McCarthy...
et al., 2009; Halstead et al., 2010; Seto et al., 2010), as has the presence of two HAdV types in archived clinical samples (Singh et al., 2012).

“Proteotyping” is a novel approach to the study of genome evolution (Obenauer et al., 2006), and has been applied to characterize recombination among HAdV-D (Robinson et al., 2013a; Singh et al., 2013). In this method, maximum likelihood trees are used to align amino acid sequences of hypervariable, frequently recombined proteins. Each amino acid is assigned a unique, arbitrary color. Consensus residues are colored white, and gaps in the alignment are colored black. A threshold of <10% sequence divergence is used to distinguish unique proteotypes. An example of proteotyping is shown (Figure 4), comparing an amino acid alignment from E3 14.7K, a highly conserved gene with one distinct proteotype, with the hypervariable E3 CR1α (Singh et al., 2013), with six distinct proteotypes observed among 38 HAdV-Ds. E3 14.7K is therefore not hypervariable and not recombinant. E3 CR1α is hypervariable and recombinant.

Another way to interpret the analyses for those proteins like E3 CR1α, with more than one proteotype is that those proteotypes containing more than one HAdV type have previously recombined in nature, while those proteotypes with
only one HAdV type are those that have not (yet) been shown to recombine in nature. HAdV-D37 and 29 fall within different hexon proteotypes (Figure 5). HAdV-D37 shares a hexon proteotype with HAdV-D13 and 30 (Robinson et al., 2013a), while HAdV-D29 shares a hexon proteotype with HAdV-D15, 56, and 69 (Singh et al., 2015). These two hexon proteotypes therefore have undergone prior homologous recombination. In contrast, the hexon proteins of HAdV-D10 and 28 are each in a proteotype with only one member; hexon recombination for these two viruses has therefore not yet been documented in nature. In sum, these data show by independent means that homologous recombination within HAdV-D is common, and confirm previously recognized patterns of homologous recombination among HAdV-D (Robinson et al., 2009a, 2011b, 2013a,b; Walsh et al., 2009, 2010a; Singh et al., 2012, 2013; Zhou et al., 2012; Gonzalez et al., 2014).

The local sequence and/or structure of DNA in regions flanking recombinogenic sites is significant for directing cellular recombination machinery to those regions. In bacteria, a signal for recombination between homologous DNA is the crossover hotspot instigator, or Chi nucleotide sequence. This was first discovered in bacteriophage lambda, then in bacterial DNA, and later shown to mediate recombination between them (Stahl, 1998). The Chi sequence in E. coli (ChiEC) is 5′-GCTGGTGG-3′ (Smith et al., 1981; Bianco and Kowalczykowski, 1997), and its presence induces the conversion of the RecBCD enzyme from a helicase to an exonuclease, producing ssDNA that can invade homologous dsDNA during recombination (Taylor et al., 1985).

**FIGURE 4** | Proteotyping analysis comparing the HAdV-D E3 14.7K (A) and CR1α (B) proteins. The 14.7K protein was conserved, while CR1α demonstrated 6 unique proteotypes. Maximum likelihood phylogenetic trees are shown to the left for each putative protein, and amino acid signatures to the right. The scale bar at the bottom left of each sub-figure denotes the phylogenetic distance reflected in horizontal dimension of the respective tree. To construct the amino acid signatures shown, each amino acid was assigned a unique color (upper right corner), consensus amino acids at each position across all 38 viruses were assigned white, and gaps in the alignment were colored black. Horizontal red lines delineate distinct proteotypes. Adapted from Singh et al. (2013) with permission.
The RecA enzyme of *E. coli* is loaded onto unwound ssDNA by RecBCD and promotes ssDNA exchange/recombination with homologous dsDNA (Cox, 1999; Smith, 2012). RecA has significant homology to eukaryotic Rad51 and its paralogs (Suwaki et al., 2011), enzymes that repair dsDNA breaks in human cells, facilitate homologous recombination, and during adenovirus infection, bind to the E2 DNA binding protein (Tookman et al., 2016). In our study of the region just 5′ to HVL2 on the penton base gene, a recombination hot-spot for HAdV-D (Robinson et al., 2009a), we found Chi-like sequences (ChiAD), e.g., 5′-ACTTCTGA-3′ in the proteotype containing HAdV-D64, and 5′-TCTCCTGA-3′ in the proteotype including HAdV-D37 (Lee et al., 2018). The putative ChiAD sequences we identified in HAdV-D were found within the GC-rich component of GC/AT transition zones that precede and include HVL2, and were conserved within each proteotype. In *vitro*, *E. coli* lysates containing RecA protein increased recombination of two HAdV-D genotypes with the same penton base HVL2 proteotype. RecA was shown by ChIP to bind specifically to ChiAD nucleotide sequence in the same regions, and also colocalize with adenovirus DNA within infected cell nuclei. These data suggest that Chi-like nucleotide sequences adjacent to the junction of conserved and hypervariable gene segments in HAdV-D may be an important signal for homologous recombination, and provide evidence in support of the idea that local bacterial flora might enhance natural recombination through Chi-like nucleotide sequences at HAdV-D recombination hotspots.

Another explanation for homologous recombination between HAdV, not exclusive of a role for ChiAD, is the potential for GC-low (AT-rich) single stranded DNA (ssDNA) to form hairpin loops (Nagy and Bujarski, 1997; Ohshima et al., 2007), a physical nonlinearity that would facilitate binding of ssDNA of one HAdV-D type to a homologous segment of ss or dsDNA from a physically adjacent but different HAdV-D type during coinfection of the same cell. Hairpin loops and other alterations in the physical configuration of ssDNA during DNA replication might also contribute to polymerase jumping (Jennings et al., 1983; Spaan et al., 1983; Pääbo et al., 1990; Viswanathan et al., 1999), in which physical constraints to polymerization lead to translocation of the DNA polymerase to an adjacent DNA from a different virus, resulting in a recombined DNA. Polymerase jumping has been shown to occur during HAdV DNA replication (King et al., 1997; de Jong et al., 2003), although it has not been suggested previously as a mechanism for HAdV-D evolution. Analysis of 38 HAdV-D whole genome sequences identified instances of 15 nucleotide-long GC-rich sequence adjacent to 15 nucleotide-long AT-rich sequence (sometimes with a 15 or 30 nucleotide-long GC-moderate sequence intervening), located...
just 5′ and 3′ to frequently recombined gene segments, and which were shown by in silico analysis of their corresponding ssDNA to form hairpin loops (Robinson et al., 2013a). Taken together, these data suggest covariant effects of nucleotide sequence and ssDNA secondary structures on homologous recombination between two HAdV-Ds.

TRANSCRIPTOME

Regions of the HAdV-D genome currently thought to be “noncoding,” may contain functional elements. Because viruses exist on the nano-scale, viral genomes are by necessity constrained by size, and “junk” nucleotide sequences represent an extravagance. The National Human Genome Research Institute project to identify functional elements in the human genome (Encyclopedia of DNA Elements, or ENCODE) identified functionality in much of the human genome previously without known utility (Consortium et al., 2007; Qu and Fang, 2013; Kellis et al., 2014). The double-stranded DNA genomes of HAdV also contain regions with no known function. Transcriptional profiling of host gene expression has been studied after HAdV infection (Dorer et al., 2011). However, although viral transcriptomes have been reported for several viruses, most notably dengue, varicella zoster, and Epstein-Barr viruses (Ortmann et al., 2008; Ertl et al., 2011; Nagel et al., 2011, 2013; Arvey et al., 2013; Sujayanont et al., 2014), a de novo HAdV transcriptome has not been reported. Wu and coworkers used deep RNA sequencing to confirm known bat AdV transcripts (Wu et al., 2013), but did not investigate “noncoding” regions. In silico ORF prediction in HAdV can be difficult due to splice variants and inconsistencies in banked gene annotations (Davison et al., 2003), but in a prior annotation of HAdV-D37, ~60 new additional ORFs were predicted using in combination, the NCBI ORF finder, TIGR annotation engine, and GeneMark Heuristic model (Robinson et al., 2008). Putative genes were found within the large regions of noncoding DNA on the complementary strand opposite to established HAdV genes (Figure 6), in smaller regions on the coding strand within established transcription units but between confirmed genes, and overlapping or completely within established genes. Work is in progress in our laboratories to identify putative new genomic elements in HAdV by high-throughput sequencing of the viral transcriptome of HAdV-D37.

TRANSCRIPTION FACTOR BINDING SITES

HAdV uses host TFs nuclear factor I and III (NF-I and NF-III) as part of the viral DNA replication complex (Pruijn et al., 1988; Mul et al., 1990; Hatfield and Hearing, 1991, 1993). Simian AdVs typically lack the NF-I binding site, while human viruses express it. It was previously reported that HAdV-E4, originally isolated in 1952, is a product of recombination between HAdV-B16 and the simian AdV, SAdV-E26. Clinical strains of HAdV-E4 isolated recently contain a NF-I binding site in the inverted terminal repeat (Houng et al., 2006; Dehghan et al., 2013a,b) that is absent in the original isolate (Purkayastha et al., 2005), suggesting that NF-I binding may be important to viral fitness in humans. To further elucidate mechanisms of viral gene expression, we are exploring novel TF binding sites on HAdV-D DNA, using ENCODE validated methodologies (Gerstein et al., 2012; Landt et al., 2012).
PROTEIN INTERACTOMES

The major HAdV capsid structural proteins—hexon, penton base, and fiber—interact directly with extracellular mediators of host immunity (Gahéry-Ségard et al., 1998; Molinier-Frenkel et al., 2002; Schoggins and Falck-Pedersen, 2006; Tamanini et al., 2006; Kalyuzhniy et al., 2008; Chintakuntlavar et al., 2010; Braddock et al., 2012; Flatt et al., 2013). The hexon, penton base, and fiber proteins also exhibit distinct amino acid signatures, characterizing discrete proteotypes (Robinson et al., 2013a). Gene products from the E3 transcription unit of HAdV function in viral immune evasion (Horwitz, 2004; Lichtenstein et al., 2004b; Windheim et al., 2004). In HAdV-D, the open reading frames for three of eight E3 genes—CR1α, CR1β, and CR1γ—are uniquely hypervariable compared to the other ORFs within the E3 transcription unit, and also segregate into discrete proteotypes (Singh et al., 2013). Highly conserved genes, such as DNA binding protein, DNA polymerase, and E3 14.7K, show no such variability (Robinson et al., 2013a; Singh et al., 2013).

While it may be assumed that hypervariability in major capsid and E3 proteins is driven through evolutionary selection by the extracellular interactome, amino acid differences in a hypervariable protein can also lead to differences in that protein’s intracellular interactome, the set of intrinsic host cell proteins which network with the viral protein, as was recently confirmed for E3 CR1 genes across HAdV species (Martinez-Martin et al., 2016). Viral capsid structural proteins are critical to virion stability. For the nonenveloped HAdV, fiber and penton base proteins on the external surface of the capsid serve as ligands for attachment to the host cell (Huang et al., 1999) and initiate viral entry (Wickham et al., 1993), respectively. HAdVs are typically internalized via endosomes. Endosomal acidification leads to structural instability of the capsid and endosomal release into the cytosol. HAdV capsid is then transported by microtubules to the nuclear membrane. Viral DNA then enters the nucleus through nuclear pores, leaving almost all the viral structural proteins in the cytosol (Henaff et al., 2011). Viral capsid proteins within the cell are eventually targeted for ubiquitination (Ko et al., 2010; Marvin and Wietthoff, 2012; Horan et al., 2013; Li et al., 2013) and degraded (Greber et al., 1993), but there are many opportunities for interaction with intracellular host cellular proteins during entry, trafficking, translation, assembly, and egress. Penton base HVL2, with its RGD motif, is critical to viral internalization through the interaction with host cell integrins; (Wickham et al., 1993) but function of penton base HVL1 is unknown, and might be revealed through knowledge of its protein interactome. The closely adjacent hexon HVL1 and 2 form the epsilon epitope that determines serum neutralization, and interactions between the hexon protein and serum coagulation factor X confers liver tropism to HAdV-C5 (Sumarheni et al., 2014). However, nothing is known about potential hexon interactions with intracellular proteins during infection.

The E3 transcription unit of HAdV codes for proteins that mediate immune evasion by the virus (Horwitz, 2004). Although E3 is labeled as an early transcription region, its transcripts are expressed both early and late during viral infection (Chow et al., 1977; Chow and Broker, 1978; Bhat and Wold, 1986), and there is evidence for at least one E3 protein that late transcripts are translated (Robinson et al., 2011a). E3 gene products are not required for viral replication in cultured cells (Morin et al., 1987), but inhibit cellular and cytokine mediated host immune responses to infection (Horwitz, 2004; Lichtenstein et al., 2004b; Windheim et al., 2004). Almost all of what is known about the function of specific E3 proteins derives from studies on HAdV-C. For example, HAdV-C2 E3 CR1α directs another E3 protein (19K) to the endoplasmic reticulum of cytotoxic T cells (Wilson-Rawls et al., 1994), where 19K binds to and retains MHC class I proteins (Jefferys and Burgert, 1990), preventing presentation of viral peptides within MHC class I at the cell surface (Burgert and Kvist, 1985, 1987; Andersson et al., 1987; Burgert et al., 1987; Cox et al., 1991). CR1α, RIDα, and RIDβ proteins cooperate to evade TNFα-related apoptosis through TRAIL (Elsing and Burgert, 1998; Tollefson et al., 1998; Benedict et al., 2001; Lichtenstein et al., 2004a). CR1β (Wold et al., 1984), also called the adenovirus death protein (Tollefson et al., 1992), is required for cell lysis (Tollefson et al., 1996) and viral spread (Dororin et al., 2003). The ORF size of each E3 gene varies across HAdV species (Figure 7) (Robinson et al., 2011c). Similarly, immune evasion functions of E3 gene products may not be the same across HAdV species, or function similarly in all cell types (Routes and Cook, 1990). Windheim and coworkers recently showed that the CR1β protein of the eye pathogen HAdV-D64 suppresses natural killer cell function (Windheim et al., 2013). The E3 CR1 genes are uniquely hypervariable within HAdV-D, and as predicted, overlapping but distinguishable intracellular interactomes across proteotypes were recently reported by Martinez-Martin and colleagues, who used protein microarrays to identify novel CR1β binding partners (Martinez-Martin et al., 2016).

![Comparison of E3 transcription unit from HAdV-C and -D. Note in particular the difference in ORF size between CR1β of the two HAdV species. Adapted from Robinson et al. (2011c) with permission.](https://example.com/figure7.png)
CONCLUSIONS

HAdV was critical to the dual discoveries of viral oncopogenesis and RNA splicing (Berget et al., 1977; Chow et al., 1977; Whyte et al., 1988). HAdV is also a significant agent of disease for which there is no approved treatment. Recent mining of HAdV genomes has been highly productive, and there is ample evidence to suggest that further whole genome analyses will elucidate new and fundamental mechanisms in HAdV biology. In the last decade, of 27 newly identified HAdVs, 19 were HAdV-Ds, suggesting the continuing evolution of new pathogens from species D. Analyses of 38 fully sequenced HAdV-D whole genomes identified homologous recombination of specific regions within the hexon, penton base, fiber, and E3 CR1 genes as the major mechanism behind HAdV-D evolution, a new finding (Robinson et al., 2013a; Singh et al., 2013). Stereotypical reductions in GC content at the junction of conserved and hypervariable regions, along with Chi-like sequence motifs (also a new finding), appear likely to augment the intrinsic tendency of HAdV to undergo homologous recombination in vivo (Lee et al., 2018).

Recently, the whole genome sequences of 85 HAdVs from archives and current collections were determined, including both historical and circulating strains, respectively (Ismail et al., 2018). Of these, 3 novel recombinants within HAdV-B and 15 within HAdV-D were identified. Only two of the 15 HAdV-Ds were found to contain novel genes (penton base and fiber); these were subsequently typed as HAdV-D71 and 72. Isolates of HAdV-D53 and HAdV-D58, two novel genotypes recently recognized, were also identified, adding confidence in their clinical importance. Fully genotyped HAdVs now number 90, with more awaiting type numbers, and the scientific community has a 10-fold larger database of unique HAdV genomes than available only 15 years ago. Published and validated ENCODE methodologies can now be applied, and comparisons made across disparate HAdV genomes. We suggest that the HAdV genome contains previously uncharacterized functional elements, and that every HAdV protein has pleiotropic interactions. Current technologies should afford a wave of new and important discoveries that may lead to needed therapies against adenoviral diseases.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This work was funded by National Institutes of Health grants EY013124, EY021558, and EY014104, a Senior Scientific Investigator Award grant (to JC) from Research to Prevent Blindness, Inc., New York, NY, The Falk Foundation, and the Massachusetts Lions Eye Research Fund.
cell specificity in epidemic keratoconjunctivitis. J. Virol. 90, 9598–9607. doi: 10.1128/JVI.01010-16

Jeffries, W. A., and Burgert, H. G. (1990). E3/19K from adenovirus 2 is an immunosubversive protein that binds to a structural motif regulating the intracellular transport of major histocompatibility complex class I proteins. J. Exp. Med. 172, 1653–1666. doi: 10.1084/jem.172.6.1653

Jennings, P. A., Finch, J. T., Winter, G., and Robertson, J. S. (1983). Does the higher order structure of the influenza virus ribonucleoprotein guide sequence rearrangements in influenza viral RNA? Cell 34, 619–627. doi: 10.1016/0092-8674(83)90394-X

Jones, M. S., Harrach, B., Ganac, R. D., Gouzum, M. M., Dela Cruz, W. P., Riedel, B., et al. (2007). New adenovirus species found in a patient presenting with gastroenteritis. J. Virol. 81, 3978–3984. doi: 10.1128/JVI.02650-06

Kalyuzhniy, O., Di Paolo, N. C., Silvestry, M., Hothers, E. S., Barry, M. A., Jones, M. S., Harrach, B., Ganac, R. D., Nagel, M. A., Choe, A., Traktinskiy, I., Cordery-Cotter, R., Gilden, D., and Cohrs, R., et al. (2011). Genetic analysis of a novel human adenovirus with a serologically unique hexon and a recombinant fiber gene. PLoS ONE 6:e24491. doi: 10.1371/journal.pone.0024491

Kang, J., Shenk, T. (1984). Adenovirus tripartite leader sequence enhances translation of mRNAs late after infection. Proc. Natl. Acad. Sci. U.S.A. 81, 3655–3659. doi: 10.1073/pnas.78.12.3655

Mahadevan, P., Seto, J., Tilghman, S. M., and Seto, D. (2010). Natural variants of human adenovirus type 3 provide evidence for relative genome stability across time and geographic space. Virology 397, 113–118. doi: 10.1016/j.virol.2009.10.052

Martinez-Martin, N., Ramani, S. R., Hackney, J. A., Tom, I., Wranik, B. J., Chan, M., et al. (2016). The extracellular interactome of the human adenovirus family reveals diverse strategies for immunomodulation. Nat. Commun. 7:11473. doi: 10.1038/ncomms11473

Marvin, S. A., and Wiethoff, C. M. (2012). Emerging roles for ubiquitin in adenovirus cell entry. Biol. Cell 104, 188–196. doi: 10.1111/boc.201100096

McCarthry, T., Lebeck, M. G., Capuano, A. W., Schnurr, D. P., and Gray, G. C. (2009). Molecular typing of clinical adenovirus specimens by an algorithm which permits detection of adenovirus coinfections and intermediate adenovirus strains. J. Clin. Virol. 46, 80–84. doi: 10.1016/j.jcv.2009.06.008

Meinschad, C., and Winnacker, E. L. (1980). Recombination in adenovirus. I. Analysis of recombinant viruses under non-selective conditions. J. Gen. Virol. 48, 219–224. doi: 10.1099/0022-1317-48-1-219

Moliner-Frenkel, V., Lengagene, R., Gaden, F., Hon, S. S., Choppin, J., Gahery-Segard, H., et al. (2002). Adenovirus hexon protein is a potent adjuvant for activation of a cellular immune response. J. Virol. 76, 127–135. doi: 10.1128/JVI.76.1.127–135.2002

Morin, J. E., Lubec, M. D., Barton, J. E., Conley, A. J., Davis, A. R., and Hung, P. P. (1987). Recombinant adenovirus induces antibody response to hepatitis B virus surface antigen in hamsters. Proc. Natl. Acad. Sci. U.S.A. 84, 4626–4630. doi: 10.1073/pnas.84.13.4626

Moyer, C. L., Besser, E. S., and Nemerow, G. R. (2016). A single maturation cleavage site in adenovirus impacts cell entry and capsid assembly. J. Virol. 90, 521–532. doi: 10.1128/JVI.02014-15

Moyer, C. L., Wiethoff, C. M., Maier, O., Smith, J. G., and Nemerow, G. R. (2011). Functional genetic and biophysical analyses of membrane disruption by human adenovirus. J. Virol. 85, 2631–2641. doi: 10.1128/JVI.02321-10

Mul, Y. M., Verrijzer, C. P., and van der Vliet, P. C. (1997). Dissociation of the protein primer and DNA polymerase after initiation of adenovirus DNA replication. J. Biol. Chem. 272, 24617–24623. doi: 10.1074/jbc.272.39.24617

Morin, J. E., Lubec, M. D., Barton, J. E., Conley, A. J., Davis, A. R., and Hung, P. P. (1987). Recombinant adenovirus induces antibody response to hepatitis B virus surface antigen in hamsters. Proc. Natl. Acad. Sci. U.S.A. 84, 4626–4630. doi: 10.1073/pnas.84.13.4626

Moyer, C. L., Besser, E. S., and Nemerow, G. R. (2016). A single maturation cleavage site in adenovirus impacts cell entry and capsid assembly. J. Virol. 90, 521–532. doi: 10.1128/JVI.02014-15

Moyer, C. L., Wiethoff, C. M., Maier, O., Smith, J. G., and Nemerow, G. R. (2011). Functional genetic and biophysical analyses of membrane disruption by human adenovirus. J. Virol. 85, 2631–2641. doi: 10.1128/JVI.02321-10

Mul, Y. M., Verrijzer, C. P., and van der Vliet, P. C. (1997). Dissociation of the protein primer and DNA polymerase after initiation of adenovirus DNA replication. J. Virol. 64, 5510–5518.

Munz, P. L., Young, C., and Young, C. S. (1983). The genetic analysis of adenovirus recombination in triparental and superinfection crosses. Virology 126, 576–586. doi: 10.1016/S0042-6822(83)80014-2

Nagel, M. A., Choe, A., Gilden, D., Traina-Dorge, V., Cohrs, R. J., and Mahalingam, R. (2013). GeXPS multiplex PCR analysis of the simian varicella virus transcriptome in productively infected cells in culture and acutely infected ganglia. J. Virol. Methods 193, 123–130. doi: 10.1016/j.jviromet.2013.07.002

Nagel, M. A., Choe, A., Traktinskiy, I., Cordery-Cotter, R., Gilden, D., and Cohrs, R. J. (2011). Varicella-zoster virus transcriptome in latently infected human ganglia. J. Virol. 85, 2276–2287. doi: 10.1128/JVI.01862-10

Nagy, P., and Bujarski, J. J. (1997). Engineering of homologous recombination hotspots with AU-rich sequences in brome mosaic virus. J. Virol. 71, 3799–3810.

Nemerow, G. R. (2000). Cell receptors involved in adenovirus entry. Virology 274, 1–4. doi: 10.1006/viro.2000.0468

Nemerow, G. R., Pache, L., Reddy, V., and Stewart, P. L. (2009). Insights into adenovirus host cell interactions from structural studies. Virology 384, 380–388. doi: 10.1016/j.viro.2008.10.016

Neumann, R., Generisch, E., and Eggers, H. J. (1987). Detection of adenovirus nucleic acid sequences in human tonsils in the absence of infectious virus. Virus Res. 7, 93–97. doi: 10.1016/0168-1702(87)90060-8

Obenauer, J. C., Denson, J., Mehta, P. K., Su, X., Mukatira, S., Finkelstein, D. B., et al. (2006). Large-scale sequence analysis of avian influenza isolates. Science 311, 1578–1580. doi: 10.1126/science.1121586

Oshima, K., Tomita, Y., Wood, J. T., Minematsu, Y., Kajiyama, H., Tomimura, K., et al. (2007). Patterns of recombination in turnip mosaic viral genomic sequences indicate hotspots of recombination. J. Gen. Virol. 88, 298–315. doi: 10.1099/vir.0.82335-0

Ottman, A. C., Brumfield, S. K., Walther, J., McInerney, K., Brouns, S. J., van de Werken, H. J., et al. (2008). Transcriptome analysis of infection of the archaeon...
Sulfolobus solfataricus with Sulfolobus tumrred icosahedral virus. J. Virol. 82, 4874–4883. doi: 10.1128/JVI.02583-07
Padio, B., Irwin, D. M., and Wilson, A. C. (1990). DNA damage promotes jumping between templates during enzymatic amplification. J. Biol. Chem. 265, 4717–4724.
Pruijn, G. J., van Miltenburg, R. T., Claessens, J. A., and van der Vliet, P. C. (1998). Interaction between the octamer-binding protein nuclear factor III and the adenovirus origin of DNA replication. J. Virol. 62, 3092–3102.
Purkayastha, A., Su, J., McGraw, J., Ditty, S. E., Hadfield, T. L., Seto, J., et al. (2005). Genomic and bioinformatics analyses of HAdV-4ac and HAdV-7ac, two human adenovirus (HAdV) strains that constituted original prophylaxis against HAdV-related acute respiratory disease, a reemerging epidemic disease. J. Clin. Microbiol. 43, 3083–3094. doi: 10.1128/JCM.43.7.3083-3094.2005
Qu, H., and Fang, X. (2013). A brief review on the human encyclopedia of DNA elements (ENCODE) project. Genomics Proteomics Bioinformatics 11, 135–141. doi: 10.1016/j.gpb.2013.05.001
Ramke, M., Lee, J. Y., Dyer, D. W., Seto, D., Rajaiya, J., and Chodosh, J. (2017). The 5’UTR in human adenoviruses: leader diversity in late gene expression. Sci. Rep. 7:618. doi: 10.1038/s41598-017-00747-y
Robinson, C. M., Rajaiya, J., Walsh, M. P., Seto, D., Dyer, D. W., Jones, M. S., et al. (2009a). Computational analysis of human adenovirus type 22 provides evidence for recombination among species D human adenoviruses in the penton base gene. J. Virol. 83, 8980–8985. doi: 10.1128/JVI.00786-09
Robinson, C. M., Rajaiya, J., Zhou, X., Singh, G., Dyer, D. W., and Chodosh, J. (2011a). The E3 CR1-gamma gene in human adenoviruses associated with epidemic keratoconjunctivitis. Virus Res. 160, 120–127. doi: 10.1016/j.virusres.2011.05.022
Robinson, C. M., Seto, D., Jones, M. S., Dyer, D. W., and Chodosh, J. (2011c). Molecular evolution of human species D adenoviruses. Infect. Genet. Evol. 11, 1208–1217. doi: 10.1016/j.meegid.2011.04.031
Robinson, C. M., Shariati, F., Gillaspy, A. F., Dyer, D. W., and Chodosh, J. (2008). Genomic and bioinformatics analysis of human adenovirus type 37: new insights into corneal tropism. BMC Genom. 9:213. doi: 10.1186/1471-2164-9-213
Robinson, C. M., Shariati, F., Zaitshik, J., Gillaspy, A. F., Dyer, D. W., and Chodosh, J. (2009b). Human adenovirus type 19: genomic and bioinformatics analysis of a keratoconjunctivitis isolate. J. Virol. 83, 3083–3094. doi: 10.1128/JVI.00969-11
Singh, G., Zhou, X., Lee, Y. Y., Yousuf, M. A., Ramke, M., et al. (2015). Recombination of the epsilon determinant and corneal tropism: human adenovirus species D types 15, 29, 56, and 69. Virology 485, 452–459. doi: 10.1016/j.virology.2015.08.018
Smith, G. R. (2012). How RecBCD enzyme and Chi promote DNA break repair and recombination: a molecular biologist's view. Microbiol. Mol. Biol. Rev. 76, 217–228. doi: 10.1128/MMBR.05026-11
Smith, G. R., Kunes, S. M., Schultz, D. W., Taylor, A. L., and Triman, K. L. (2011). Structure of chi hotspots of generalized recombination. Cell 24, 429–436. doi: 10.1016/0006-8477(81)90333-0
Spana, W., Delius, H., Skinner, M., Armstrong, J., Rottier, P., Smeekens, S., et al. (1983). Coronavirus mRNA synthesis involves fusion of non-contiguous sequences. EMBO J. 2, 1839–1844.
Stahl, F. W. (1998). Recombination in phage lambda: one geneticist's historical perspectives. Gene 223, 95–102. doi: 10.1016/S0378-1119(98)00246-7
Suyajyant, P., Chinnimanu, K., Tassaneestrith, B., Tanghawkornchaikul, N., Malasit, P., and Suriyaphol, P. (2014). Comparison of phi29-based whole genome amplification and whole transcriptome amplification in dengue virus. J. Virol. Methods 195, 141–147. doi: 10.1016/j.jviromet.2013.10.005
Sumarheni, S., Hong, S. S., Josserand, V., Coll, J. L., Boulanger, P., Schoenh, G., et al. (2014). Human full-length coagulation factor X and a GLA domain-derived 40-mer polypeptide bind to different regions of the adenovirus serotype 5 hexon capsomer. Hum. Gene. Ther. 25, 339–349. doi: 10.1038/hum.2013.222
Suwaki, N., Kaire, L., and Tarsounas, M. (2011). RAD51 paragol role in DNA damage signalling, recombinational repair and tumorigenesis. Semin. Cell Dev. Biol. 22, 898–905. doi: 10.1016/j.semcdb.2011.07.019
Tamanini, A., Nicolis, E., Bonizzato, A., Bezzerrvi, V., Melotti, P., Assael, B. M., et al. (2006). Interaction of adenovirus type 5 fiber with the cowpoxvirus and adenovirus receptor activates inflammatory response in human respiratory cells. J. Virol. 80, 11241–11254. doi: 10.1128/JVI.00721-06
Tan, D., Zhu, H., Fu, Y., Tong, F., Yao, D., Walline, J., et al. (2016). Severe community-acquired pneumonia caused by human adenovirus in immunocompetent adults: a multicenter case series. PLoS ONE 11:e0151199. doi: 10.1371/journal.pone.0151199
Taylor, A. F., Schultz, D. W., Ponticelli, A. S., and Smith, G. R. (1985). RecBC enzyme nicking at Chi sites during DNA unwinding: location and orientation-dependence of the cutting. Cell 41, 153–163. doi: 10.1016/0092-8674(85)90070-4
Tollefsen, A. E., Hermiston, T. W., Lichtenstein, D. L., Colle, C. F., Tripp, R. A., Dimitrov, T., et al. (1998). Forced degradation of Fas inhibits apoptosis in adenovirus-infected cells. Nature 392, 726–730. doi: 10.1038/33712
Tollefsen, A. E., Scaria, A., Hermiston, T. W., Ryser, J. S., Wold, L. J., and Wold, W. S. (1996). The adenovirus death protein (E3-11.6K) is required at very late stages of infection for efficient cell lysis and release of adenovirus from infected cells. J. Virol. 70, 2296–2306.
Tollefsen, A. E., Scaria, A., Saha, S. K., and Wold, W. S. (1992). The 11,600-MW protein encoded by region E3 of adenovirus is expressed early but is greatly amplified at late stages of infection. J. Virol. 66, 3633–3642.
Tookman, L. A., Browne, A. K., Connell, C. M., Bridge, G., Ingemarsdotter, C. K., Dowson, S., et al. (2016). RAD51 and BRCAR2 enhance oncolytic adenovirus Type 5 activity in ovarian cancer. Mol. Cancer Res. 14, 44–55. doi: 10.1158/1541-7786.MCR-15-0188-T
Torres, S., Chodosh, J., Seto, D., and Jones, M. S. (2010). The revolution in viral genomics as exemplified by the bioinformatic analysis of human adenoviruses. Viruses 2, 1367–1381. doi: 10.3390/v2071367
Visswanathan, V. K., Krcmarik, K., and Cianciotto, N. P. (1999). Template secondary structure promotes polymerase jumping during PCR amplification. Biotechniques 27, 508–511. doi: 10.2144/99273s004
