1. Executive Summary *(Please limit to 500 words.)*

Provide an executive summary of the proposal.

Human adenoviruses (HAdV) are infectious disease pathogens, causing different diseases in the United States as well as globally, with recent and past outbreaks in both the civilian and military populations. In this proposal, we describe our plan to analyze 150 HAdV genomes from diverse and documented sources to understand HAdV biology, especially evolution and the emergence of new and highly pathogenic serotypes. These genomes represent prototypes as well as recently isolated and archived clinical strains collected under documented circumstances, which were characterized using classical methods. Recent genome studies of several HAdVs have provided insight into the molecular mechanisms of evolution and provided understanding of emerging HAdV pathogens. This project extends these single genome studies, and serves as a model in understanding a pathogen with a wide range of pathologies in extraordinary detail based on genomics and bioinformatics, complementing the vast but now limited and "dated" resources already provided by a multitude of researchers.

As "applied research" data, these will lead to effective vaccines, with a measure of the epitopes: hexon, fiber and penton. Additionally, pathogen genome signature probes for clinical diagnostics and pathogen surveillance will be made available. These data will support the continued development and use of HAdV as vectors in gene therapy and gene delivery applications.

HAdV was characterized as a respiratory pathogen in 1953, with subsequent serotypes implicated in many diseases, including ocular, gastrointestinal, renal and metabolic. In addition to causing death, HAdV is associated with high morbidity rates, which is acknowledged as an important problem in the U.S. military.

No similar project has been attempted with any other DNA virus. Furthermore, the adenovirus is unique in the broad range of illnesses it causes at multiple mucosal sites. This suggests the potential for direct correlations between genomic sequence and tissue tropism not possible with most other pathogens. Although their 36,000 base dsDNA genomes are relatively similar, there are many sequence differences (for example, species A is 58% identical to species C). These account for their tissue tropism, virulence, pathogenicity, host response/immune systems evasion and other biological characteristics. There are 31 prototype and 37 clinical isolates genomes deposited in GenBank, with 55 defined total prototypes partitioned into seven species. We propose to sequence the remaining 24 prototypes, which are needed references for recombination analysis. The 126 characterized clinical isolates, from outbreaks and sporadic cases, will provide further insight and clarity into the pathoepidemiological dynamics of this human pathogen, explaining HAdV evolution, pathogenicity, virulence and host/tissue adaptation. These dynamics include multiple genomes from a single outbreak and from outbreaks stemming from an original outbreak to characterize genome changes in the pathogen population, including natural variation, hotspots and molecular evolution. Archived strains present snapshots of evolution of a prototype across time. As an example, the recent report of recombination and its role in the emergence of a new pathogen, HAdV-53, causing highly contagious epidemic keratoconjunctivitis (EKC), was possible using prototype genomes deposited in GenBank. This, in turn, was immediately supplemented by clinical data, demonstrating the pathogenicity of this emergent HAdV.

2. Justification

Provide a succinct justification for the sequencing or genotyping study by describing the significance of the problem and providing other relevant background information.

This section is a key evaluation criterion.

1. *State the relevance to infectious disease for the organism(s) to be studied; for example the public health significance, model system etc.*

Relevance of HAdV genomes

HAdVs have been recognized as an infectious disease pathogen since 1953 (31), and have been studied from the perspectives of virology, biology, medicine, public health and biotechnology in these five decades. They are a model organism in the research laboratory for cell and molecular biology. Ongoing public health, medical and clinical studies emphasize the importance of HAdVs as current, relevant pathogens and human health concerns. Aside from a pressing global concern, the U.S. military has special interests in them as respiratory pathogens and is funding the redevelopment and redeployment of vaccines. Ironically, despite this pathogen role, HAdVs are used also as vectors for gene therapy and gene delivery.

Goals

This project has as a practical goal to complete the genome analysis of the remaining 25 prototype genomes not sequenced and to analyze a selection of 126 clinical isolates collected from different current and historical outbreaks. The prototype genomes, of which there are now 55 accepted by the community and noted in the literature (out of which 31 genomes are deposited in GenBank) (7, 12, 14, 36), will provide: 1) reference genomes for recombination analysis; 2) identification of genome recombinations and other genome changes; and 3) historical snapshots of HAdVs isolated and characterized using classical methods from 1953 to the present. These include three new serotypes discovered and characterized recently by genomics and bioinformatics (12, 14, 36). The 125 clinical isolates to be sequenced are from six collections and were individually characterized using classical methods and are distinct potentially interesting genomes, representing a broad range of epidemic conditions: collected at various times during an outbreak; sampled as the outbreak spread geographically; sporadic cases; and even as the foci of outbreaks. They include pathogens of different HAdV-associated diseases, but particularly respiratory and ocular, as these are of the greatest concern and are of the largest numbers in these collections. Both prototype and clinical isolate genomes will provide: 1) appropriate diagnostic probes for molecular typing assays; 2) pathogen genome signatures for microarray development; 3) understanding of epitope stability for vaccine development and application; 4) SNP data for natural variation studies and 5) database of genomes relevant to the continuing use and development of HAdV as gene therapy and gene delivery vectors.

The theoretical goal is to understand HAdV biology and pathology in detail at the genome level, especially the molecular evolution of serotypes and species; the role of recombination in new serotype evolution; the role that the genome plays in HAdV biology, including cell tropism; the roles and variability of the proteome; and the role of genome contributions to the pathogenicity and infectivity of strains. As noted, some HAdV cause mild symptoms and some are even asymptomatic (7). Analogous to several recent reports (12, 14, 29, 36, 37), the emergence and establishment of new and highly pathogenic serotypes may be understood in the context of genome recombination events. Intellectually, this adenovirus genome project is unique in that no similar project has been attempted with any other DNA virus. Furthermore, the adenovirus is

unique in the broad range of illnesses it causes at multiple mucosal sites. This suggests the potential for direct correlations between genomic sequence and tissue tropism which is not possible with most other pathogens."

Aside from the basic biology questions, this work will lead to better and more effective vaccines, an understanding of vaccine effectiveness, applicable and relevant pathogenic genome signature probes for clinical diagnostics and pathogen surveillance, and will support the continued development and application of HAdV as vectors in gene therapy and gene delivery.

Public Health and novel HAdV strains: Application to Genomics

As a public health pathogen of interest, surveillance programs monitor HAdVs, including a government-sponsored ocular pathogen program in Japan (Univ. of Pittsburgh has a similar collection) and a respiratory pathogen program in the U.S. (Naval Health Research Center (NHRC)). These do not include genome sequencing and analysis as part of their routine surveillance protocols; instead, they rely on limited and now "dated" molecular and immunological techniques to characterize the collected clinical isolate. Existing surveillance programs serve as repositories and archive these isolates. Other research groups have collections of respiratory and ocular HAdV isolates as usual isolates is the lack of funding for "pure" sequencing projects of this sort, even in the context well, for example, the University of Iowa (College of Public Health), University of Pittsburgh (Dept. of Ophthalmology) and the Vision Research Foundation (Chennai, India) who are collaborators on this project. One limitation to the routine analysis of genomes from interesting and unusual isolates is the lack of funding for "pure" sequencing projects of this sort, even in the context of funded applicable clinical research such as for eye diseases. A recent study combining genomics and clinical studies using mice characterized an emergent ocular HAdV pathogen in great detail (36). Other novel, emergent, reemergent and recombinant genomes have been documented and reported (12, 14, 29, 30, 36, 37).

Human infectious agents

HAdVs are associated with a wide range of human infectious diseases, including respiratory, ocular, gastrointestinal, renal and metabolic (1, 2, 4, 7-9, 13, 14, 16, 20-22, 24-30, 34-36, 39). HAdV infections can be highly contagious, causing high morbidity rate and mortality rates and result in outbreaks. The particular disease caused by a particular HAdV is likely due to the tissue tropism of the particular serotype or strain of HAdV, which is encoded in the genome. While the HAdV genomes share genes in common, there are large differences both in sequence and in the presence or absence of certain genes (14, 26-30, 34, 36). For example, while the coat surface proteins, i.e., penton, hexon and fiber, are involved in cell receptor binding and cell entry, the small sequence differences amongst them allow a given serotype to infect the lungs, the cornea or the GI tract. There are also differences and similarities in the genome landmarks, including the critical "Inverted Terminal Repeats" (ITR), which contain viral replication and transcription functions. In addition, the "E3 genes" allow a particular HAdV to subvert the host immune system (17), supporting a view that differences in the proteome may reflect the tropism and virulence of the types.

HAdV as a zoonotic concern

Adenoviruses affect all vertebrates (10). Ones from chimpanzees are of interest for two reasons: 1) use as vectors for gene therapy and vaccine delivery (32, 33) and 2) may be reservoirs for emergent HAdV infectious agents. The sole human host member of species E is HAdV-E4, which has become embedded as an Acute Respiratory Pathogen (ARD) pathogen in the U.S. military (3, 16, 24). Sampling the serotypes and their contribution to the ARD profile has shown that HAdV-E4 was a minor contributor to ARD prior to the 1960s, becoming a major contributor in the 1970s (a vaccine was made against it) and constituting 99.9% of the HAdV ARD pathogens up to 2007, when HAdV-B14 reemerged briefly after an absence of fifty years (16, 21, 24) and became a contributing pathogen for ARD for a short time before subsequently disappeared. Genomic analysis of the prototype HAdV-E4 shows that its origin is likely a zoonosis from chimpanzees (26). Similar genomic analysis of the remaining unsequenced prototypes may reveal similar origins for human HAdVs. For reference, there are also several monkey genomes available in GenBank.

Respiratory diseases and vaccine development

The success of a genome sequencing and analysis approach is demonstrated in an analysis and re-characterization of the pathogen responsible for a recent large outbreak in China causing high morbidity and a fatality (36, 37, 39). Typically, subspecies B1 members are respiratory pathogens, in contrast to closely related subspecies B2, which are kidney, bladder and urinary tract pathogens. Exceptions are HAdV-B11a and HAdV-B14, which are respiratory agents (2, 21, 34, 39). Genomic analysis of the "HAdV-B11a" from this outbreak shows that it actually contains a HAdV-B14 genome with a recombinant HAdV-B11 hexon (37), and should be recognized as a novel emergent prototype, as HAdV-B55 (manuscript submitted). Hexon is a cell tropism determinant and a change can alter the pathology! A practical application suggests that the development of a HAdV-B11 vaccine may be useful for this emergent pathogen.

Species HAdV-B, -E and -C are respiratory disease pathogens. Outbreaks of acute respiratory disease (ARD) and respiratory diseases caused by HAdV are important global concerns (7), with recent epidemics noted in Asia (4, 39), South America (15), Europe (5, 11) and the United States (9). Within young and previously healthy U.S. basic military trainees (BMTs), ARD is a major concern. HAdV-B14 caused a large outbreak across several bases and in several U.S. cities recently (2, 16, 21, 24). Within the military, HAdVs constituted the largest percentage of the respiratory pathogens (2, 24) before the introduction of vaccines. This led to the development and deployment of two HAdV vaccines in the 1970s, which were highly effective. However, the vaccines were removed from deployment after cessation of vaccine manufacturing, which led to a rapid reemergence of ARD due to HAdV. The deployment of reconstituted vaccines, based on the original vaccines, is planned to commence in the beginning of 2010. The vaccine target is the original 1970s HAdV-E4 and -B7 hexons. Following the molecular evolution of this hexon gene, as well as the genomes of previous outbreaks, is important in maintaining the effectiveness of this 39-year old vaccine. Understanding how the entire genome evolves is also important, especially given the recombination events recently noted for the hexon and penton in other circulating and clinically relevant HAdVs (8, 13, 29, 36). Two members of the consortium proposed for this project, David Metzgar of the Naval Health Research Center (NHRC) and Adriana Kajon of the Lovelace Respiratory Research Institute, have interests in examining respiratory pathogens of species B and E strains that are circulating in the U.S. military and will continue using molecular biology techniques to characterize the clinical isolates sequenced.

These unique sets of genomes encompassing prior outbreaks, along with "linked" outbreaks (movement of personnel from other bases), will be used to track mutations and strain variation to understand the molecular evolution of HAdVs, as well as to follow the epidemiology of outbreaks, as BMTs are assigned to other military bases. Therefore, an important benefit of this comprehensive study of selected genomes is to understand the genome changes that occur in circulating HAdV-B7 and HAdV-E4 to which vaccines are being redeveloped, and to ensure these vaccines are appropriate.

HAdV-C; gene therapy and gene delivery vectors

Kajon also has interest in clinical isolates from civilian pediatric cases of ARD, specifically HAdV-E4 and several species C clinical isolates from South America. Her collection of respiratory HAdV, as well as another collection of respiratory HAdV (University of Iowa), provides a "balancing" set of civilian HAdVs. Analysis of these allows an understanding of the global spread and establishment of mutations. Species HAdV-C types are particularly interesting, as they may cause respiratory disease or, in many cases, may not- having no clinical manifestation. One long-standing hypothesis is that most adults carry HAdV-C asymptomatically from earlier infections in childhood. Recently, species HAdV-C genomes have also been reported to exhibit evidence for prior recombination (22); perhaps this is a general pathway of molecular evolution of HAdV pathogens leading to new or more significant clinical manifestations. HAdV in species HAdV-C are also of importance because both HAdV-C2 and -C5 have been used as vectors in human gene therapy trials, as well as for gene delivery experiments. Better understanding of all species C genomes will benefit their use in biotechnology.

Ocular disease and HAdV bioreactors

HAdV is also a devastating ocular pathogen. Species D contains several members (HAdV-D8, -D19, -D37, -D53 and -D54) that are associated with eye infections, e.g., epidemic keratoconjunctivitis (EKC) (13, 29, 30) globally. However, there are 33 members of species D, characterized only by serum neutralization, which targets a small portion of the hexon protein. Most of these HAdV-D serotypes are noted as gastrointestinal tract isolates. As many of the new serotypes are reported as isolates from stool samples of AIDS patients (6, 7, 18), perhaps the immunocompromised state and the recently reported phenomenon of co infections of multiple and simultaneous HAdVs (25, 35) present a pathway to the molecular evolution of emerging HAdV pathogens, including potential emergent eye pathogens. Genomic analysis will resolve this hypothesis of the human host being a "bioreactor" for new serotypes.

One graphic example of the power of the genomics and bioinformatics approach to studying adenoviruses is the new, previously "not-described" and emergent HAdV causing highly contagious EKC that was isolated in Germany in 2006 (8). This pathogen was characterized using genomics and bioinformatics recently by several members of this consortium (36). It is a new type that contains several genome recombination events, with a penton from HAdV-D37, a hexon partially from HAdV-D22, as well as a fiber and 3' region from HAdV-D8. There also appears to be a genome region recombined from an unsequenced or unknown prototype or strain (hence the need for the rest of the prototype genomes). Based on serum neutralization, it was initially identified as a variant of HAdV-D22, which does not normally cause EKC. However, this new type, HAdV-D53, is a robust highly contagious eye pathogen, causing EKC. It has also

been isolated, recently, in Japan as part of their routine surveillance, highlighting the global and rapid spread of emergent pathogens, although a complete genome analysis of this Japan strain has not yet been reported. In fact, two other examples of emergent HAdV eye pathogens have been reported recently from the Japan program (13), one of which contains a recombination event; the other as well, upon further evaluation (Seto, unpublished). A member of this consortium, James Chodosh (Massachusetts Eye and Ear Infirmary, Harvard Medical School), has access to a University of Pittsburgh collection, which contains many clinical isolates that have been collected from patients afflicted with eye diseases. Understanding the role of recombination in molecular evolution of HAdV eye pathogens, with perhaps identification of genome hotspots, may lead to better surveillance and diagnostic protocols, and may lead to vaccine development against specific determinants that differentiate pathogenicity and virulence in these recombinants.

Other diseases

HAdV is also highly associated with the gastrointestinal tract, both symptomatically and apparently asymptomatically. Recent new types of HAdV have been isolated from the stools of immunocompromised individuals, some with gastrointestinal disease. One recent example is HAdV-52, which was characterized as a new type and also a member of a new species (14). Genome sequencing and analysis has led to new definitions of HAdV taxonomy. Morris Jones (David Grant Medical Center, Travis AFB) has access to novel HAdVs, some of which are from the gastrointestinal tracts of patients from the U.S. military and the California Public Health Department (as well as additional respiratory HAdV). Jones is currently sequencing HAdV-36, which has been implicated in human obesity, presumably interfering with the metabolic physiology (1).]

Model for understanding pathogen biology

The sequencing and analysis of the rest of the HAdV prototypes will provide an invaluable reference and resource for the research community. Completion of whole genome sequences for all existing serotypes will expedite our ability to identify newly recognized and emergent viruses, and to document genome recombination in both new and old HAdVs. Having “snapshots” of original circulating strains from the pre-2009, some dating to the 1950s allows the examination of HAdV evolution and taxonomy. The application of these resources to the continuing analysis of interesting clinical isolates, including genome sequencing and analysis, is expected to lead to understanding at the molecular level of the differences between “highly contagious” versus “mildly virulent” circulating strains. These genomes will allow rapid identifications of recombination events, to provide insight into potential hotspots of recombination and/or to highlight important genome components of pathogenicity and virulence. In doing so, comprehensive genome analysis of a large sample size of HADV genomes, including all of the prototypes and a large number of current circulating clinical isolates, will provide a high-resolution map of HAdV molecular evolution and pathogenicity. Critical components of the genome and the proteome will be characterized, allowing for the development of effective surveillance methods, based on unique and common genome signatures, and of effective vaccines. Genome sequencing and computational analysis of the relatively small HAdV genomes, will permit a very broad and comprehensive database, of a group of specific, defined and related viral pathogen, and is a model for the analysis of other pathogens, including viral and bacterial. The occurrence of emerging, “not seen before”, and of reemerging, “seen before, but

not circulating", HAdV pathogens within a baseline of previously characterized symptomatic and asymptomatic HAdV strains and diseases allow a unique and demonstrable effective application of the power of high-throughput DNA sequencing technology and bioinformatic analysis. This genome sequence data set will be uniquely complemented by clinical documentation of the strains, as well as with molecular biological and clinical research across two major areas of clinical interest (respiratory and ocular).

2. *Are there genome data for organisms in the same phylum / class / family / genus? What is the status of other sequencing / genotyping projects on the same organism? Provide information on other characteristics (genome size, GC content, repetitive DNA, pre-existing arrays etc.) relevant to the proposed study. Have analyses been carried on the raw data already generated/published?*

There are genome sequences available for at least one prototype (31 total sequenced) from all of the seven species. One of these has recently been deposited and its reevaluation as a prototype is under review (37) (Seto, in review). One member of this consortium has sequenced and analyzed all of the "then-remaining" members of species B1, B2, C and E, including several recent clinical strains (19, 26-28). We propose to do the same for species A and D. There are also genomes representing clinical isolates from several species (37 sequenced) in GenBank. Currently, there are small-scale genome sequencing projects, within this group of collaborators, looking at single unusual clinical strains and putative novel HAdV. These 68 GenBank genomes have been or are currently being analyzed, allowing meaningful computational methods to be applied, for example systematic whole genome phylogeny.

Genome data: Sizes range from 34,125 (HAdV-A12) to 36,015 (HAdV-E4) nucleotides. The GC content range is from 46.5% (HAdV-A12) to 57.7% (HAdV-E4). Inverted repetitive sequences occur at both ends, ca. 150 bases, with no other significant repetitive sequences in the genome (a single very short, ca 50 bases, repeat in found in some species HAdV-B genomes). There are no pre-existing arrays that are relevant to this study; there are representative genome sequences, as noted, and a planned database designed as an AdenovirusWiki site that will serve as a community data resource (http://www.binf.gmu.edu/wiki/index.php/Main_Page) (Seto). Several genome sequence and analysis publications are in the literature, including recent ones, from the collaborators listed, either individually or collaboratively. However, none of these requested genomes have been sequenced nor analyzed beyond preliminary clinical, epidemiological and molecular biological characterizations.

3. *If analyses have been conducted, briefly describe utility of the new sequencing or genotyping information with an explanation of how the proposed study to generate additional data will advance diagnostics, therapeutics, epidemiology, vaccines, or basic knowledge such as species diversity, evolution, virulence etc. of the proposed organism to be studied.*

As noted, the prototype genomes to be sequenced are urgently needed as reference genomes for the characterization of recombination events in emerging, reemerging, pathogenic and novel HAdV isolates and to understand pathways of evolution. Some of these previously uncharacterized, and currently archived as 'historical isolates' in collections of the consortium, may actually be novel prototypes that have contributed to the recombination events of subsequent and current outbreak pathogens. The lack of genome analysis in the past, particularly the lack of support for acquiring high-throughput instrumentation and the associated costs, has hindered our progress. The development of advanced diagnostics technology, including

microarray-based techniques, is dependent on having unique signature genome sequences of the pathogens. Following the evolution of outbreak strains may allow predictions of useful probes for diagnostics and surveillance. Vaccines and their continued development are also dependent upon the hexon and other coat proteins, as will be understood better with the genome sequences proposed. An understanding of the stability of hexon epitopes will expedite the development of appropriate vaccines, much like the current vaccines for HAdV-4 and -7, and the recent HAdV-3 (23, 38). Conserved epitopes allowing widely applicable vaccines are cost-effective for manufacturers to develop and market. Species diversity, pathoepidemiology and understanding the detailed molecular evolution of HAdV, especially in the context of emergent and reemergent pathogens, including changes in tropism and virulence, have been shown on a very limited scale.

3. Rationale for Strain Selection

4. *Provide the rationale behind the selection of strains and the number of strains proposed in the study. The focus of the program is on potential agents of bioterrorism or organisms responsible for emerging or re-emerging infectious diseases. Non-select agents or non-pathogenic organisms will be considered when they can provide insight into these scientific areas.*

Understanding natural variants of pathogens and their nonpathogenic counterparts is important especially for documented outbreaks. How do new variants arise? How are they maintained in the population? Are there genome mutation hotspots? HADV is one such organism with a depth and breadth of strains relatively well defined (up to now into seven species, HAdV-A through –G, that are correlated with clinical and biological data. Human diseases range from respiratory, ocular, gastrointestinal, renal and metabolic, all of which are presumably encoded in the genomes and proteomes that are related yet different amongst the 55 serotypes. Individual strains representing a lineage of a serotype, with archived available samples, provide insight into the molecular evolution through time. Short-term variations and possible timing of mutations in outbreaks, as well as identification of genome hotspots, are approachable by examining multiple, especially “outlier” (clinically and based on serum neutralization) genomes from a single outbreak, isolated at time intervals and geographic space. There are isolates from sporadic cases. Within the collections, genomes from “linked” outbreaks, that are from consecutive or concurrent epidemics due to personnel movement, are available for analysis. Unusual clinical observations, for example severity of disease or mixed identification, may provide additional interest in the particular genome, as natural variants in these populations. As noted, the prototype genomes and possibly “yet-to-be-sequenced prototype” genomes will provide an invaluable resource to the research community especially for the characterization of recombinants that comprise emergent and reemergent pathogen population. Comparisons of genomes, particularly of recombinants, may lead to the identification of hotspots, particularly in the hexon- the epitope for clinical diagnosis and for vaccine targeting. In summary, all of the information gained on recombination, population genome comparison, pathogen emergences and even pathogenicity can be applied to the studies of other pathogens, particularly other viral pathogens.

These genomes, coupled with additional biological, clinical and epidemiological data, allow a unique opportunity to understand a collection of well-studied viruses with a diversity of genomes and diseases. The approach is to “sequence” a population rather than an “individual”.

Adenoviruses were one of the first respiratory viruses to be studied and were model organisms for molecular biology, lending insight into eukaryotic molecular mechanisms. As a model for naturally occurring and clinically relevant, pathogens and for biothreat agents, adenoviruses are still important. A member of this consortium has participated in producing, validating and deploying a military-sponsored project resulting in a successful "advanced diagnostic" microarray-based assay for surveillance of naturally-occurring respiratory agents and biothreat agents (Epidemic Outbreak Surveillance Project 2002-2005; USAF Surgeon General, DoD and DTRA). This included the adenoviruses, as a model of a reemergent pathogen in the context of the withdrawal of the vaccines by the military. As part of the probes identification protocol, the genomes of all prototype respiratory HAdV pathogens were sequenced and analyzed for unique genome signatures. HAdV co infections were first identified, characterized and documented during the course of this work. Co infection is likely a pathway for molecular recombination in the evolution of emergent and reemergent HAdV pathogens.

4a. Approach to Data Production: Data Generation

5. *State the data and resources planned to be generated. (e.g draft genome sequences, finished sequence data, SNPs, DNA/protein arrays generation, clone generation etc.)*

150 whole HAdV genomes will be newly sequenced and analyzed, representing approximately 5,400,000 bases. A whole genome shotgun strategy coupled with 454 sequencing technology will be used. Included will be the remaining 24 prototype genomes, out of 55 recognized prototypes, and 126 clinical isolates, including 1) 50 ocular isolates from Chodosh and Univ. of Pittsburgh; 10 ocular from India; 42 civilian respiratory, ocular and fecal samples (25 Lovelace Institute and 17 Univ. of Iowa), 20 military respiratory samples (NHRC) and 3 respiratory and 1 fecal sample from civilian patients (David Grant USAF Medical Ctr.). With this consortium of researchers, should there be very interesting new isolates, they may be swapped out with some of the above within the time line of this collaboration. We require high-quality finished sequences. Given the frame of extant finished sequences across all species and given the importance and well-defined protocols for sequencing, it may be cost-effective for a funding agency to support high-quality finished sequences. Within this unique population data, SNPs will be identified and probes from unique genome signatures will be available.

In order to determine the optimal sequencing strategy for 150 HAdV whole genomes we will conduct an initial pilot project on 10 isolates to test several different sequencing protocols. These methods will include but will not be limited to: whole-genome shotgun, Roche/454 and Solexa sequencing. Based on the outcome of the pilot project and taking into account overall quality of genomic assemblies, cost effectiveness and speed of data generation a protocol will be chosen and optimized for the remaining HAdV samples.

We intend to publish this work as a significant survey of genomes that comprise a single group of related yet dissimilar human pathogens, as a demonstration of a "population comparative genomics" project. Again, this adenovirus genome project differs in the diversity, variety and range of human diseases caused by a single pathogen, HAdV, and in the variations and similarities of the putatively more stable DNA-based genomes, across seven species and fifty-years of study and collection.

4b. Approach to Data Production: Data Analysis

6. *Briefly describe the analysis (value-add) envisioned to be performed subsequently by the community and the potential to develop hypotheses driven proposals given the datasets and resources produced by this work.*

Given the genome sequences, we can perform genome annotations based on our extensive experiences with these genomes, and we can complement JCVI's automated annotations of this large set of genomes. All genome sequences will be deposited in GenBank and accessible to the user community via AdenovirusWiki. Annotation data would be also submitted to the NIAID Viral Bioinformatics Resource Center (BRC). Genome computational analyses and in silico proteome analyses will provide detailed insights into the differences and similarities of these genomes, across the differences in biology, epidemiology and clinical observations, across time and space. Knowing how recombination drives pathogen emergence and reemergence, for example after 50-years of senescence, will be invaluable. Knowledge gained through the availability of new sequence data will drive hypotheses to minimize the disease impact and will especially be important to public health, in applications related to disease epidemiology and the development of diagnostics probes and tools, surveillance arrays, and vaccines and therapeutics.

5. Community Support and Collaborator Roles:

7. *Provide evidence of the relevant scientific community's size and depth of interest in the proposed sequencing or genotyping data for this organism or group of organisms.*

The latest "International Adenovirus Meeting", the Ninth, was held in April 09 in Dobogoko, Hungary, attracting over 170 researchers from all parts of the world and including presentations ranging from basic biology and evolution to applied biotechnology, including human gene therapy vector development. In between the two extremes were many scientifically diverse areas including virus structure, epidemiology, medicine, infectious disease, and vaccine development. Several of the collaborators noted here gave talks and were approached for collaborations; the adenovirus community is, on the whole, friendly and cooperative.

The public health and medical sectors, both civilian and military, are actively working on HAdV pathogens, defining outbreaks and trying to identify the pathogens. An extensive set of reference and clinical isolate genomes, as proposed here, will expedite their work by providing "state-or-the-art" data and resources to complement and extend their observations. The development of rational and appropriate vaccines is one goal in limiting the effects of HAdV globally.

8. *List all project collaborators and their roles in the project*

Donald Seto. Department of Bioinformatics and Computational Biology, George Mason University. Manassas, VA.
Principal Investigator and coordinator; Liaison with JCVI; Bioinformatics.

Adriana Kajon. Lovelace Respiratory Research Institute. Albuquerque, NM.
Co-Investigator and samples provider; Molecular biology analysis.

David Metzgar. Naval Health Research Center. San Diego, CA.
Co-Investigator and samples provider; Molecular biology analysis.

David Dyer. University of Oklahoma. Oklahoma City, OK.
Co-investigator; Genome analysis and sequencing data resource.

Morris Jones. David Grant USAF Medical Center. Travis AFB, CA.
Co-Investigator and samples provider; Molecular biology and bioinformatics analysis.

James Chodosh. Massachusetts Eye and Ear Infirmary, Harvard Medical School. Boston, MA.
Co-Investigator and samples provider; coordinator (below); Clinical and molecular biology analysis.

Regis Kowalski. University of Pittsburgh. Pittsburgh, PA. (associated with Chodosh). Samples provider; clinical analysis

H.N. Madhavan. Vision Research Foundation. Chennai, India. Samples provider; clinical analysis (associated with Chodosh).

University of Iowa, College of Public Health offers a HAdV collection from different sources.

9. *List availability of other funding sources for the project.*

The National Eye Institute (NEI/NIH) is a potential funding source for the ocular HAdV isolates; however, they are likely not to support genome sequencing but may support clinical studies based on the genomes. NIAID is a funding source for both ocular and respiratory isolates, potentially funding clinical studies based on the genomes. The military has limited funding for sequencing HAdV but has supported a small-scale genome project of 16 HAdV genomes in the past (2002) for probe identification and the development of diagnostic tools, as well as single genome analyses. They do support surveillance, archiving and serotyping of HAdV relevant to military training bases, which provides limited HAdV research.

6. Availability & Information of Strains:

10. *Indicate availability of relevant laboratory strains and clinical isolates. Are the strains/isolates of interest retrospectively collected, prepared and ready to ship?*

All samples are available as virus stocks, including the ATCC prototype strains (acquired by the consortium) with the exception of ten Indian ocular isolates- which require International permit and cost reimbursement. The University of Iowa strains need to be purchased at $85 each. Attached are two letters of collaboration and sample availability (University of Pittsburgh and Vision Research Foundation of Chennai, India).

11. *Attach relevant information, if available in an excel spreadsheet for multiple samples: e.g*

An Excel spreadsheet containing descriptions of the genomes to be sequenced is attached. All genomes are double-stranded DNA. In many cases, the particular "serotype" has been determined using serum neutralization. They are stored as virus isolates. None has been deposited in a third-party repository.

12. *What supporting metadata and clinical data have been collected or are planned on being collected that could be made available for community use?*

All metadata relevant to the strains sequenced through this proposal will be deposited in the NIAID Viral BRC.

7. Compliance Requirements:

7a. Review NIAID's Reagent, Data & Software Release Policy

Done.

7b. Public Access to Reagents, Data, Software and Other Materials:

13. *State plans for deposit of starting materials as well as resulting reagents, resources, and datasets in NIAID approved repositories.*

The genomes and their annotations will be deposited in GenBank upon the completion of the annotations and confirmation of genome sequence quality. A short embargo period will be requested to allow the consortium members time to analyze and publish their findings. However, the genome and annotations will be available, by request, to the research community. ATCC (Manassas, VA) will archive some of the important strains at a reduced fee, if funding is provided. We are receptive to depositing the strains in a NIAID approved repository.

7c. Internal Review Board (IRB) / IACUC

NIAID review of IRB documentation is required prior to commencement of work.

14. *Indicate if the study requires IRB/IACUC review.*

Any required IRB exemption documentation required by NIAID for all the clinical samples will be obtained prior to commencement of work.

15. *If an existing IRB documentation will apply to this project please attach the application(s), the protocol(s) and the consent form(s) along with the exemption / approval letter from the participating institution(s).*

None has been required by the collaborating institutions for current work.

References

1. **Atkinson, R. L., N. V. Dhurandhar, D. B. Allison, R. L. Bowen, B. A. Israel, J. B. Albu, and A. S. Augustus.** 2005. Human adenovirus-36 is associated with increased body weight and paradoxical reduction of serum lipids. Int J Obes (Lond) **29:**281-6.
2. **Binn, L. N., J. L. Sanchez, and J. C. Gaydos.** 2007. Emergence of adenovirus type 14 in US military recruits--a new challenge. J Infect Dis **196:**1436-7.
3. **Blasiole, D. A., D. Metzgar, L. T. Daum, M. A. Ryan, J. Wu, C. Wills, C. T. Le, N. E. Freed, C. J. Hansen, G. C. Gray, and K. L. Russell.** 2004. Molecular analysis of adenovirus isolates from vaccinated and unvaccinated young adults. J Clin Microbiol **42:**1686-93.
4. **Cheng, C. C., L. M. Huang, C. L. Kao, P. I. Lee, J. M. Chen, C. Y. Lu, C. Y. Lee, S. Y. Chang, and L. Y. Chang.** 2008. Molecular and clinical characteristics of adenoviral infections in Taiwanese children in 2004-2005. Eur J Pediatr **167:**633-40.
5. **Chmielewicz, B., J. Benzler, G. Pauli, G. Krause, F. Bergmann, and B. Schweiger.** 2005. Respiratory disease caused by a species B2 adenovirus in a military camp in Turkey. J Med Virol **77:**232-7.
6. **Crawford-Miksza, L., and D. P. Schnurr.** 1996. Seroepidemiology of new AIDS-associated adenoviruses among the San Francisco Men's Health Study. J Med Virol **50:**230-6.
7. **Echavarria, M.** 2008. Adenoviruses in immunocompromised hosts. Clin Microbiol Rev **21:**704-15.
8. **Engelmann, I., I. Madisch, H. Pommer, and A. Heim.** 2006. An outbreak of epidemic keratoconjunctivitis caused by a new intermediate adenovirus 22/H8 identified by molecular typing. Clin Infect Dis **43:**e64-6.
9. **Gray, G. C., T. McCarthy, M. G. Lebeck, D. P. Schnurr, K. L. Russell, A. E. Kajon, M. L. Landry, D. S. Leland, G. A. Storch, C. C. Ginocchio, C. C. Robinson, G. J. Demmler, M. A. Saubolle, S. C. Kehl, R. Selvarangan, M. B. Miller, J. D. Chappell, D. M. Zerr, D. L. Kiska, D. C. Halstead, A. W. Capuano, S. F. Setterquist, M. L. Chorazy, J. D. Dawson, and D. D. Erdman.** 2007. Genotype prevalence and risk factors for severe clinical adenovirus infection, United States 2004-2006. Clin Infect Dis **45:**1120-31.
10. **Harrach, B., and M. Benko.** 2007. Phylogenetic analysis of adenovirus sequences. Methods Mol Med **131:**299-334.
11. **Hierholzer, J. C., A. Pumarola, A. Rodriguez-Torres, and M. Beltran.** 1974. Occurrence of respiratory illness due to an atypical strain of adenovirus type 11 during a large outbreak in Spanish military recruits. Am J Epidemiol **99:**434-42.

12. **Ishiko, H., and K. Aoki.** 2009. Spread of epidemic keratoconjunctivitis due to a novel serotype of human adenovirus in Japan. J Clin Microbiol **47:**2678-9.
13. **Ishiko, H., Y. Shimada, T. Konno, A. Hayashi, T. Ohguchi, Y. Tagawa, K. Aoki, S. Ohno, and S. Yamazaki.** 2008. Novel human adenovirus causing nosocomial epidemic keratoconjunctivitis. J Clin Microbiol **46:**2002-8.
14. **Jones, M. S., 2nd, B. Harrach, R. D. Ganac, M. M. Gozum, W. P. Dela Cruz, B. Riedel, C. Pan, E. L. Delwart, and D. P. Schnurr.** 2007. New adenovirus species found in a patient presenting with gastroenteritis. J Virol **81:**5978-84.
15. **Kajon, A. E., A. S. Mistchenko, C. Videla, M. Hortal, G. Wadell, and L. F. Avendano.** 1996. Molecular epidemiology of adenovirus acute lower respiratory infections of children in the south cone of South America (1991-1994). J Med Virol **48:**151-6.
16. **Kajon, A. E., J. M. Moseley, D. Metzgar, H. S. Huong, A. Wadleigh, M. A. Ryan, and K. L. Russell.** 2007. Molecular epidemiology of adenovirus type 4 infections in US military recruits in the postvaccination era (1997-2003). J Infect Dis **196:**67-75.
17. **Kajon, A. E., W. Xu, and D. D. Erdman.** 2005. Sequence polymorphism in the E3 7.7K ORF of subspecies B1 human adenoviruses. Virus Res **107:**11-9.
18. **Koopmann, J., F. Dombrowski, J. K. Rockstroh, U. Pfeifer, T. Sauerbruch, and U. Spengler.** 2000. Fatal pneumonia in an AIDS patient coinfected with adenovirus and Pneumocystis carinii. Infection **28:**323-5.
19. **Lauer, K. P., I. Llorente, E. Blair, J. Seto, V. Krasnov, A. Purkayastha, S. E. Ditty, T. L. Hadfield, C. Buck, C. Tibbetts, and D. Seto.** 2004. Natural variation among human adenoviruses: genome sequence and annotation of human adenovirus serotype 1. J Gen Virol **85:**2615-25.
20. **Lin, B., Z. Wang, G. J. Vora, J. A. Thornton, J. M. Schnur, D. C. Thach, K. M. Blaney, A. G. Ligler, A. P. Malanoski, J. Santiago, E. A. Walter, B. K. Agan, D. Metzgar, D. Seto, L. T. Daum, R. Kruzelock, R. K. Rowley, E. H. Hanson, C. Tibbetts, and D. A. Stenger.** 2006. Broad-spectrum respiratory tract pathogen identification using resequencing DNA microarrays. Genome Res **16:**527-35.
21. **Louie, J. K., A. E. Kajon, M. Holodniy, L. Guardia-LaBar, B. Lee, A. M. Petru, J. K. Hacker, and D. P. Schnurr.** 2008. Severe pneumonia due to adenovirus serotype 14: a new respiratory threat? Clin Infect Dis **46:**421-5.
22. **Lukashev, A. N., O. E. Ivanova, T. P. Eremeeva, and R. D. Iggo.** 2008. Evidence of frequent recombination among human adenoviruses. J Gen Virol **89:**380-8.
23. **Lyons, A., J. Longfield, R. Kuschner, T. Straight, L. Binn, J. Seriwatana, R. Reitstetter, I. B. Froh, D. Craft, K. McNabb, K. Russell, D. Metzgar, A. Liss, X. Sun, A. Towle, and W. Sun.** 2008. A double-blind, placebo-controlled study of the safety and immunogenicity of live, oral type 4 and type 7 adenovirus vaccines in adults. Vaccine **26:**2890-8.
24. **Metzgar, D., M. Osuna, A. E. Kajon, A. W. Hawksworth, M. Irvine, and K. L. Russell.** 2007. Abrupt emergence of diverse species B adenoviruses at US military recruit training centers. J Infect Dis **196:**1465-73.
25. **Metzgar, D., M. Osuna, S. Yingst, M. Rakha, K. Earhart, D. Elyan, H. Esmat, M. D. Saad, A. Kajon, J. Wu, G. C. Gray, M. A. Ryan, and K. L. Russell.** 2005. PCR analysis of egyptian respiratory adenovirus isolates, including identification of species, serotypes, and coinfections. J Clin Microbiol **43:**5743-52.

26. **Purkayastha, A., S. E. Ditty, J. Su, J. McGraw, T. L. Hadfield, C. Tibbetts, and D. Seto.** 2005. Genomic and bioinformatics analysis of HAdV-4, a human adenovirus causing acute respiratory disease: implications for gene therapy and vaccine vector development. J Virol **79:**2559-72.
27. **Purkayastha, A., J. Su, S. Carlisle, C. Tibbetts, and D. Seto.** 2005. Genomic and bioinformatics analysis of HAdV-7, a human adenovirus of species B1 that causes acute respiratory disease: implications for vector development in human gene therapy. Virology **332:**114-29.
28. **Purkayastha, A., J. Su, J. McGraw, S. E. Ditty, T. L. Hadfield, J. Seto, K. L. Russell, C. Tibbetts, and D. Seto.** 2005. Genomic and bioinformatics analyses of HAdV-4vac and HAdV-7vac, two human adenovirus (HAdV) strains that constituted original prophylaxis against HAdV-related acute respiratory disease, a reemerging epidemic disease. J Clin Microbiol **43:**3083-94.
29. **Robinson, C. M., J. Rajaiya, M. P. Walsh, D. Seto, D. W. Dyer, M. S. Jones, and J. Chodosh.** 2009. Computational analysis of human adenovirus type 22 provides evidence for recombination between human adenoviruses species D in the penton base gene. J Virol.
30. **Robinson, C. M., F. Shariati, A. F. Gillaspy, D. W. Dyer, and J. Chodosh.** 2008. Genomic and bioinformatics analysis of human adenovirus type 37: new insights into corneal tropism. BMC Genomics **9:**213.
31. **Rowe, W. P., R. J. Huebner, L. K. Gilmore, R. H. Parrott, and T. G. Ward.** 1953. Isolation of a cytopathogenic agent from human adenoids undergoing spontaneous degeneration in tissue culture. Proc Soc Exp Biol Med **84:**570-3.
32. **Roy, S., G. Gao, D. S. Clawson, L. H. Vandenberghe, S. F. Farina, and J. M. Wilson.** 2004. Complete nucleotide sequences and genome organization of four chimpanzee adenoviruses. Virology **324:**361-72.
33. **Roy, S., G. Gao, Y. Lu, X. Zhou, M. Lock, R. Calcedo, and J. M. Wilson.** 2004. Characterization of a family of chimpanzee adenoviruses and development of molecular clones for gene transfer vectors. Hum Gene Ther **15:**519-30.
34. **Seto, J., M. P. Walsh, P. Mahadevan, A. Purkayastha, J. M. Clark, C. Tibbetts, and D. Seto.** 2009. Genomic and bioinformatics analyses of HAdV-14p, reference strain of a re-emerging respiratory pathogen and analysis of B1/B2. Virus Res **143:**94-105.
35. **Vora, G. J., B. Lin, K. Gratwick, C. Meador, C. Hansen, C. Tibbetts, D. A. Stenger, M. Irvine, D. Seto, A. Purkayastha, N. E. Freed, M. G. Gibson, K. Russell, and D. Metzgar.** 2006. Co-infections of adenovirus species in previously vaccinated patients. Emerg Infect Dis **12:**921-30.
36. **Walsh, M. P., A. Chintakuntlawar, C. M. Robinson, I. Madisch, B. Harrach, N. R. Hudson, D. Schnurr, A. Heim, J. Chodosh, D. Seto, and M. S. Jones.** 2009. Evidence of molecular evolution driven by recombination events influencing tropism in a novel human adenovirus that causes epidemic keratoconjunctivitis. PLoS One **4:**e5635.
37. **Yang, Z., Z. Zhu, L. Tang, L. Wang, X. Tan, P. Yu, Y. Zhang, X. Tian, J. Wang, Y. Zhang, D. Li, and W. Xu.** 2009. Genomic analyses of recombinant adenovirus type 11a in China. J Clin Microbiol.
38. **Zhang, Q., X. Su, D. Seto, B. J. Zheng, X. Tian, H. Sheng, H. Li, Y. Wang, and R. Zhou.** 2009. Construction and characterization of a replication-competent human

adenovirus type 3-based vector as a live-vaccine candidate and a viral delivery vector. Vaccine **27:**1145-53.

39. **Zhu, Z., Y. Zhang, S. Xu, P. Yu, X. Tian, L. Wang, Z. Liu, L. Tang, N. Mao, Y. Ji, C. Li, Z. Yang, S. Wang, J. Wang, D. Li, and W. Xu.** 2009. Outbreak of acute respiratory disease in China caused by B2 species of adenovirus type 11. J Clin Microbiol **47:**697-703.