Investigating how adenovirus E1A oncoproteins regulate host cell signalling pathways

BY

Asma Omran

A thesis submitted to

The University of Birmingham

For the degree of

MSc by Research Cancer and Genomic Sciences (023C)

Institute of Cancer and Genomic Sciences

The Medical School

The University of Birmingham

September 2023

Acknowledgements

I want to extend my gratitude to my supervisor Dr Andy Turnell for his continuous support and encouragement during the course of my MSc by Research. The laboratory and research skills he has taught me has been invaluable and will surely assist me in my future research endeavours. I look forward to continuing valuable research in his lab. I am thankful to my peers in the lab – Jessica Bula, Tarana Sharmin, and Sarinya Wongsanit - who have supported and guided me during my project. I am grateful for their patience, kindness, and friendship. A special thanks to my good friend Sarinya in working alongside me on Adenovirus type 2 and 12 9S E1A; her input and overall presence has been delightful. And I thank her for her unwavering belief in me. I would also like to thank my family for their constant encouragement, support, and love: the reason I am able to persevere on this journey.

I would like to extend my gratitude to Professor Joe Mymryk from The University of Western Ontario for the Adenovirus 9S-GFP plasmids.

Abstract

The adenovirus E1A 13S and 12S transcripts give rise to protein products of 289 and 243 amino acids, respectively, in Ad2/5, and differ by conserved region 3 which is present only in the 289-residue protein. These proteins are, functionally, well characterised: the CR3 region of the 13S gene product is essential for transactivation of other viral early promoters through interaction with the cellular basal transcriptional machinery, transcription factors and transcriptional regulators; whilst the 12S gene product also regulates transcription through interaction with an overlapping set of transcriptional regulators, which can bind to other regions of the E1A protein, as well as CR3, to regulate cellular transcription programmes.

To study cellular functions of the Ad5 and Ad12 13S gene product, in the absence of 12S and other E1A gene products, we have made clonal, tetracycline (doxycycline)-inducible Ad5 and Ad12 13S E1A U2OS cell lines. In this regard, we have also made tetracycline-inducible Ad5 13S E1A L1920A and Ad12 13S E1A RG2 U2OS cell lines where the function of the N-terminal region has been inactivated through mutation. Co-immunoprecipitation/Western blot studies with these cell lines revealed that E1A expressed in response to doxycycline was functional as it was able to bind known E1A-binding proteins, CBP/p300, the pRB family of proteins as well as CtBP, although we were unable to confirm the ability of potential E1A-interacting proteins, identified by mass spectrometry, to interact with the 13S E1A gene product. These cell lines do have the potential, however, to study 13S E1A function in greater detail.

In order to study the function of the 9S E1A gene product we similarly attempted to make tetracycline-inducible Ad5 and Ad12 9S E1A U2OS cell lines, though such attempts were unsuccessful. We therefore utilised GFP-tagged Ad2 and Ad12 9S E1A constructs that could be expressed transiently to study their function. GFP pulldowns from U2OS cells expressing

either GFP alone or GFP-tagged Ad12 9S E1A, coupled to mass spectrometry identified a number of cellular proteins that isolated specifically with the Ad12 9S gene product. Conventional IP-Western blot studies will need to be performed however, to identify bona fide 9S-interacting proteins before functional assays can be employed to determine the consequences of 9S interactions with these proteins.

Very little is known about the structure of E1A proteins other than that E1A function is dictated by molecular recognition features that interact with partner proteins. The 13S and 12S E1A gene products, are thought to be largely intrinsically-disordered proteins that assume structure on binding to partner proteins. We therefore utilised Alphafold 2 to determine putative structures for Ad2, Ad5 and Ad12 13S gene products. Consistent with the idea that E1A is mostly unstructured, large regions of the E1A were deemed not to have structural integrity. Interestingly, however, the proposed Zn-finger component of CR3 from these Ad species are predicted to form finger like projections, with an α-helical region separated by a short unstructured region that runs into a smaller α-helix, with the 4 Cys residues, responsible for coordinating Zn^{2+} binding, forming the base of the finger. These data suggest that CR3 from different adenovirus types are very well conserved at the structural level.

We also determined putative structures for the Ad2, Ad5 and Ad12 9S gene products, which have not been considered previously. Interestingly, 9S gene products were proposed to be highly structured, although structures generated were different for the 3 adenovirus types. Ad2 9S E1A was proposed to form a loop composed of 2 α-helices that ran parallel to each other, whilst Ad5 9S was also composed of 2 α-helices, though were proposed to be arranged perpendicular to one another. Ad12 9S was also composed of 2 α-helices that ran in series. Although these species might adopt similar structure when in association with partner proteins these data suggest that 9S from different Ad types might possess different functions.

Taken together, these studies have been useful in generating cell lines that can be used to probe E1A function in more detail and identifying putative 9S E1A-interacting proteins. It will be useful in the future to consider E1A function and interaction with cellular partner proteins in relation to specific amino acids within its predicted structure.

Contents

List of figures

List of tables

Abbreviations

2A6	Anti-Adenovirus Type 2 Fiber Monoclonal
Ad	Human Adenovirus
Ad Pol	Adenovirus DNA polymerase
AI	Artificial Intelligence
AP1	Activator protein 1
APS	APS ammonium persulfate
ATAD3A	ATPase Family AAA Domain Containing 3A
ATF	Activating transcription factor
BamHI	Restriction Enzyme From Bacillus Amyloliquefaciens H
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
BSA	Bovine serum albumin
CAR	Coxsackie virus-adenovirus receptor
CBP	CREB Binding Protein
CDK2	Cyclin-dependent kinase 2
CR	Conserved region
CtBP	C-terminal-binding protein
CTR	C-terminal region
Cys	Cysteine
DBP	DNA-binding protein
DCAF	DDB1- and CUL4- associated factor
dCMP	Deoxycytidine monophosphate
DNA	Deoxyribonucleic acid
Dyrk	Dual-specificity tyrosine-regulated kinases
E	Early
E1A	Early region 1A
E1B	Early region 1B
EBV	Epstein-Barr virus
ECL	Enhanced Chemiluminescence
EDTA	Ethylenediaminetetraacetic acid

EMBL-EBI	EMBL European Bioinformatics Institute
GFP	Green fluorescent protein
GO	Gene Ontology
HATs	Histone acetyltransferases
HBV	Hepatitis B
HCV	Hepatitis C virus
HEPES-OH	20mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HHV8	Human herpes virus 8
HIPK2	Homeodomain-interacting protein kinase 2
HPV	Human Papillomavirus
HRP	Horseradish peroxidase
hSUR2	*Herpesvirus saimiri* U RNAs
HTLV-1	Human T-lymphotropic virus-1
HVR	Hypervariable regions
I	Intermediate
IgG	Immunoglobulin G
IP	Immunoprecipitation
ITR	Inverted-terminal-repeats
KSHV	Kaposi sarcoma herpesvirus
L	Late
LB	Luria Bertani
mA	Milliamps
MLP	Major late promoter
MLTU	Major late transcription unit
MoRFs	Molecular recognition features
mRNA	Messenger ribonucleic acid
MSH	MutS homolog
NCBI	National Center for Biotechnology Information
NETN	Nuclear and cytoplasmic extraction buffer
NFI	Nuclear factor I
NP-40	Nonidet P-40
NTR	N-terminal region

NTR	N-terminal region
Oct-1	Octamer-binding protein
P/CAF	P300/CBP-associated factor
pcDNA	Plasmid cloning DNA
PCR	Polymerase chain reaction
PEI	Polyethylenimine
pFRT	Flp recombination target site
PKA RIIa	cAMP-Dependent Protein Kinase, regulatory subunit RIIa
pRB	Retinoblastoma protein
pTP	Terminal Protein precursor protein
pTP-CAT	pTP-trinucleotide intermediate forms
qRT-PCR	Quantitative reverse transcription polymerase chain reaction
R	Residue
R2G	R to G at residue 2 mutation
RNA pol	RNA polymerase
RNA	Ribonucleic acid
rpm	Revolutions per minute
rRNA	Ribosomal ribonucleic acid
RT	Room temperature
SB	Sample buffer
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	Sodium Dodecyl Sulphate Poly Acrylamide Gel Electrophoresis
SOC	Super Optimal broth with Catabolite repression
T3	Thyroid hormone
TBP	Terminal binding protein
TBS	Tris-buffered saline
TCA	Trichloroacetic acid
TEMED	N, N’, N’-Tetramethylethylenediamine
TET	Tetracycline
TP	Terminal Protein
TRRAP	Transformation/transcription domain associated protein
U2OS	Human Bone Osteosarcoma Epithelial Cells

UTB	Urea, B-mercaptoethanol, and Tris
v/v	Volume per volume
w/v	Weight per volume
WB	Western Blot
WCE	Whole-cell extract
WT	Wild type
Zn	Zinc

Chapter 1

Introduction

1.1. Oncogenic viruses

Although not part of their normal life cycle a number of human viruses do have oncogenic potential. Oncogenic viruses can induce tumourigenesis by modulating the function of the host cell genome or proteome; viral proteins regulate host cell proteins that function in cell growth and death pathways to affect their function (MacLennan and Marra, 2023). Examples of human oncogenic viruses include Epstein-Barr virus (EBV), Hepatitis B (HBV) and C virus (HCV), Human Papillomavirus (HPV), Human T-lymphotropic virus-1 (HTLV-1), Kaposi sarcoma herpesvirus (KSHV/HHV8) and adenovirus. EBV is a causative agent of Burkitt's and Hodgkin lymphoma as well as oropharyngeal epithelial cancers, whilst HBV promotes the most common liver cancer, hepatocellular carcinoma (Jiang et al., 2021). HTLV-1 is associated with Adult T-cell leukaemia (ATL;Sharma et al., 2018), HPV can induce cervical, vulvar, and head and neck carcinomas (Szymonowicz and Chen, 2020), and KSHV can promote Kaposi's sarcoma (Chang et al., 2017). Human Adenovirus (Ad) type 12 (Ad12) was the first human virus found to be tumourigenic (Trentin et al., 1962; Yabe et al., 1964). Ad12 can induce tumours in newborn rodents but given that Ad is predominantly a lytic virus it is not thought to promote human tumourigensis. However, Ad infection, in some instances, can be persistent and it has been shown that human Ads can promote cellular transformation *in vitro* by a 'hit and run' mechanism, whereby infection is oncogenic but the maintenance of the transformed phenotype requires the activation of cellular oncogenes; Ad has been found in childhood brain tumours and leukaemic cells suggesting that it might have a role in human tumourigenesis (Ip and Dobner, 2020; Kosulin et al., 2007; Gustafsson et al., 2007).

1.2. Adenovirus classification

Human adenoviruses of the *mastadenoviridae* genus that infects mammals encompass 111 types that have been classified into 7 distinct groups, A to G, based on their ability to agglutinate red blood cells and their antibody-neutralisation properties, which broadly map to their genomic sequences and phylogenetic relationships (Human adenovirus working group; http://hadvwg.gmu.edu; Xu et al., 2000). Group A adenoviruses are oncogenic and includes Ad12, 18, 31, and 31 (Dhingra et al., 2019); whilst the moderately oncogenic Group B includes Ad3, 7, 7a, and 11; the non-oncogenic group C includes Ad1, 2, 5, 6, and 57; non-oncogenic group D includes Ad8, 9, 13, and Ad30; non-oncogenic group E includes Ad4; group F includes Ad40, 41, and 40/41; and group G includes Ad52. The ability of group F and G viruses to induce oncogenesis is not known. Although most adenoviruses do not have oncogenic potential the E1 and E4 genes from most groups do possess the ability to transform human and rodent cells *in vitro* (Ip and Dobner, 2020).

1.3. Adenovirus structure

Adenoviruses (Ad) are non-enveloped viruses comprising of a linear double-stranded DNA core and an outer protein capsid with icosahedral structure ranging from 65-90 nm in size (Kulanayake and Tikoo, 2021). The capsid is composed of 252 capsomeres, comprising 240 hexon proteins, and 12 penton bases that have protruding fibres. Hexons are the predominant structural component of the capsid; they are homotrimers of over 900 residues and comprise hypervariable regions (HVR) and a conserved base (Crawford-Miksza and Schnurr, 1996). The seven loops of the HVR facilitate binding to cells and possesses antibody neutralising activities (Marquez-Martinez et al., 2023). The penton fibres are required for cell entry via the coxsackie virus-adenovirus (CAR) receptor, whilst the penton base facilitates virus entry into the cell through its association with cell surface integrins. Additionally, there are minor proteins called

IIIa, IX, VI, and VIII which help stabilise the capsid. There are seven A ‘packaging domain’ repeats (5′-TTTGN8CG-3′) in the Ad genome (nucleotides 200-380), which function to package the genome into the capsid. Ad capsid proteins L1-52/55K, L4-22K, and IVa2 all interact with the Ad genome via its packaging domain to allow for DNA packaging in the virion. In this regard, L4-22K interacts with the packaging domain via the TTTG motif whilst IVa2 interacts via the CG motif (Ma and Hearing, 2011). Ad packaging also requires the viral proteins IIIa. Indeed, it has been shown that IIIa interacts with L1-52/55K to enhance Ad packaging specificity (Ma and Hearing, 2011; Pérez-Berná et al., 2014).

The Ad genome consists of early (E), intermediate (I), and late (L) transcription units. E1A, E1B, E2, E3, E4 are the early region transcripts. E2-late, IVa2, and IX are the intermediate transcription units (Marquez-Martinez et al., 2023), and the major late transcription unit (MLTU) yields five mRNA families, L1 to L5, which encode the capsid proteins; the largest E1A 13S mRNA species activates E promoters whilst 13S E1A and IVa2 activates the major late promoter (MLP). Both the intermediate transcription unit and MLTU are activated in response to adenoviral DNA replication onset (Farley et al., 2004).

1.4. Adenovirus replication

The Ad genome possesses two identical, inverted-terminal-repeats (ITR) of approximately 100 base pairs (bp) located at either end of the virus genome; these regions contain the origin viral DNA replication. The terminal 18 bp makes up the minimal origin and the remaining sequence functions as an auxiliary origin. The 5’ ends of the adenovirus genome are capped by Terminal Protein (TP), which are presumed to prevent cellular exonucleases from degrading the viral genome and facilitating DNA unwinding. Adenoviral replication requires three viral proteins encoded by the E2 genes: Terminal Protein precursor protein (pTP; encoded by E2B, which

can be processed by a viral protease to TP); Ad DNA polymerase (Ad Pol; encoded by E2B); and DNA-binding protein (DBP; encoded by E2A). These viral proteins bind to two, cellular transcription factors that bind the auxiliary origin: nuclear factor I (NFI); and octamer-binding protein (Oct-1), which allow for origin bending and enhance initiation more than 100-fold. TF association with viral origins is regulated by DBP, which also facilitates double-stranded DNA unwinding and DNA elongation.

Protein-primed DNA synthesis initiated by this preinitiation complex starts with the covalent addition of deoxycytidine monophosphate (dCMP) to a serine residue in pTP which becomes the first nucleotide of the newly synthesized viral DNA; in Ad5, Ser580 of pTP is used as a primer for the covalent attachment of dCMP by Ad Pol to pTP. Replication initiation occurs at position four of the origin sequence 3'-G_1TAG_4TA; subsequently, pTP-trinucleotide intermediate forms (pTP-CAT). This pTP-CAT intermediate jumps back three bases and pairs with template residues 1-3 (seq 3'-GTA); this jumping mechanism results in short 3 base pair repeat sequences in the first 10 bp of all Ad origins. The pTP-CAT intermediate is then elongated by Ad Pol, facilitated by DBP, to form a new duplex genome as the non-template strand is displaced**.** New Ad genome duplexes (replication or complementary displaced single strand annealing) enter subsequent replication rounds, are processed and encapsidated/packaged into new virions (King and van der Vliet, 1994; Hoeben and Uil, 2013).

1.5. Adenovirus E1A structure and function

The major adenovirus 13S E1A and 12S E1A gene products are identical except for conserved region 3 (CR3, which is evolutionarily conserved between E1A species from different Ad types) which is present in the 289-residue species but not the 243-residue species. Both proteins possess three other conserved regions, CR1, CR2 and CR4 that, like CR3, are similarly

conserved between E1A from different Ad types. CR's are highly modular in function, and possess short amino acid motifs, termed molecular recognition features (MoRFs) that selectively bind to different cellular proteins and dictate function. These E1A species also possess a less well conserved N-terminal region (NTR), but studies suggest they also bind to a common set of cellular proteins (Gallimore and Turnell, 2001).

Although Ad E1A species do not bind DNA or RNA their primary role is to regulate host and viral transcription programmes to facilitate viral replication. E1A, in the presence of cooperating oncogenes such as E1B-55K or *ras* promotes cellular transformation and has therefore been used as a model system to understand the molecular events that regulate transformation. In the absence of cooperating oncogenes, E1A can in isolation induce apoptosis by up-regulating the tumour suppressor protein, p53 (Lowe and Earl Ruley, 1993).

The functions of E1A are determined by E1A's ability to interact with a number of cellular transcriptional regulators (Figure 1.1). The N-terminal region (NTR) of E1A, in cooperation with CR1, binds to the histone acetyltransferases (HATs), CBP, p300 and P/CAF; p300 was first identified through its interaction with E1A. Dependent on cellular and promoter context, and the different multiprotein complexes containing E1A formed, E1A interaction with these HATs can either stimulate or inhibit HAT-dependent transcription. Important studies by Arnie Berk's laboratory showed that the 12S E1A gene product promoted the hypoacetylation of histone H3 K18 to inhibit global transcription programmes requiring this modification (Horwitz et al., 2008; Ferrari et al., 2008). Indeed, 12S E1A was shown to bind to promoters of cell cycle genes and antiviral genes to modulate histone acetylation in a temporal manner to either activate or repress transcription (Ferrari et al., 2008, 2014). The ability of E1A to transform cells is dependent upon its ability to interact with CBP and p300; a single mutation within the NTR, R to G at residue 2, ablates E1A's ability to interact with CBP/p300 and promote transformation (Wang et al., 1993). Indeed, the ability of E1A to induce p53-

dependent apoptosis is also dependent upon its ability to interact with CBP/p300 (Querido et al., 1997). The NTR of 12S E1A also binds to TBP to repress RNA pol II-dependent transcription (Boyd et al., 2002) and p400 and TRRAP at different residues within the NTR to promote cellular transformation (Fuchs et al., 2001).

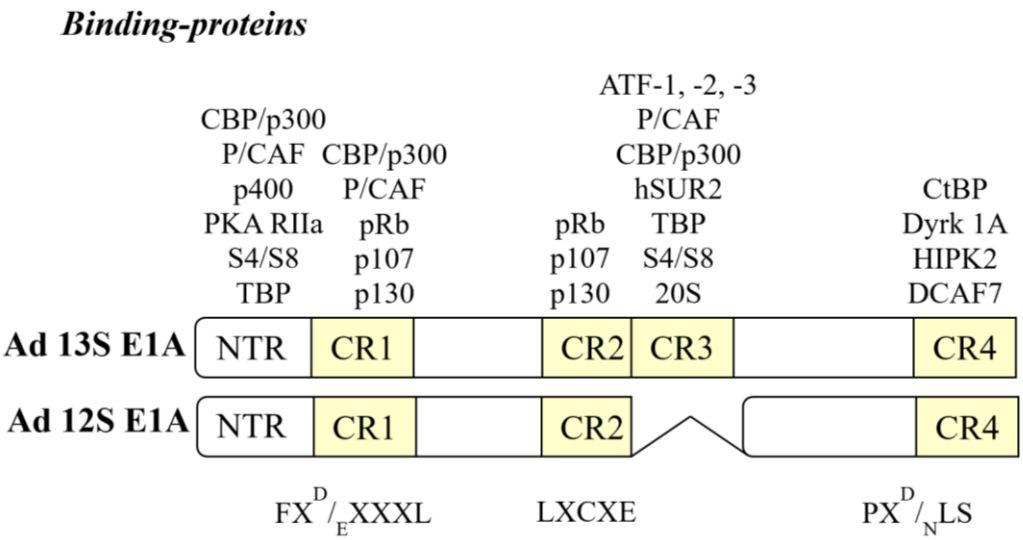

Figure 1.1. E1A domain structure, E1A-binding proteins and functional binding motifs: This schematic picture indicates the conserved regions of the larger 12S and 13S E1A gene products and the key cellular proteins known to bind to different regions of the E1A proteins. The binding motifs in CR1, CR2 and CR4 are also shown. Modified from Gallimore and Turnell, 2001.

CR1 and CR2 bind to the pRB family of proteins, namely pRB, p107 and p130 through a conserved LXCXE motif located in CR2 (Egan et al., 1988). In binding to pRB, E1A displaces pRB from E2F, allowing the free E2F to stimulate E2F-dependent transcription programmes that promote S-phase entry. In this regard E1A mimics Cyclin-dependent kinase 2 (CDK2)-dependent phosphorylation of pRB, which similarly serves to activate E2F transcription and promote S phase entry. As such, a major role of E1A during infection is to promote S-phase entry to provide the cellular factors required for viral DNA replication (Gallimore and Turnell, 2001).

CR4 is located within the C-terminal region (CTR) of E1A. The CTR of E1A contains a PLDLS motif that associates with C-terminal binding proteins (CtBP1/2) which function as transcriptional repressors; E1A association with CtBP modulates E1A-induced cellular transformation (Zhao et al., 2008). CR4 also interacts with protein kinases Dyrk1A and HIPK2 through interaction with the adaptor protein DCAF7 (Glenewinkel et al., 2016). Transformation studies with CTR mutants that are unable to bind these proteins indicates that they suppress E1A + *ras* transformation such that E1A mutants unable to bind Dyrk1A, HIPK2 and DCAF7 promote hyper-transformation (Cohen et al., 2013).

CR3 is unique to larger 13S E1A species and is required for transactivation of viral early and late promoters and so is essential for a productive viral infection (Lillie and Green, 1989). In Ad5, CR3 comprises a Zinc Finger between residues 147 and 177, with essential cysteine residues at 154, 157, 171 and 174 coordinating a single Zn^{2+} ion (Figure 1.2). The CTR of CR3 (183-188) is responsible for promoter targeting and recruiting cellular Transcription Factors such as ATF-2, and Transcription Associated Factors- TAFs to promoter regions (Geisberg et al., 1994, Webster and Ricciardi, 1991). The Zn Finger has been shown to bind TBP, a component of the general transcription machinery; Sur2, a component of the human Mediator complex (Boyer et al., 1999); and the 26S proteasome through both 19S RP and 20S components (Rasti et al., 2006). In this regard, CR3 mutants that do not bind Sur2 or the proteasome are transactivation-defective, indicating their importance in E1A-mediated transactivation and viral replication. A role for CBP/p300 and P/CAF in CR3-dependent transactivation has also been established as these proteins all associate specifically with the Zn finger domain (Pelka et al., 2009a, 2009b).

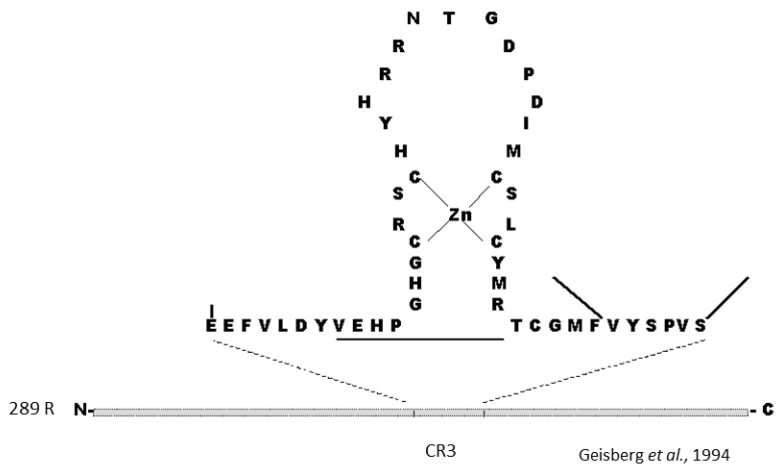

Figure 1.2. Schematic illustration showing the zinc finger motif in CR3 of adenovirus 13S E1A. The four cysteines are Cys 154, 157, 171, and 174 and coordinate a single Zn^{2+} ion. The zinc finger region spans residues 147 to 177 in Ad2/5 (Geisberg et al., 1994).

1.6. Adenovirus 9S E1A

The 55-residue gene product of the Ad2/Ad5 9S E1A transcript, produced by differential splicing of E1A exons, was first identified 45 years ago (Berk and Sharp, 1977; Chow et al., 1979). It is the most divergent of E1A splice variants, sharing the first 26 amino acids of the N-terminal region with all E1A species but through differential splice-site usage has a distinct C-terminal region, constituting a unique 29 amino acids not present in other E1A variants (Figure 1.3).

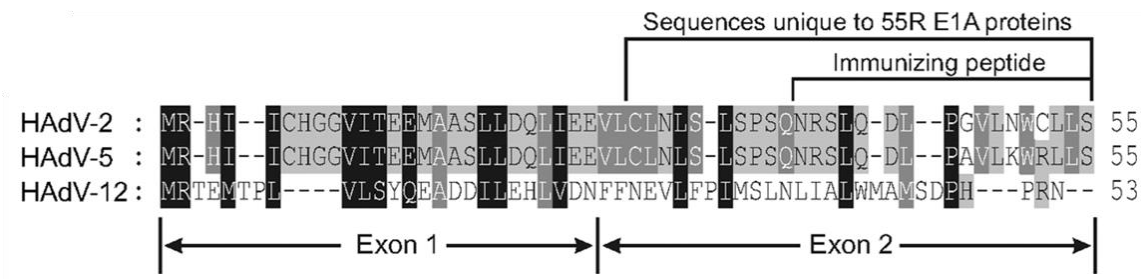

Figure. 1.3. Sequence alignment of the Ad2, Ad5 and Ad12 55-R E1A species, showing conserved and similar residues (Taken from Miller et al., 2012).

RNA analysis of Ad2-infected HeLa cells has indicated that in contrast to other E1A splice variants, 9S transcript levels are low early during infection but it is the most abundant E1A

species during late infection (Ulfendahl et al., 1987). Despite these early studies the function of the 9S E1A gene product has not been studied in great detail and, as such, the functions of the 9S E1A gene product are not well described. A more recent study that was able to generate a polyclonal antibody that recognises 9S E1A alone (see Figure for epitope) and functions in WB, IP and IF confirmed that maximal expression of the 55-residue protein was late during infection and that it was localised both in the cytoplasm and nucleus (Miller et al., 2012).

These studies also determined that the N-terminal region of the E1A 55-R species, like 12S and 13S E1A gene products (Turnell et al., 2000; Rasti et al., 2006), interacted with the S8 component of the 19S regulatory particle of the proteasome (Miller et al., 2012). Moreover, a follow-up study determined that like the 13S and 12S E1A gene products, the 9S E1A gene product interacts with the unliganded thyroid hormone receptor through a CoRNR box motif, and acts to stimulate thyroid hormone receptor activity, that is suppressed by the addition of thyroid hormone, T3 (Arulsundaram et al., 2014). Despite these insights into 55-R E1A function, virtually nothing is known about the function of the 29 C-terminal amino acids.

1.7. Aims and Objectives

The main aims of this project were to:

1. Generate and validate TET-inducible Ad13S E1A and Ad13S E1A-RG2 U2OS clonal cell lines so that we could study the function of Ad5 and Ad12 CR3 in the absence of viral infection and 12S E1A expression.

2. Generate and validate TET-inducible Ad9S E1A-GFP U2OS clonal cell lines so that we could study the function of both the Ad2 and Ad12 55-R E1A species in isolation and without the need for a 55-R specific antibody.

3. Determine the E1A 55-R interactome by GFP-pulldown coupled to mass spectrometry so that we could identify, and potentially characterise new E1A 55-R-interacting proteins and elucidate new functions for E1A 55-R, in particular the unique C-terminal region.

In this regard, these aims necessitate the cloning of the respective genes of interest into a TET-inducible expression plasmid and validation by sequencing, and the generation, selection and expansion of clonal cell lines using antibiotic resistance. An underlying objective of the project was to create these new reagents so that they could be used for other E1A studies in the laboratory, in addition to those outlined here.

Chapter 2

Materials and Methods

2.1. Cell biology methods

2.1.1. Cell culture

All of the cell lines used during this study were maintained in DME-HEPES, 8% (v/v) foetal calf serum and 2mM glutamine (complete medium; Sigma-Aldrich) and incubated in a NuAire™ DH Autoflow CO_2 Air-Jacketed Incubator at 37°C. Normally, the osteosarcoma cell line, U2OS (HTB-96, ATCC) was used for transfection, and FRT-U2OS cells (ThermoFisher), which express the TET-repressor and have one Flp recombination in the cellular genome, were used to facilitate the generation of TET-inducible cell lines. For cell splitting medium from the plates was discarded, and then washed with 8 ml of PBS, twice. TrypLE trypsin was then added to the cells (ThermoFisher), which were then incubated at 37°C for 5 minutes in the NuAire™ DH Autoflow CO_2 Air-Jacketed Incubator. Plates were removed and rocked to allow to visualise the detached cells. Fresh medium was added to the plates and the cells were fully resuspended and replated at the required density. Optional step: the contents were transferred to a conical flask and centrifuged at 1600 g for 3 minutes at 24°C. This formed a pellet and the supernatant was discarded. The pellet was resuspended in the required amount of fresh medium, cells were counted and replated at the required cell number. Plates were swirled to mix and then incubated at 37°C in the incubator until required.

2.1.2. Generation of inducible cell lines

The specific E1A splice variant of interest was cloned into the TET-responsive expression plasmid, pFRT/TO (ThermoFisher) and validated by sequencing (see section.2.3.12). For transfection onto a 90% confluent 10cm dish of FRT-U2OS cells 1.5µg of pFRT/TO-E1A and 13.5µg of POG44 (plasmid expressing recombinase) was incubated with 15µl of Lipofectamine 2000 (ThermoFisher) in 0.5ml low serum, Optimem (Thermofisher) for 45 min at RT, and then incubated with FRT-U2OS cells for 6 h in 5.5ml (total volume) of Optimem.

After this time, the transfection mixture was replaced with complete medium for 48h, after which cells were incubated in hygromycin (100μg/ml) for approximately 3-4 weeks (medium was changed every 3 days). After this extended time in selection medium single cells grew into discrete, rounded colonies that were then picked with a sterile 1ml pipette tip, in a sterile flow-cabinet using low-power microscopy. These clones were maintained in hygromycin and expanded for further study.

2.1.3. Transient transfections

For transient transfection we used Polyethylenimine (PEI). Initially, 100mM PEI stocks were generated in 20mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES-OH), vortexed, filter sterilized through a 0.22μm filter and stored at -80C until needed. For optimisation of transfection, we mixed different ratios of plasmid to PEI into 500μl, 20mM HEPES-OH (pH 7.4) containing 150mM NaCl, vortexed and left to stand for 30 min at RT. The PEI-DNA mixture was then added to U2OS cells in Optimem medium at approximately 50% confluency and incubated at 37ºC for 4-6 h. The medium was then discarded and replaced with completed medium and incubated until required. Alternatively, Lipofectamine 2000 (Life Technologies) was mixed with the appropriate amount of DNA in Optimem and then were similarly incubated with the cells at 37ºC for 4-6 h, after which time the medium containing the transfection mix was replaced with complete medium.

2.1.4. Fluorescence microscopy

In order to visualise GFP expression and cellular localisation of GFP-tagged proteins, our GFP-U2OS cells, and GFP-Ad2-9S U2OS cells and GFP-Ad12-9S U2OS cells were visualised

under Nikon Eclipse TS100 Microscope and images captured using EVOS M5000 microscope Invitrogen ThermoFisher Scientific Fluorescent Microscope.

2.2. Protein biochemistry methods

2.2.1. Cell lysate preparation

Cell viability and confluency were checked under the light microscope prior to harvesting. To generate cell lysates, the medium was discarded, and the cells washed with ice-cold PBS.

For Western Blot (WB) analyses cells were lysed in UTB (9M urea; 150 mM B-mercaptoethanol, 50 mm Tris (pH 7.4)) and left for 5 min (until the consistency of the solution became viscous). Cell lysates were then scraped with a cell-scraper in downward motions whilst rotating the plate; so that as much of the lysate could be collected). The scraper was tapped on the dish to remove any excess and the lysate was collected into an Eppendorf. Cell lysates were then sonicated to disrupt DNA using a Misonix Microson Ultrasonic Cell Disruptor at setting 5 for 20 sec, and then centrifuged at 4°C for 30 min at 16,000rpm and then stored at -80°C until needed. For immunoprecipitation (IP) or GFP pulldown cells were lysed in NETN (150mM NaCl, 1.0mM EDTA (pH 8.0) 25mM Tris-Cl (pH 7.4), 1% (v/v) Nonidet P-40 (NP-40) at 4°C and isolated with a scraper as outlined above. Cell lysates were similarly disrupted with a sonicator and centrifuged to remove insoluble material, after which lysates were used fresh, without freezing, for IP studies. With a needle, the solution in the eppendorfs were carefully taken up taking care not to disturb the film on top of the solution, the white streak in the solution, and the pellet. The remaining solution in the eppendorfs were combined and centrifuged again and the method was repeated until as much solution as possible was taken up.

2.2.2. Protein concentration

For the Bradford Assay, Eppendorfs were prepared with the following amounts of Bovine serum albumin (BSA): 0, 5, 10, 20, 30 µg; 1 ml of Bradford solution was then added to each sample. Eppendorfs were also prepared for the lysates: typically 4 µl of samples (so that readings were generally in the middle of the standard curve) and 1 ml of Bradford were mixed together. The eppendorfs were then vortexed for a few seconds to mix. The spectrophotometer was set at 595nm and blanked with distilled water. The absorbance of the samples were then measured and recorded, and the cuvette washed between samples with Bradford reagent alone. A standard curve was produced on Excel using the BSA samples to generate an equation for the line so that we could calculate the volume of sample required for 50 µg of protein sample for WB, or 1mg+ for IP.

2.2.3. SDS-PAGE

10% (w/v) polyacrylamide gels were made using Acrylamide (Acrylamide Bis-Acrylamide Stock Solution 40% (w/v) Acrylamide Ratio 37.5:1 Bis-Acrylamide, Severn Biotech Ltd, Product no. 20-3600-10, Batch no. 22798); Tris (0.1M); Bicine (0.1M); SDS (0.1% (w/v)); N, N, N', N'-Tetramethylethylenediamine (TEMED) (99%, Sigma-Aldrich, Lot no. BCCF2825) (0.3% (v/v)); ammonium persulfate (APS) (0.06% (w/v)). TEMED was added next to last then APS was added last to polymerise the gel; the gel was left to set for 30 minutes. Wells were washed with dH_2O and then filled with running buffer (0.1M Tris-Bicine, 0.1% (w/v) SDS) to prevent the gel from drying.

To prepare samples for SDS-PAGE they were mixed with an equal amount of sample buffer (SB; 6M Urea, 33.3mM Tris pH 7.5, 3.33% (w/v) Sodium Dodecyl Sulphate (SDS), 0.1% (w/v) Bromophenol Blue). Eppendorf lids were then pierced with a needle, and samples were heated

on a heat block at 95°C for 5 min. Samples were then centrifuged for 1 min at 13,000 rpm and loaded onto the gel. 10 µl of stained molecular weight marker was added to the first well. The powerpack was set between 16 and 22 milliamps (mA) and run for approximately 18 h depending on the size of the proteins to be studied.

2.2.4. Western blotting

To prepare for Western blot transfer, the following layers were sandwiched between two plastic cassettes in this order: two sponges, filter paper, nitrocellulose membrane, the gel, filter paper, one sponge. All layers were soaked in transfer buffer (25 mM Tris, 20% (v/v) methanol, 190 mM glycine). The cassette lid was placed on top and the cassette clamped closed and placed in the buffer tank. The tank was filled with Tris buffer and the machine was set to 280 mA and left for 5 h so that the proteins transferred from the gel to the nitrocellulose in the direction of the anode. Following transfer efficiency could be evaluated qualitatively by staining the gel with 0.1% (w/v) Ponceau S and 3% (w/v) trichloroacetic acid (TCA) to visualise the bands on the nitrocellulose. Ponceau S was removed from the nitrocellulose by incubation in TBST (Tris Buffered Saline, 0.1% (v/v) Tween 80).

2.2.5. Antibodies

Blots were blocked in 5% (w/v) milk in TBST for 30 min to 1 h. Blots were then incubated with primary antibodies in 5% (w/v) milk in TBST (see Table 2.1) in the cold room (4°C) overnight on a rocker. Blots were then washed in TBST and incubated in secondary antibodies for 3 hours on the rocker at RT. The secondary antibodies used were mouse or rabbit (see Table 2.1) depending on what primary antibody was used. Blots were then washed in TBST three times (15 min each) prior to Enhanced Chemiluminescence. ECL substrate (Immobilon®

Western Chemiluminescent HRP substrate, Millipore, Sigma-Aldrich, P90720, cat no. WBKLS0500, lot no. 2213702) was prepared with equal parts Peroxide solution and Luminol Enhancer solution. Blots were incubated in ECL for 2 minutes. Blue-Sensitive X-Ray films (Wolf laboratories) and the Compact X4 developer (X-ograph Imaging System) were used to visualise the blots.

Table 2.1: antibodies used in this study

Antigen	Antibody	Dilution	Origin	Supplier
CBP	767	1:500	Rabbit	In-house
CtBP	M18	1:20	Mouse	In-house
E1A – Ad12	#13 (WB)	1:20	Mouse	In-house
E1A – Ad12	#5 (IP)	1:20	Mouse	In-house
E1A- Ad5	M58	1:10,000	Mouse	In-house
E1A – Ad5	610	1:5000	Rabbit	In-house
E1B55K–Ad12	XPH9	1:4000	Mouse	In-house
E1B55K– Ad5	2A6	1:4000	Mouse	In-house
GFP	B-2	1:1000	Mouse	Santa Cruz Biotechnology
Mouse IgG	Anti-mouse HRP Ref# P0447	1:4000	Goat	Dako
MSH2	SC-494	1:1000	Rabbit	Santa Cruz Biotechnology
MSH6	A300-023A	1:2000	Rabbit	Santa Cruz Biotechnology
p107	C-18, sc-318,	1:500	Rabbit	Santa Cruz Biotechnology
p300	N-15, sc-584,	1:500	Rabbit	Santa Cruz Biotechnology
p53	DO-1	1:40	Mouse	Original source, David Lane
pRB	IF-8, sc-102	1:1000	Mouse	Santa Cruz Biotechnology

Rabbit IgG	Anti-rabbit HRP Ref# P0399	1:4000	Swine	Dako
ATAD3A	A-4, sc-376185	1:1000	Mouse	Santa Cruz Biotechnology
β-actin	AC-74	1:50,000	Mouse	Sigma-Aldrich

2.2.6. Immunoprecipitation

For IP from E1A-expressing cells we first induced the expression of E1A in clonal FRT-U2OS cell lines by incubating cells for 24 h in complete medium containing doxycycline (0.1μg/ml). Cells were then lysed and prepared in NETN (as described section 2.2.1) and proteins concentrations measured (section 2.2.2). The appropriate IP'ing antibody or IgG control was incubated with equal amounts of protein (>1mg/sample) overnight on a rotator at 4°C. Thirty-μl of packed protein G Sepharose beads (KPL) was then added to the IP for an additional 3 h on a rotator at 4°C. Beads were then precipitated by centrifugation at 4°C then washed 5 times with NETN buffer. After the final wash, a fine gauge needle was used to remove residual buffer, before being mixed with SB and prepared for SDS-PAGE.

2.2.7. GFP-pull down

Cell lysates, from either GFP-U2OS cells or GFP-Ad12 9S expressing cells, were prepared as described in section 2.2.1. Typically over 1mg clarified protein lysate per sample, was incubated with 5μl packed GFP-TRAP beads (Chromotek) and left to mix overnight on a rotator at 4°C. Beads were then precipitated by centrifugation at 4°C then washed 5 times with NETN buffer. After the final wash, a fine gauge needle was used to remove residual buffer, before being mixed with SB and prepared for SDS-PAGE.

2.2.8. Mass spectrometry

After SDS-PAGE, the gel was fixed in 50% v/v methanol, 20% v/v acetic acid for 1h. After this time gel slices were excised and processed for mass spectrometry (MS). Gel slices were then washed for 1 h in 50 % (v/v) acetonitrile (Millipore) and 50 mM ammonium bicarbonate (Fisher) at 37 °C. Next, gel slices were treated with 50 mM DTT in 10 % (v/v) acetonitrile, 50 mM ammonium bicarbonate for 1 h at 55 °C to reduce proteins, after which they were carboxymethylated with 100 mM iodoacetamide (Sigma Aldrich) for 30 min in the dark in a buffer containing 10 % (v/v) acetonitrile and 50 mM ammonium bicarbonate. Slices were then washed with 10 % (v/v) acetonitrile, 40 mM ammonium bicarbonate and then dried by washing in 100% v/v acetonitrile. Gel slices were then rehydrated in 10 % (v/v) acetonitrile, 40 mM ammonium bicarbonate in the presence of sequence-grade modified trypsin (Promega), and incubated at 37 °C for 18 h, such that in-gel trypsinisation of proteins occurred. Supernatants, containing peptides, were then removed from the samples, and transferred to low protein binding Eppendorf tubes, after which gel slices were washed twice with 3 % (v/v) formic acid (Sigma Aldrich) to maximise tryptic peptide extraction from the gel. Washes were combined with the original supernatant and dried under vacuum and stored at – 20 °C until analysed.

Samples to be analysed were resuspended in 100 µl 2 % (v/v) acetonitrile, 1 % (v/v) formic acid and loaded onto an Ultimate 3000 HPLC column (Dionex) for peptide separation before entry into a maXis Impact time of flight mass spectrometer (Bruker Daltronics) for peptide isolation, fragmentation and detection. Ms/ms data was analysed by Bruker Daltronic software ProteinScape by comparison to the MASCOT database (Matrix Science).

2.2.9. Generation of protein structures using Colab AlphaFold2 and ChimeraX

Artificial intelligence (AI) software is useful in predicting three-dimensional structures of proteins of unknown structure. An example of such software is DeepMind and EMBL-EBI AlphaFold AI programme (https://alphafold.com/; Jumper et al., 2021; Varadi et al., 2022). This software can be used to predict the structure of over 200 million proteins. As AlphaFold did not have a predictive image of any E1A species we used Google's Colab AlphaFold2 program to input 13S and 9S protein sequences for Ad2, Ad5, and Ad12, so that we could generate 3D images of E1A. To further analyse the side chains and conformation of different E1A species, PDB-files generated by Colab AlphaFold2 were uploaded to UCSF ChimeraX (http://www.rbvi.ucsf.edu/chimerax/; Pettersen et al., 2021; Goddard et al., 2018). "Molecular graphics and analyses performed with UCSF ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from National Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases".

2.3. Molecular biology methods

2.3.1. PCR

To amplify E1A gene sequences for cloning PCR was performed. For a 50µl reaction, 50ng template DNA (in 0.5µl) was added to 10 µl of 5x Q5 reaction buffer, 1µl of 10mM dNTPs, 2.5 µl of 10µM forward primer, 2.5 µl of 10µM reverse primer, 0.5 µl Q5 high fidelity DNA polymerase, and 32.5 µl of nuclease-free water. Samples were run using a 2720 Thermal cycler PCR machine (Applied Biosystems) which was typically set to complete 35 cycles of the following: 98ºC (30s), 98ºC (10s), 60ºC (30s), 72ºC (20-30s per kb), 72ºC (2 minutes), 4ºC.

Oligonucleotide primers used during this study to amplify E1A species by PCR or for sequencing:

Ad12 R2G BamHI Forward primer: TGCATGGATCCATGGGAACTGAAATGACTCCCT

Ad12E1A XhoI Reverse: TGCATCTCGAGTTAATTACATCTAGGGCGTTTCAC

Ad12 E1A Sequencing primer, TGCACCCTGAAGATATGGATTTAT

pcDNA5 FRT Sequencing primer: ATAGAAGACACCGGGACCGATCCA

2.3.2. PCR clean-up

Firstly, the columns and collection tubes were prepared, and 500 µl of Column Preparation Solution was added to the centre of the columns. The columns were centrifuged at 13,000 rpm for 1 min and the flow-through was discarded. Next, 250 µl of Binding Solution was added to the PCR product (1:5 of PCR product to Binding Solution). Once combined, the mix was added to the column and centrifuged at 13,000 rpm for 1 min and the flow-through was discarded. Next the column was washed with 500 µl of Wash solution and centrifuged at 13,000 rpm for 1 min and the flow-through again discarded. The column was then centrifuged for an additional 2 min to remove any excess ethanol from the filter. The collection tube was then replaced with a fresh 2 ml collection tube. 50 µl of nuclease-free water was added to the centre of the column and left to incubate at RT for 1 minute. It was then centrifuged for 1 minute at 13,000 rpm to elute the DNA.

2.3.3. Preparation of media and plates

To prepare Luria Bertani (LB) agar plates, LB agar (1.5% (w/v) agar: pre-made and autoclaved) was melted, in a microwave at low power for 10 min. Once cooled (for 15-20 minutes), ampicillin was added to the LB agar solution (at a final concentration of 100µg/ml) in the GMAG sterile flow hood and mix thoroughly. The LB agar was then poured into 10 cm plates, allowed to cool and set for 15 minutes. Once translucent, the plates were placed upside down for an additional 5 min to complete the drying process. The plates were then wrapped in Nesco Film, placed in a sterile plastic bag and sealed tightly. The plates were stored in the cold room (4°C).

2.3.4. Transformation of bacteria

DNA plasmids were obtained from the -80ºC and left to thaw on ice. They were centrifuged, placed on ice, and subsequently heated on a heat block at 65°C for 20 min. Samples were returned to ice and clean eppendorfs were prepared. 10 µl of competent bacteria (DH5α) was combined with 5 µl of mini-prep DNA and placed directly on ice for 30 min to 1 h. Samples were then heat-shocked in a water bath at 42°C for 1 min and left on ice for 5 min. 200 µl of super optimal broth with catabolite repression (SOC) medium was then added to each sample, which were inverted to mix and placed in the orbital shaker (37°C) at 200 rpm for 1 h.

Samples were taken out of the orbital shaker and pipetted onto respective LB agar plates and spread across the plates using a plastic hockey-stick. The plates were left to dry for 5-10 min in a sterile flow hood and stacked upside down and left in the incubator (37°C) overnight.

2.3.5. DNA Purification - Miniprep

Single colonies were picked into 5ml LB medium containing 100µg/ml ampicillin in a 15 ml tube. Colonies were grown overnight in the orbital shaker at 37ºC. The bacteria were pelleted by centrifugation at 4000 rpm at 4ºC for 5 min. The supernatant was discarded, after which the pellets were resuspended in 200 µl of mP1 resuspension buffer (containing RNAse A). Samples were transferred to eppendorfs and 200 µl mp2 lysis buffer was then added; samples were inverted 10 times to mix and left to stand for 2 minutes. Next, 300 µl mP3 neutralisation buffer was added and samples inverted. The samples were centrifuged at 13,000 rpm for 5 minutes or until the supernatant was clear. The supernatants were transferred to FastGene mP columns (without disturbing the pellet) which were centrifuged at 13,000 rpm for 1 min, whereupon the liquid in the collection tube was discarded. Next, 400 µl of mP4 First wash buffer was added to the FastGene mP columns and samples were centrifuged at 13,000 rpm for 1 min. The liquid in the collection tube was again discarded. Then, 600 µl of mP5 second wash buffer was added and samples were centrifuged again at 13,000 rpm for 1 min. The contents of the collection tube were similarly discarded. The columns were then spun for 2 min to remove any remaining buffer. Columns were then placed into fresh eppendorfs and 100 µl of nuclease-free water was added. Samples were left to stand for 5 minutes and centrifuged at 13,000 rpm for 5 min to collect the eluted DNA.

2.3.6. DNA Purification - Maxiprep

200ml of bacteria grown overnight was transferred to centrifugation flasks and centrifuged for 5000rpm at 4°C for 15 min. The supernatant was discarded, and the bacterial pellets were fully resuspended in 12 ml of Buffer RES, after which 12 ml Buffer LYS was added and the mixture was left to stand at RT for 5 minutes. The columns and filters were then prepared, as 25 ml of Buffer EQU was added to the rim of the column in circular manner to ensure wetting of the

tube. The sample was then neutralised with 12 ml Buffer NEU and mixed thoroughly until colourless. The bacterial lysate was inverted thrice, and then loaded onto the NucleoBond Xtra Column Filter. To wash the sample, 15 ml Buffer EQU was added to the rim of the column in circular motions. After all the solution had dripped through the column, the column filter was discarded. For the second wash, 25 ml of Buffer Wash was added to the column. 15 ml Buffer ELU was then added to the column to elute the DNA, which was collected in a centrifuge tube. To precipitate the DNA, 10.5 ml of isopropanol was added to this tube, and the sample was left to stand at RT for 30 minutes. Finally, samples were centrifuged at 16,000rpm for 30 min. After centrifugation and in a sterile hood, the supernatant was discarded, and 1 ml of 70% ethanol was added to the flask to detach the DNA pellet from the flask. The pellet and ethanol were transferred to an Eppendorf, inverted, and centrifuged at 13,000 rpm for 1 min. Samples were washed twice with 1 ml of 70% ethanol and centrifugation for 1 min. Residual ethanol was removed, and the pellets were left to air dry for 10-15 minutes. Pellets were resuspended in ~500 µl of nuclease-free water, depending on the size of the pellet.

2.3.7. Measuring DNA concentration

DNA concentrations were measured using the Thermo Scientific NanoDrop™ 1000 Spectrophotometer. The NanoDrop was first blanked with distilled water and then 2 µl of DNA sample was used to measure its concentration and 260/280nm ratios calculated.

2.3.8. DNA digestions

To prepare the digests, 10 µl of the purified Miniprep DNA was combined with 2 µl Cutsmart Buffer, 1 µl BamHI, 1 µl XhoI, and 6 µl nuclease-free water. For the undigested samples, 10 µl of Miniprep DNA was used alone. Samples were incubated in a water bath at 37°C for 3 h.

Samples were then transferred to a 65°C heat block for 15 min and subsequently centrifuged for 30 sec prior to analysis upon an agarose gel.

2.3.9. DNA ligations

For DNA ligation, insert DNA and plasmid DNA that had been treated with appropriate restriction enzymes were typically mixed in a molar ratio of 3:1 in the presence of 1x T4 buffer and 1 unit of T4 DNA ligase (NEB) and incubated overnight at 14°C. After this time reactions were treated at 65°C for 20 min to inactivate the ligase. Typically, 1/10th of the ligation mix was then used to transform bacteria to isolate clonal colonies that possessed constructs containing plasmid and insert.

2.3.10. DNA agarose gels

To prepare an 0.8% (w/v) agarose gel, 0.48 g agarose powder (Sigma-Aldrich) was dissolved in 60 ml of 1x TBE (0.13M Tris (pH 7.6), 45mM Boric acid, 2.5mM EDTA) in the microwave. For gel purification, 1µl of SYBR green (Merck) was added to the agarose solution in the flask before pouring the gel. In this case DNA was subsequently visualised using a blue light transilluminator (GeneFlow). For analytical gels, and gel imaging using the Gene Flash Syngene Bio Imaging equipment, ethidium bromide (1µg/ml; Sigma-Aldrich) was added to the agarose solution in the flask before pouring. After the gel had set DNA samples were mixed with 6x sample buffer (30% (v/v) glycerol, 0.25% (w/v) bromophenol blue, and 0.25% (w/v) xylene cyanol, loaded into the wells of the gel and electrophoresed at a constant 60 V, for 30-40 minutes in the presence of 1x TBE. 1 kb ladder (GIBCO BRL® Life Technology) was used to gauge the size of DNA.

2.3.11. Gel extraction of DNA

Protocol to extract/purify 70 bp to 10 kb DNA

Following the protocol included with the kit, 600 µl of Buffer QG solubilisation buffer was added to each gel sample: then, samples were heated on a heating block set to 50°C for 10 minutes to dissolve the gel, whereby samples were inverted every 2-3 min. Next, 200 µl of isopropanol was added and eppendorfs were inverted to mix. QIAquick® spin columns were then placed in 2 ml collection tubes and samples were added to the column in 700 µl increments and centrifuged for 1 min at 13,000rpm after each addition. Flow-through in the collection tube was discarded and 500 µl Buffer QG was added to each column as a further washing step. The columns were then washed with 750 µl Buffer PE and centrifuged for 1 min at 13,000rpm with the flow-through being discarded. The columns were centrifuged for a further min to remove excess ethanol. Columns were then placed in clean 1.5 ml microcentrifuge tubes with the lids cut off. To elute the DNA, 100 µl of nuclease-free water was added to each column. Columns were left to stand for 4 minutes and were subsequently centrifuged for 1 min at 13,000rpm to elute the DNA. The DNA concentration was then measured using the Nanodrop.

2.3.12. Sanger Sequencing of DNA

After gel imaging, the correctly identified samples were chosen for Sanger sequencing. Typically, one sequencing reaction of 20µl contained 200ng of mini-prep plasmid DNA, 10ng/µl of the appropriate sequencing primer; 4 µl 5x sequencing buffer; 1 µl Big Dye™ terminator V3.1 (Thermo Fisher); 9 µl nuclease-free water. PCR was then performed (25 cycles): 96°C for 10°C seconds; 55°C for 5 seconds; 60°C for 4 minutes. After PCR 62.5 µl 100% (v/v) ethanol, 3 µl 3M sodium acetate (Na-Acetate), and 14.5 µl nuclease-free water (80 µl) were added to the PCR reaction, vortexed to mix, left for 30 min at RT and then centrifuged

at 13,000rom to precipitate the DNA. The supernatant was the removed carefully without touching the side of the Eppendorf where the pellet had collected. Samples were then washed with 100 µl of 70% (v/v) ethanol, vortexed, and then centrifuged at 13,000 rpm for 15 min. Residual ethanol was removed, and samples were washed once more with 100 µl of 70% (v/v) ethanol, vortexed, and centrifuged at 13,000 rpm for 15 min. The residual ethanol was removed and the samples were air-dried. The PCR samples were resuspended in 11 µl of Hi-Di, vortexed, and heated at 100°C for 5 minutes in the PCR machine and subsequently placed on ice to cool. Samples were then loaded onto a 96-well PCR plate (ensuring no bubbles) and DNA sequence analysed using a 3500xl Genetic Analyzer (Applied Biosystems). The results from sequencing were annotated and stored using Chromas Lite analysed against reference sequences using the NCBI Nucleotide BLAST program (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=MegaBlast&PROGRAM=blastn&BLAST_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch&BLAST_SPEC=blast2seq&DATABASE=n/a&QUERY=&SUBJECTS=).

Chapter 3

Results

3.1. Generation of clonal Ad5 and Ad12 13S E1A TET-inducible U2OS cell lines

Most of the studies, to date, investigating Ad E1A 13S function in isolation have relied upon either the generation of mutant viruses that only express the 13S E1A species or transient transfection studies, where 13S E1A cDNA cloned into a mammalian expression plasmid is transfected into the cell. To study 13S E1A function and interaction with the host cell more conveniently, the primary aim of this part of the project was to generate clonal Ad5 and Ad12 E1A 13S E1A TET-inducible U2OS cell lines and then validate their usefulness as a tool to study 13S E1A function. To do this, we used the FlpIn system whereby the gene to be expressed is cloned into a mammalian expression vector and transfected into an FRT-TO U2OS cell line along with a recombinase that facilitates the integration of the gene of interest into the host genome at an engineered recombination locus; regulated expression is controlled by the TET repressor, which is constitutively expressed in the FRT-TO U2OS cell line.

3.1.1. Cloning of 13S E1A species into pcDNA5/FRT/TO expression vector

The first step towards generating these cell lines was to clone wild-type Ad5 and Ad12 13S E1A and NTR 13S E1A inactive mutants (Ad5, L1920A where L is mutated to A; Ad12, R2G where R is mutated to G) into the pcDNA5/FRT/TO expression vector. For Ad12 13S E1A species this first necessitated PCR with primers that introduced a BamHI restriction site immediately preceding the 5’ ATG (Ad12-E1A-BamHI or Ad12-E1A-R2G-BamHI) and a 3’ XhoI site immediately following the stop codon (Ad12-E1A-Xho). PCR and gel electrophoresis for validation was performed as described (see Materials and Methods sections 2.3.1 and 2.3.9 respectively). For Ad5 13S E1A and 13S-E1A-L1920A species 5’ BamH1 and 3’ XhoI sites had already been incorporated into the cDNA and validated by Sanger sequencing such that PCR was not needed to amplify these genes.

Next, we treated the Ad5 E1A PCR products and the Ad12 E1A expression plasmid with BamH1 and XhoI to excise the E1A species, which were then purified following the separation of DNA by agarose gel electrophoresis and DNA concentration measured. We similarly treated the pcDNA5/FRT/TO expression vector with these restriction enzymes to create sites for ligation of E1A species into this expression vector (see Materials and Methods sections 2.3.1-2.3.11 for details). We next performed ligation reactions whereby the digested DNA inserts were cloned into the pcDNA5/FRT/TO vector (see section 2.3.9). Following ligation, *E. coli* DH5α bacteria were transformed with ligation samples so that the new constructs could be replicated and analysed further. Transformed bacteria were grown and plated onto LB-agar in the presence of ampicillin and individual bacterial colonies picked and grown for further analysis. Following the miniprep, the DNA concentration was measured using the Nanopore and the DNA was treated in the absence or presence of restriction enzymes BamHI and XhoI and underwent gel electrophoresis to determine whether the selected clones had the 13S E1A cDNAs inserted. DNA gels were visualised using Syngene Bio Imaging equipment.

Figures 3.1 A (Ad5 13S and 13S L1920A) and B (Ad12 13S and 13S R2G) illustrate that in the absence of restriction enzymes only 1 major DNA species was detected. However, in the majority of cases in the presence of BamHI and XhoI, a second smaller species was detected that corresponded to E1A 13S species, suggesting that the cloning procedure had been successful. Indeed, 13S E1A species identified were approximately 900 bp in size which compares favourably with the predicted size of 867 bp.

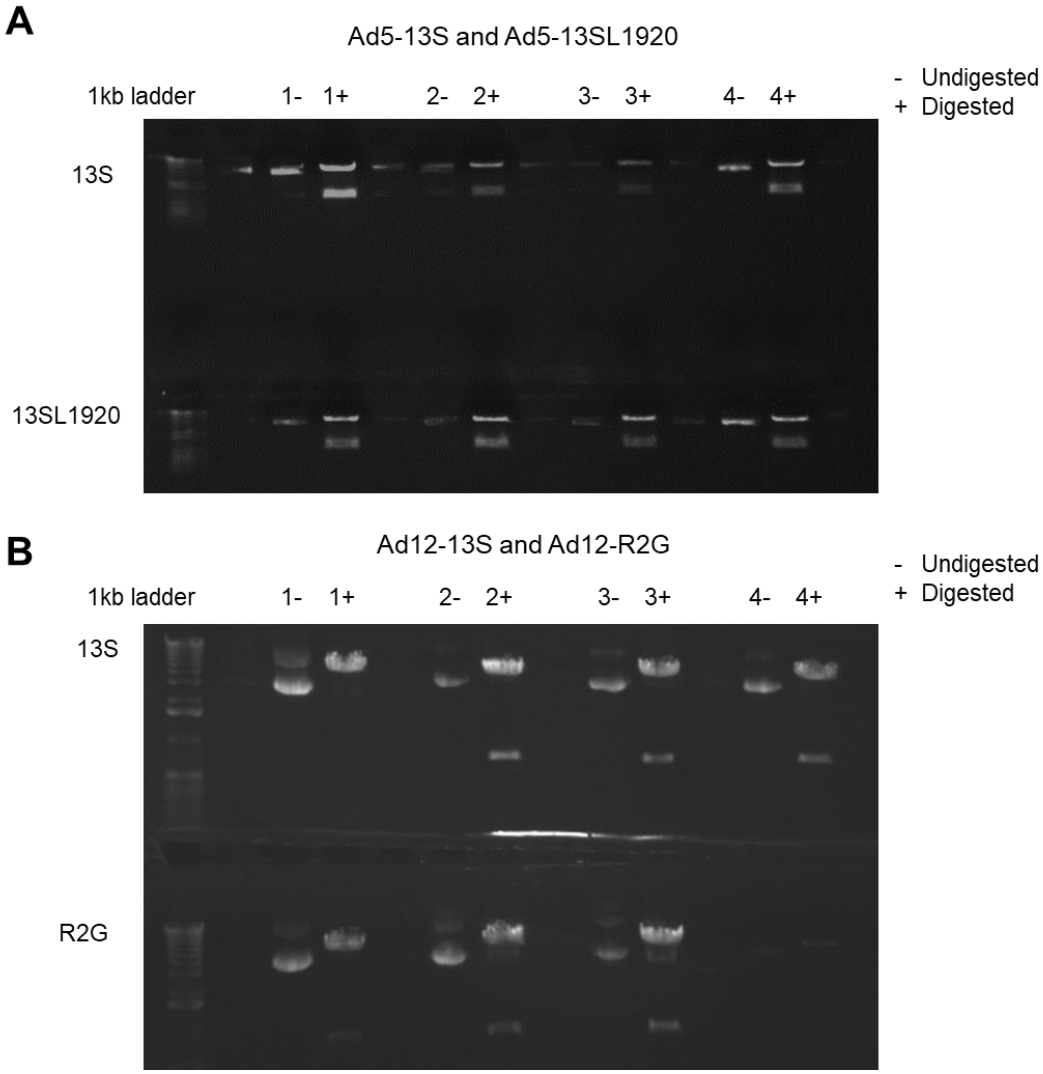

Figure 3.1. Double digest showing successful cloning of 13S E1A cDNA into the vector. Gel electrophoresis images of (A) Ad5-13S WT and Ad5-13SL1920 and (B) Ad12-13S WT and Ad12-R2G following BamHI and XhoI digestion. Images were taken using Syngene Bio Imaging equipment.

3.1.2. Sanger sequencing of isolated pcDNA5/FRT/TO clones harbouring 13S E1A

Gel electrophoresis determined the successful cloning of Ad5-13S (Fig. 3.1A) and Ad12-13S (Fig. 3.1B) into the pcDNA5/FRT/TO vector. Successful clones from the Ad12 cloning experiment were subsequently selected and sequenced by Sanger Sequencing using the appropriate sequencing primers (see Materials and Methods section 2.3.12). As the generation of 13S E1A and 13S E1A-L1920A pcDNA5/FRT/TO constructs required no PCR step and

were generated by the subcloning of previously sequenced clones, the positive clones identified did not require to be sequenced.

The results from sequencing were annotated and stored using Chromas Lite and analysed against reference sequences using the NCBI BLAST program. Through sequencing, the full Ad12 13S E1A sequence, restriction sites, START and STOP codons were successfully identified through the use of NCBI BLAST and Chromas Lite image indicating the integrity of the clones (Fig. 3.2 and data not shown). The results helped to select the best samples for downstream experiments i.e., the generation of inducible cell lines.

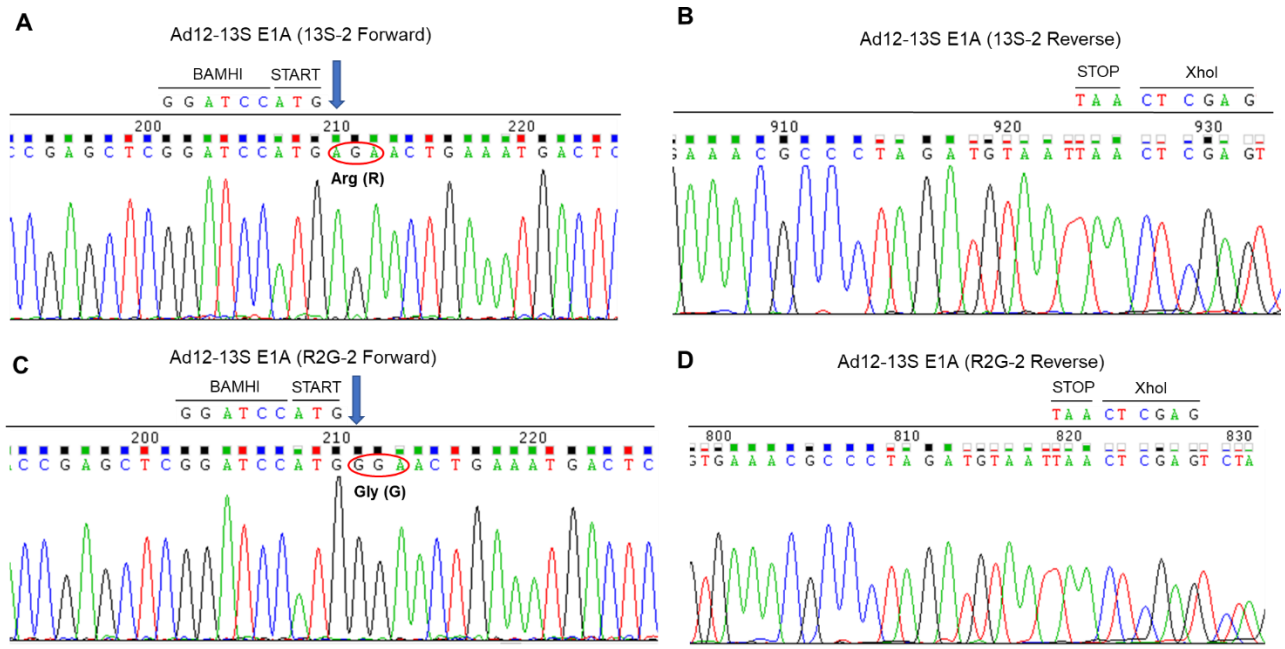

Figure 3.2. Identification of Ad12 13S E1A positive clones with wild-type sequence. Sanger Sequencing of (A) Ad12-13S E1A (13S-2 Forward) with BAMHI and XhoI; (B) Ad12-13S E1A (13S-2 Reverse); (C) Ad12-13S E1A (R2G-2 Forward); (D) Ad12-13S E1A (R2G-2 Reverse). The arrows indicate the codon change from AGA (A) to GGA (C).

3.1.3. Transfection and selection of FRT/TO U2OS cells

To generate clonal Ad5-13S E1A and Ad12 13S E1A- inducible U2OS cell lines a 90% confluent 10cm dish of FRT/TO-U2OS cells were transfected with the appropriate pFRT/TO-E1A construct and POG44 (plasmid expressing recombinase) in a 1:9 ratio using Lipofectamine 2000 (ThermoFisher) for 6 h in Optimem medium (Life Technologies). After this time, the transfection mixture was replaced with complete medium for 48h, after which cells were incubated in hygromycin (100µg/ml) for approximately 2 weeks (medium was changed every 3 days) until individual colonies could be picked under the light microscope and expanded for further study (section 2.1.2). Clonal cell lines were maintained in hygromycin. Interestingly, the number of clones that were recovered from the selection procedure for both Ad5 and Ad12 was very low (approximately 6 for each condition). This compares unfavourably with the expression of other genes of interest being studied in the laboratory, using the FlpIn U2OS system. The potential reasons for this will be discussed in more detail in the discussion.

3.1.4. Validation of clonal Ad5-13S E1A and Ad12 13S E1A- inducible cell lines

Next, we wanted to establish whether the selected Ad5 and Ad12 clones expressed E1A in TET-responsive manner. To do this, we treated cells in the absence or presence of 0.1 µg/ml doxycycline, which is a more stable, but functionally similar, analogue of tetracycline, for 24 h. After this time cells were harvested in UTB buffer, sonicated to shear DNA, and centrifuged to remove insoluble material. After protein quantification proteins were loaded onto SDS-PAGE gels and were subjected to WB analysis for Ad5 13S E1A or Ad12 13S E1A.

All five WT Ad5-13S E1A clones showed strong expression of E1A when treated with doxycycline (Fig. 3.3A), whilst four Ad5-13S E1A L1920 (mutant) clones also showed strong E1A expression in the presence of doxycycline; very little, but noticeable, E1A expression was

observed in the absence of doxycycline for both WT and mutant Ad5-13S E1A (Figure 3.3A). WB analysis revealed that for WT Ad12-13S, five clones strongly expressed 13S E1A when treated with doxycycline (Fig. 3.3B). Similarly, most of the Ad12-13S E1A L1920A cell lines expressed 13S E1A when treated with doxycycline, the exception being clone 6 that expressed substantial E1A, even in the absence of doxycycline (Fig. 3.3B).

As the smaller 12S E1A species has previously been shown to induce apoptosis by up regulating p53 we investigated whether 13S E1A had similar properties. WB analyses revealed however that p53 expression was unaffected by doxycycline treatment, although it was clear there was variability in p53 levels between different clones (Fig. 3.3C and D). Taken together these data indicate that we have successfully generated multiple cell lines that express E1A in response to doxycycline that could be valuable for studying the function of this E1A species. The generation of the NTR inactive mutants would be valuable in this regard to determine the function of CR3 in the absence of NTR activity, which binds to a similar set of cellular proteins.

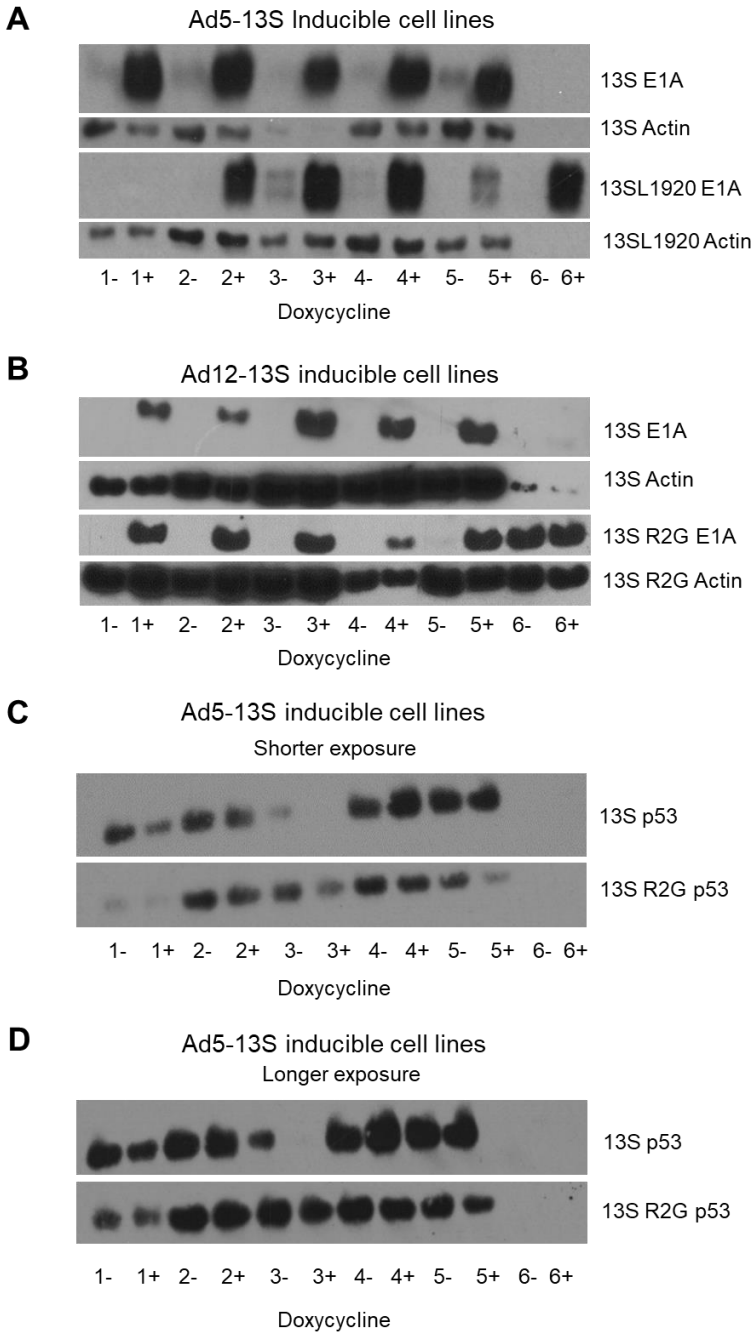

Figure 3.3. Identification of clonal cell lines that express Ad5 and Ad12 13S E1A in response to doxycycline. (A) Ad5 and Ad12 13S and 13SL1920 E1A SDS-PAGE. Ad5-13SL1920 inducible cell lines. (B) Ad12-13S and Ad12-13SL1920. (C) Ad12-13S and Ad12-13SL1920 - blotted for p53 (shorter exposure). (D) Ad12-13S and Ad12-13SL1920 - blotted for p53 (longer exposure).

3.1.5. Functional validation of TET-inducible Ad5 13S E1A U2OS cell lines

E1A does not possess enzymic activity and functions solely through interaction with cellular proteins to form a number of multiprotein complexes that, depending on the cellular context,

regulate transcriptional activation or repression programmes to modulate biological processes such as cell growth, cellular transformation or apoptosis. As described in the introduction there are a number of known E1A-binding protein (section 1.5). To validate the function of our TET-inducible WT Ad5 and Ad12 13S E1A clonal cell lines, we decided to investigate whether E1A expressed in response to doxycycline interacted with some of its well-known interacting proteins.

To study this, we needed to isolate proteins in complex with E1A. Ad5-13S E1A was therefore immunoprecipitated from doxycycline-treated cells with anti-Ad5 E1A antibodies M58, and M73 antibodies or IgG (control). Thus, following induction, cells were lysed in NETN and prepared for IP (as described section 2.2.1, 2.2.2 and 2.2.6. After IP and SDS-PAGE WB revealed that E1A was successfully immunoprecipitated specifically from cells along with known binding partners p107, pRB, CBP and p300 (Fig. 3.4A and B). This suggests that the Ad5 13S E1A species expressed in response to doxycycline is functional as it has the ability to interact with proteins that associate with the NTR, CR1, CR2 and CR3.

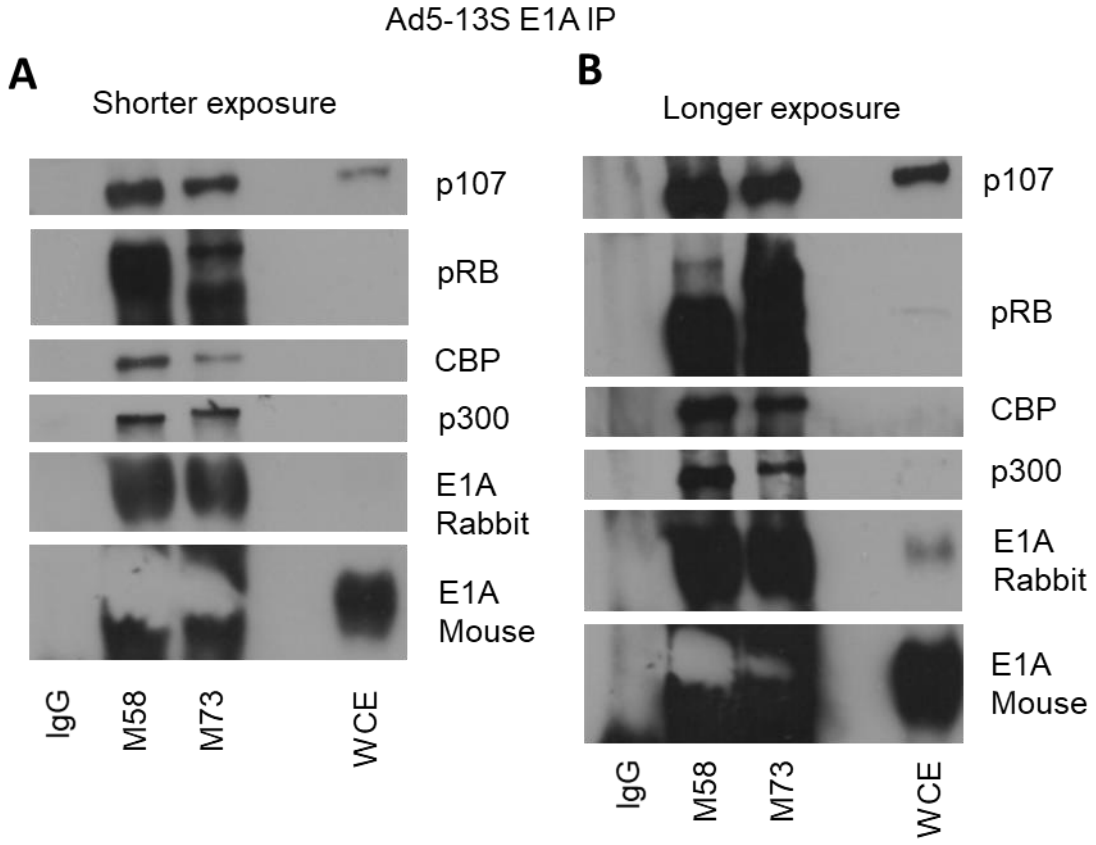

Figure 3.4. Ad5-13S E1A interaction with partner proteins. Ad5-13S E1A was immunoprecipitated using IgG, M58, and M73 antibodies. p107, pRB, CBP, p300, and E1A were observed through WB analysis; WCE, whole-cell extract. Note that the E1A signal was so strong we had negative-staining.

3.1.6. Functional validation of TET-inducible Ad12 13S E1A U2OS cell lines

To validate that the Ad12-13S E1A produced in doxycycline-treated cells was also functional we immunoprecipitated E1A from doxycycline-treated cells with anti-Ad12 E1A antibodies 160/173 or IgG (control). Following induction, cells were lysed in NETN and prepared for IP (as described section 2.2.1, 2.2.2 and 2.2.6. After IP and SDS-PAGE WB revealed that E1A was successfully immunoprecipitated specifically from cells along with known binding partners p107, pRB, p130, CBP, p300 and CtBP (Fig. 3.4B). This suggests that the Ad12 13S E1A species expressed in response to doxycycline is functional as it has the ability to interact with proteins that associate with the NTR, CR1, CR2, CR3 and CR4.

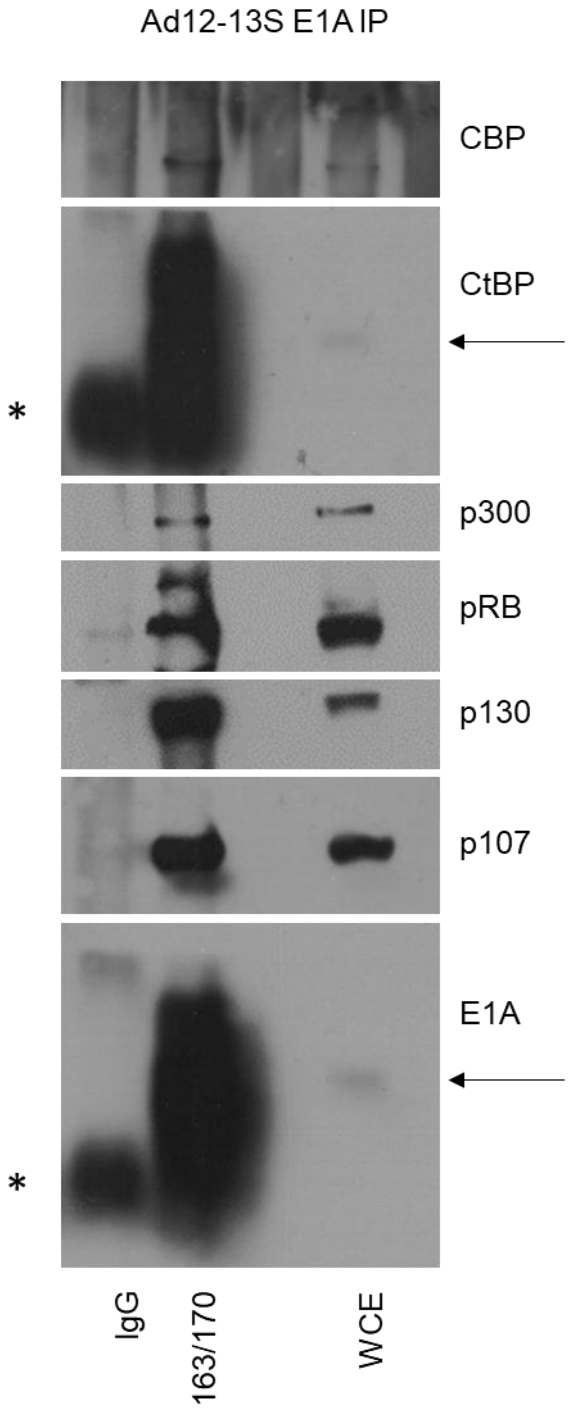

Figure 3.5. Ad12-13S E1A interaction with partner proteins. Ad12-13S E1A was immunoprecipitated using IgG and 163/170 antibodies. CBP, CtBP, p300, pRB, p130, p107, and E1A were observed through subsequent western blotting; WCE, whole-cell extract. Arrow indicates position of CtBP or E1A in WCE. *, non-specific band

3.1.7. Attempt to identify novel 13S E1A interacting proteins by IP-WB

Previous unpublished mass spectrometry studies in the laboratory identified a number of potential cellular E1A-interacting proteins (data not shown). We therefore decided to perform IPs from doxycycline-treated Ad5 and Ad12 13S E1A cell lines to see if we could confirm any novel interactions. Following induction, cells were lysed in NETN and prepared for IP (as described section 2.2.1, 2.2.2 and 2.2.6. IP-WB revealed however a lot of non-specific binding in the IgG controls as well as the anti-E1A IPs making it difficult to discern any specific binding (Figure 3.6 A and B). Thus, in summary, although we could not validate any new E1A interactors we have developed an inducible system to investigate 13S E1A function, in isolation, and the absence of expression of other viral proteins.

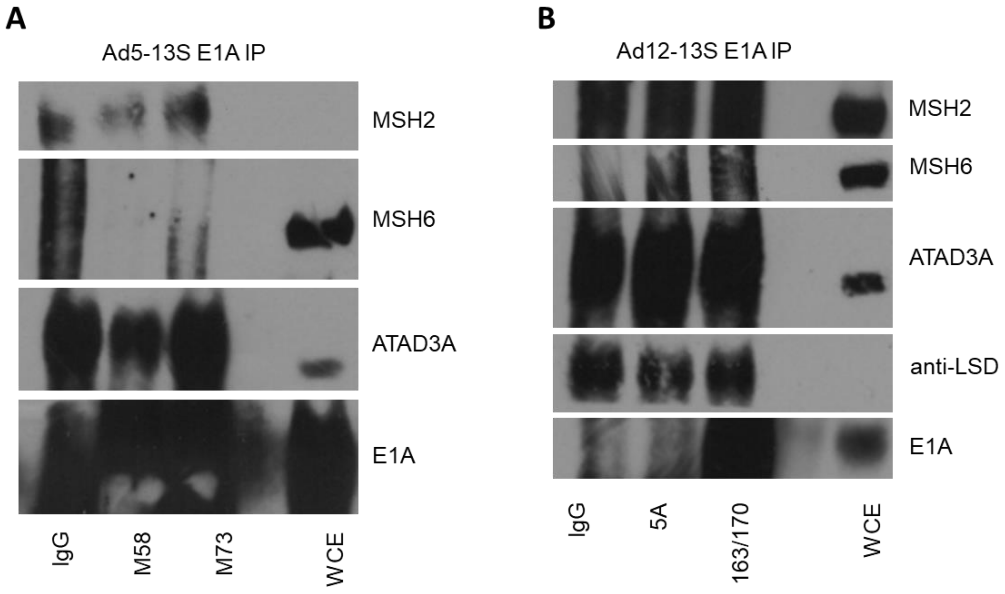

Figure 3.6. (A) Ad5-13S E1A immunoprecipitation; western blotting was unable to confirm interactions with MSH2, MSH6 or ATAD3A. (B) **Ad12-13S E1A immunoprecipitation**; western blotting was unable to confirm interactions with MSH2, MSH6, ATAD3A or LSD. Note all the non-specific bands.

3.1.8. Predictive structures of Ad2, Ad5 and Ad12 E1A CR3

Despite its relatively small size the overall structure of E1A is largely unknown, though some attempts using NMR have been made to determine the structure of portions of the smaller 12S E1A protein in complex with CBP and pRB (Ferreon et al., 2009). It has been suggested that E1A is a largely unstructured, and presumably flexible, protein that possesses Molecular Recognition Features (MoRFs) that defines protein-protein interactions. We therefore decided to use AlphaFold II, the AI platform used to predict and generate protein structures (Jumper et al., 2021; Varadi et al., 2022; see section 2.2.9) in attempt to determine predicted structures for Ad2, Ad5 and Ad12 13S E1A species. As can be seen from Table 3.1 a number of predicted structural regions and unstructured regions are proposed for each of the 3 proteins under investigation.

E1A	Missing structures	Identified structures
Ad2 13S E1A	MET 1 to THR 12 ASP 30 to THR 42 ASP 48 to ASN 58 PHE 66 to PRO 67 GLU 76 to CYS 154 GLY 165 to CYS 171 VAL 183 to CYS 268 LEU 272 to ASP 281 LYS 285 to PRO 289	GLU 13 to ALA 29 LEU 43 to TYR 47 GLU 59 to ILE 65 ASP 68 to GLN 75 ARG 155 to THR 164 SER 172 to PHE 182 ILE 269 to ASP 271 LEU 282 to CYS 284
Ad5 13S E1A	MET 1 to THR 12 ASP 30 to THR 42 ASP 48 to ASN 58 PHE 66 to PRO 67 GLU 76 to CYS 154 GLY 165 to CYS 171 VAL 183 to CYS 268 LEU 272 to ASP 281 LYS 285 to PRO 289	GLU 13 to ALA 29 LEU 43 to TYR 47 GLU 59 to ILE 65 ASP 68 to GLN 75 ARG 155 to THR 164 SER 172 to PHE 182 ILE 269 to ASP 271 LEU 282 to CYS 284
Ad12 13S E1A	MET 1 to SER 11 ASN 28 to SER 40 TYR 45 to ASN 57 PHE 65 to PRO 66 GLU 75 to HIS 101 MET 105 to MET 128 ARG 144 to CYS 159 GLY 170 to CYS 176 ILE 188 to SER 240 GLN 252 to ASP 257 LYS 261 to ASN 266	TYR 12 to PHE 27 LEU 41 to LEU 44 GLU 58 to PHE 64 GLU 67 to SER 74 PRO 102 to ASP 104 ALA 129 to GLU 143 LYS 160 to THR 169 SER 177 to PHE 187 ILE 241 to GLU 251

Table 3.1. Predicted structured and unstructured regions of Ad2, Ad5, Ad12 13S E1A species generated via ChimeraX (http://www rbvi.ucsf.edu/chimerax/; Pettersen et al., 2021; Goddard et al., 2018).

Closer inspection of the predicted structures for Ad2, Ad5 and Ad12 13S E1A species revealed perhaps not surprisingly, given their very high sequence identity (only 3 divergent amino acids), that Ad2 and Ad5 13S E1A species adopted very similar structures, with secondary structures being identical (cf. Figs. 3. 7 and 3.9; see Table 3.1). It was also clear however, that there were large portions of the protein that did not have any secondary structure, such that the overall structure was not well defined. Interestingly, the Zn-finger portions of CR3 from Ad2 and Ad5 were virtually identical, in the context of the overall protein, which is not surprising given 100% amino acid identity over this region (cf. Figs. 3.8C and 3.10C). As such the predicted Zn-finger does form a very distinct structure for interacting with an identical set of partner proteins. It is also interesting to note, in this regard, the relative positions of the 4 Cys residues at the base of the predicted Zn finger (residues 154-174) that are responsible for coordinating Zn^{2+}, which presumably enhances the structural integrity of this region of CR3.

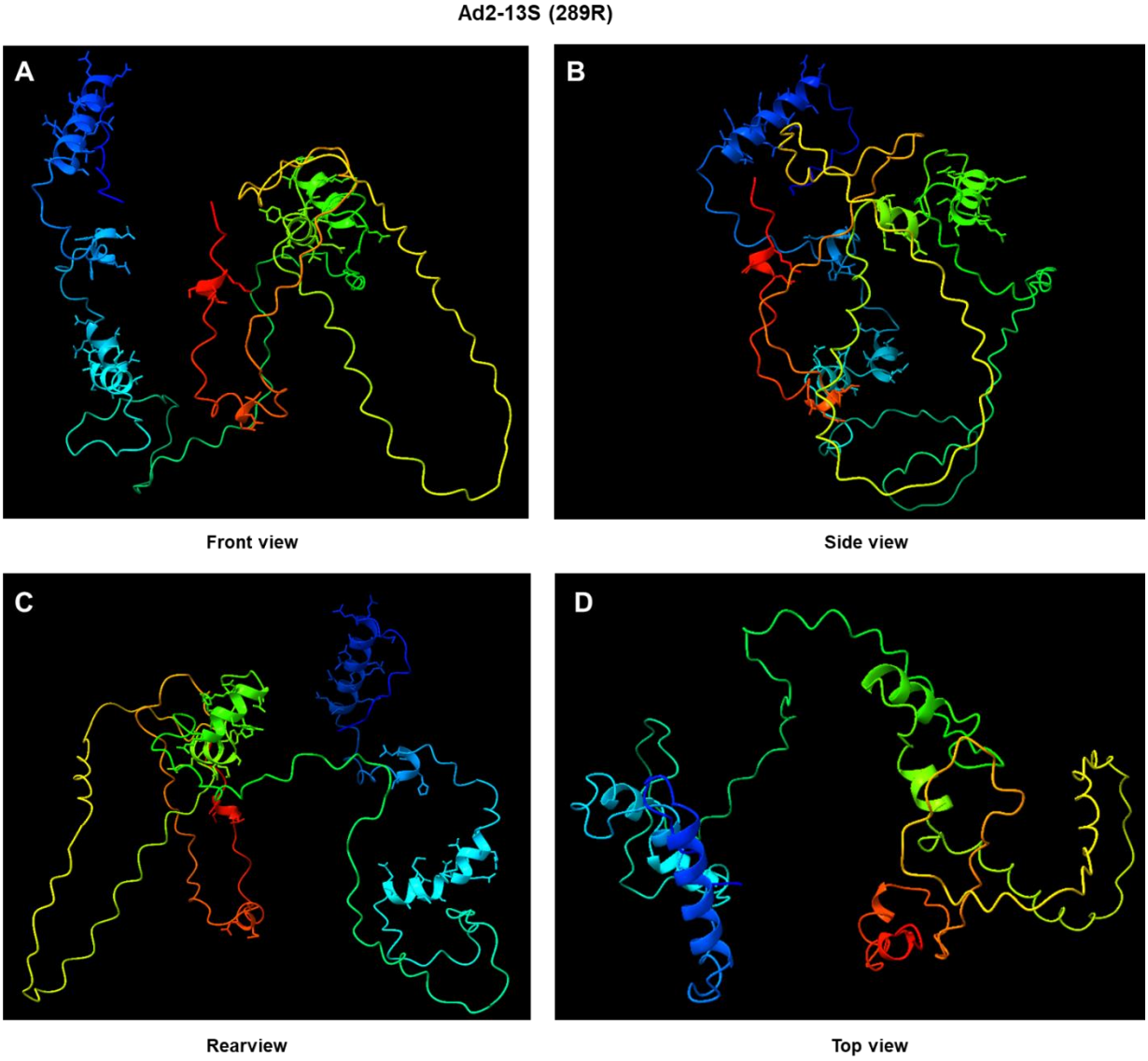

Figure 3.7. 3D predicted structures of Ad2-13S E1A (289R) generated using Colab AlphaFold 2 and ChimeraX software: (A) Front view; (B) Side view; (C) Rear view; (D) Top view. Figure generated using UCSF ChimeraX (http://www rbvi.ucsf.edu/chimerax/; Pettersen et al., 2021; Goddard et al., 2018).

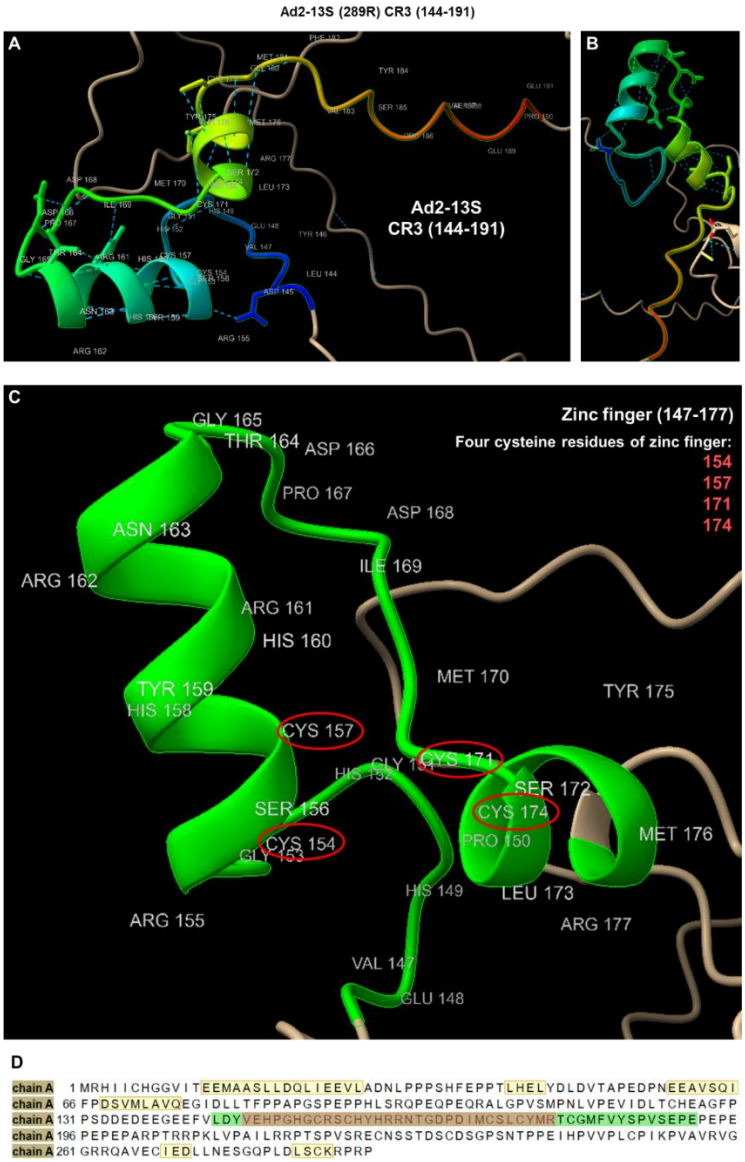

Figure 3.8. 3D predicted structures of Ad2-13S E1A (289R) CR3 (144-191) generated using Colab AlphaFold 2 and ChimeraX software: (A) Entire Ad2-13S E1A CR3 structure; (B) Side view; (C) Zinc finger (147-177) with labelled residues – four cysteines of the zinc finger are circled in red. Figure generated using UCSF ChimeraX (http://www rbvi.ucsf.edu/chimerax/; Pettersen et al., 2021; Goddard et al., 2018).

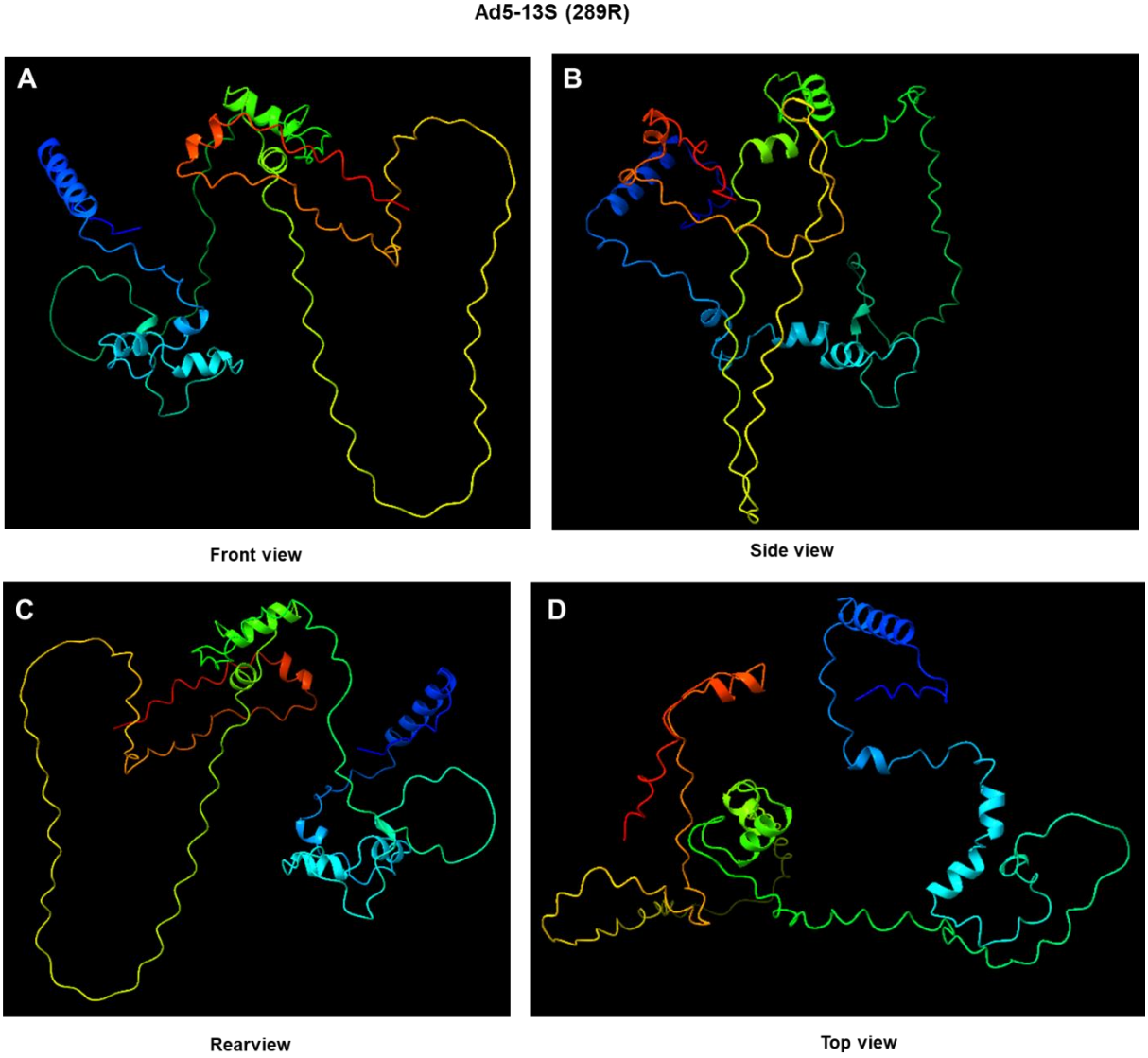

Figure 3.9. 3D predicted structures of Ad5-13S E1A (289R) generated using Colab AlphaFold 2 and ChimeraX software: (A) Front view; (B) Side view; (C) Rear view; (D) Top view. Figure generated using UCSF ChimeraX (http://www rbvi.ucsf.edu/chimerax/; Pettersen et al., 2021; Goddard et al., 2018).

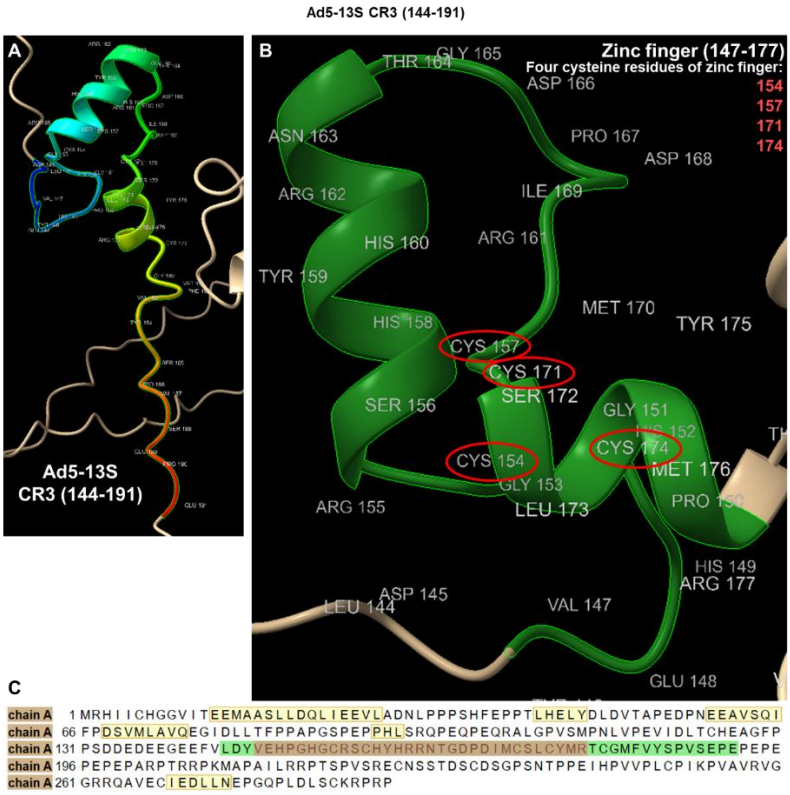

Figure 3.10. 3D predicted structures of Ad5-13S E1A (289R) CR3 (144-191) generated using Colab AlphaFold 2 and ChimeraX software: (A) Entire Ad5-13S E1A CR3 structure; (B) Zinc finger (147-177) with labelled residues – four cysteines of the zinc finger are circled in red. Figure generated using UCSF ChimeraX (http://www rbvi.ucsf.edu/chimerax/; Pettersen et al., 2021; Goddard et al., 2018).

Unlike Ad2 and Ad5-13S which consist of 289 residues, Ad12-13S consists of only 266 residues. It is clear from the analyses undertaken that Ad12 13S E1A adopts more secondary and tertiary structure when compared to Ad2 and Ad5 13S E1A, with smaller regions that are unstructured (see Table 3.1 and Fig. 3.11). Interestingly, despite much less sequence conservation in the Zn Finger region between Ad12 and Ad2/5 they adopt very similar structures with 2 α-helices connected by a short unstructured region, with 4 cysteine residues at the base of the proposed fingers, suggesting that the Zn finger of Ad12 13S E1A will bind a similar set of cellular proteins to Ad2 and Ad5 13S E1A (Fig. 3.12B).

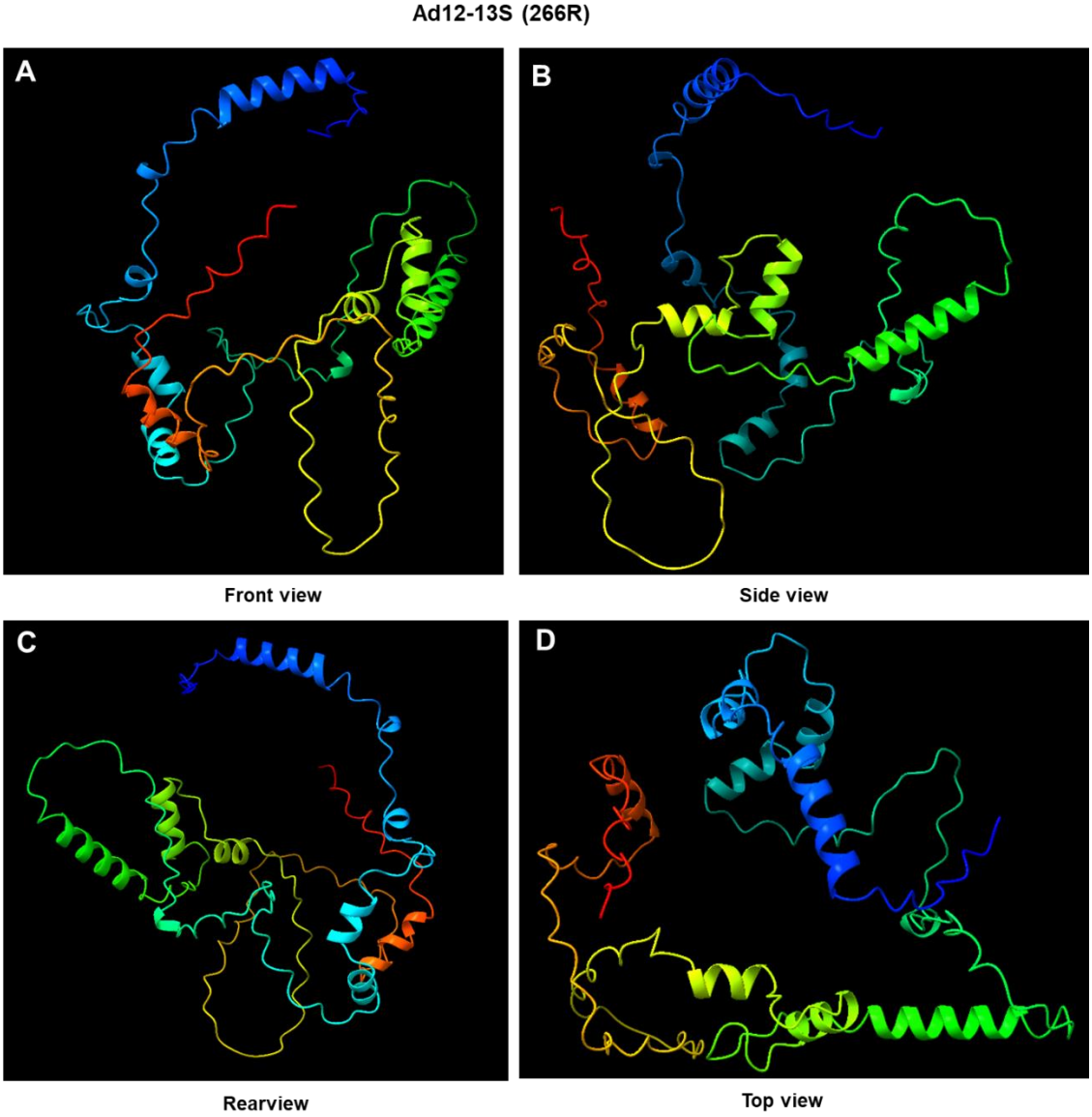

Figure 3.11. 3D predicted structures of Ad12-13S E1A (266R) generated using Colab AlphaFold 2 and ChimeraX software: (A) Front view; (B) Side view; (C) Rear view; (D) Top view. Figure generated using UCSF ChimeraX (http://www rbvi.ucsf.edu/chimerax/; Pettersen et al., 2021; Goddard et al., 2018).

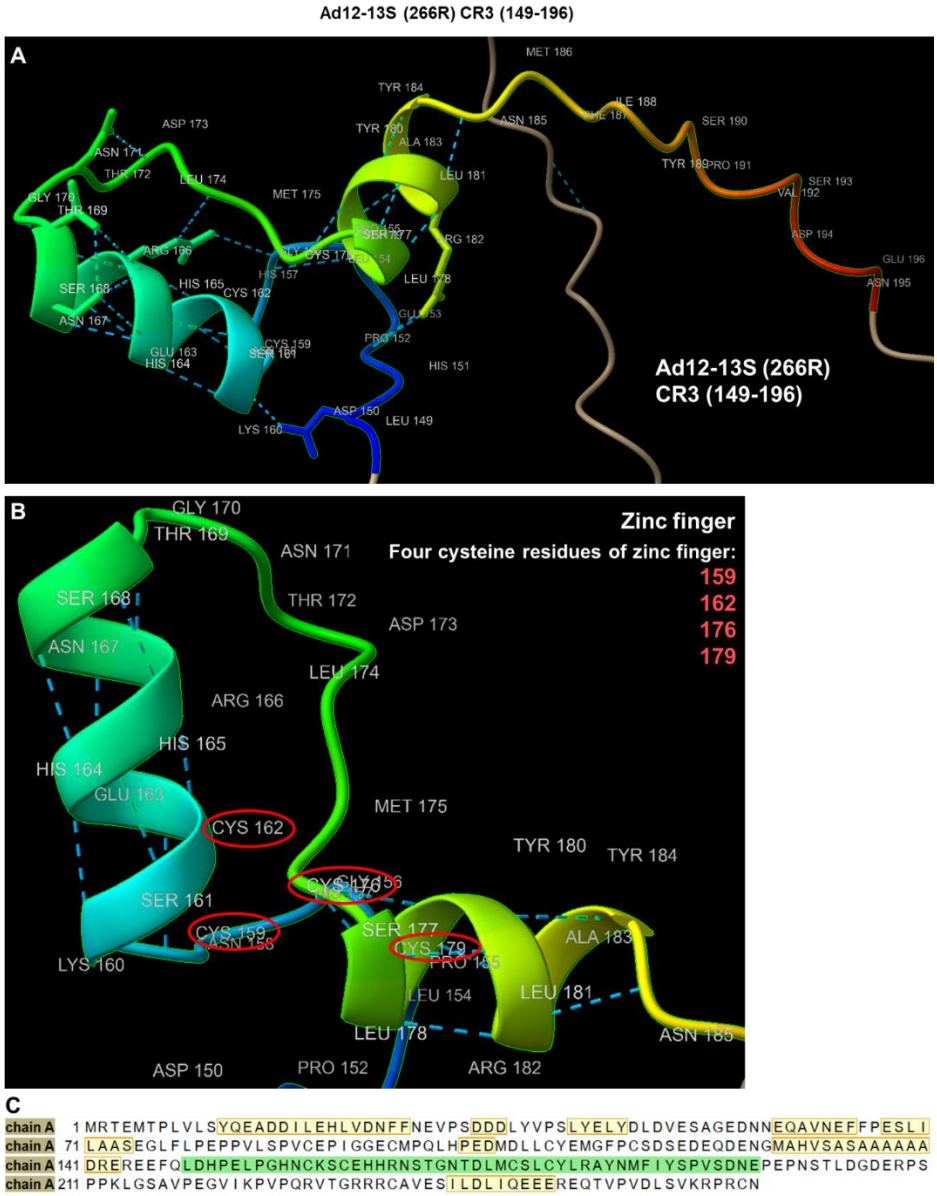

Figure 3.12. 3D predicted structures of Ad12-13S E1A (266R) CR3 (149-196) generated using Colab AlphaFold 2 and ChimeraX software: (A) Entire Ad12-13S E1A CR3 structure; (B) Zinc finger with labelled residues – four cysteines of the zinc finger are circled in red. Figure generated using UCSF ChimeraX (http://www rbvi.ucsf.edu/chimerax/; Pettersen et al., 2021; Goddard et al., 2018).

3.2. Attempt to generate clonal Ad2/Ad12 9S E1A TET-inducible U2OS cell lines

Next, we wished to study the function of the smallest E1A species, 9S, in more detail. Given the success generating Ad5 and Ad12 13S E1A TET-inducible cell lines we decided to make clonal TET-inducible U2OS cell lines that express GFP-tagged Ad2 or Ad12 9S E1A species. As we don't have access to an antibody that recognizes the 9S E1A gene product the generation

of such cell lines would help facilitate studies to identify cellular proteins that interact with 9S E1A and determine the cellular function of these viral proteins.

Initially therefore PCR was used to amplify Ad2 and Ad12 9S species that were tagged with GFP (see section 2.3.1). After PCR clean-up DNA was quantified and then digested with BamHI and XhoI to subclone into the pcDNA5/FRT/TO vector, similarly digested with these restriction enzymes. After ligation of double-digested PCR construct and the pcDNA5/FRT/TO vector, *E. coli* DH5α bacteria were transformed with the ligation mix, and plated-out onto LB-agar petri dishes containing 100μg/ml ampicillin so that single bacterial colonies could be expanded for a miniprep. After the miniprep DNA was measured, and an analytical digest performed with BamHI and XhoI, to identify by gel electrophoresis those colonies with inserted the GFP-tagged 9S E1A clone. The analytical digest determined that all clones were positive for either Ad2 9S or Ad12 9S as determined by the presence of both the upper vector band and the lower insert (9S) band on the agarose gel (Figure 3.13). Positive samples were selected for subsequent Sanger Sequencing (see section 2.3.11). Those pFRT/TO Ad 9S clones that following sequencing were identified as WT for either Ad2 or Ad12 9S were purified by maxi-prep and use to transfect FRT/TO-U2OS cells, in the presence of the recombinase in an attempt to generate TET-inducible U2OS cell lines that express GFP-tagged Ad2 or Ad9 9S E1A species (see section 3.1.3 for consideration of how transfection was performed). Despite numerous attempts we were not able to isolate any stable colonies following transfection. The possible reasons for this are considered in Chapter 4.

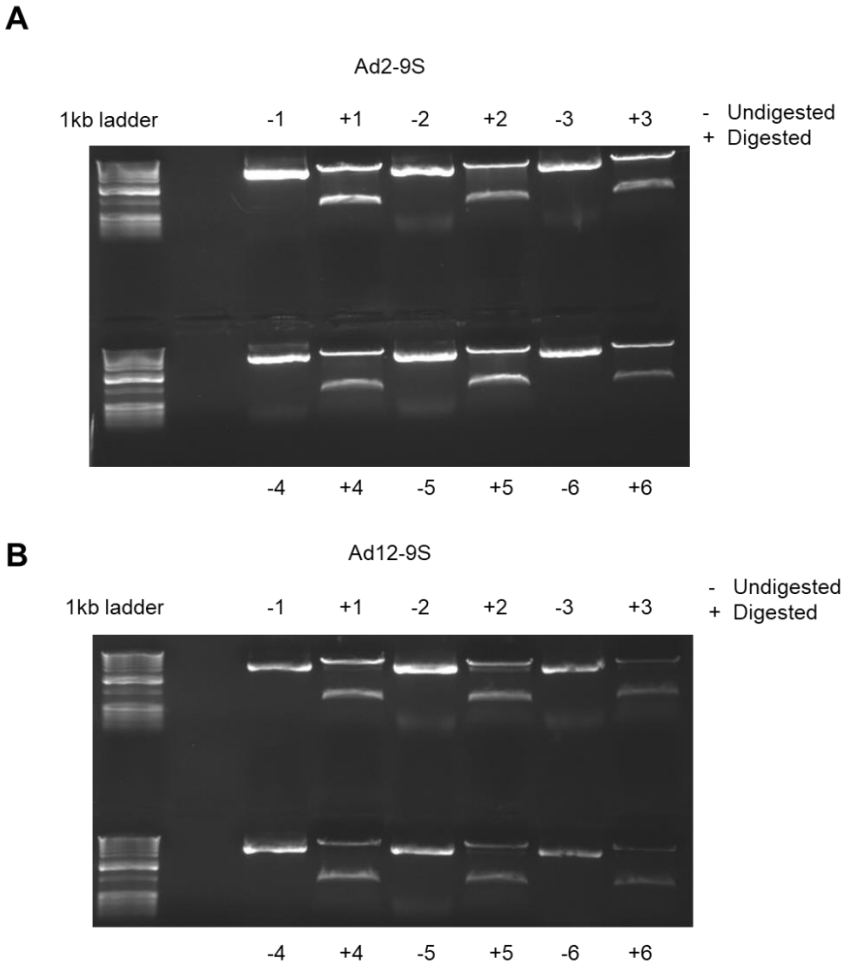

Figure 3.13. Double digest showing successful cloning of 9S E1A cDNAs into the vector. Gel electrophoresis images of (A) Ad2-9S (B) Ad12-9S after treatment with restriction endonucleases. Images were taken using Syngene Bio Imaging equipment.

3.3. Investigating the molecular functions of Ad 9S E1A

A major aim of the current study is to investigate the molecular function of the E1A 9S gene products as very little is known about their function (see section 1.6). As we were unable to generate any TET-inducible U2OS cell lines that express GFP-tagged Ad2 or Ad9 9S E1A species we decided to use previously established Ad2 and Ad12 GFP-tagged 9S constructs that could be expressed transiently in mammalian cells so that we could visualise 9S localisation in the cell and also determine the cellular interactome for Ad 9S, which might give important

insights into 9S function in the cell. In this regard it is important to note that 9S E1A species are relatively small; Ad2 and Ad5 consist of 55-residues and Ad12 consists of 53 residues.

3.3.1. Cellular localisation of Ad2 and Ad12 9S

To study the cellular localization of Ad2 and Ad12 9S gene products in the cell GFP-tagged Ad 9S species were transfected into U2OS cells (section 2.1.3). After 24h images were captured using an EVOS M5000 imaging system (ThermoFisher Scientific).

Initially, we optimised the transfection procedure for GFP-tagged Ad2 9S using PEI and Lipofectamine 2000 transfection reagents and compared expression levels by immunofluorescent microscopy, 24h post transfection. Initial observations revealed that GFP-tagged Ad2 9S species were expressed well in transfected U2OS cells in both the presence of PEI and Lipofectamine 2000 (Figure 3.14). To study the cellular localisation of GFP-tagged Ad2 9S higher magnification images were taken (Figure 3.15). These images revealed that GFP-labelled Ad2 9S was present in both the cytoplasm and nucleus. The levels of expression in these two compartments differed from cell to cell. Interestingly, however, large GFP foci were observed in both the cytoplasm and nucleus of transfected cells suggesting that Ad2 9S might localise to specific sub-compartments or distinct macromolecular complexes in the cell. As a control we utilised a stable U2OS clone that expressed GFP alone. Immunofluorescent microscopy revealed that GFP expression was equally distributed throughout the cell in both the cytoplasm and nucleus with no distinct staining pattern (Figure 3.15). Taken together, these data suggest that Ad2 9S has functions in both the nucleus and the cytoplasm.

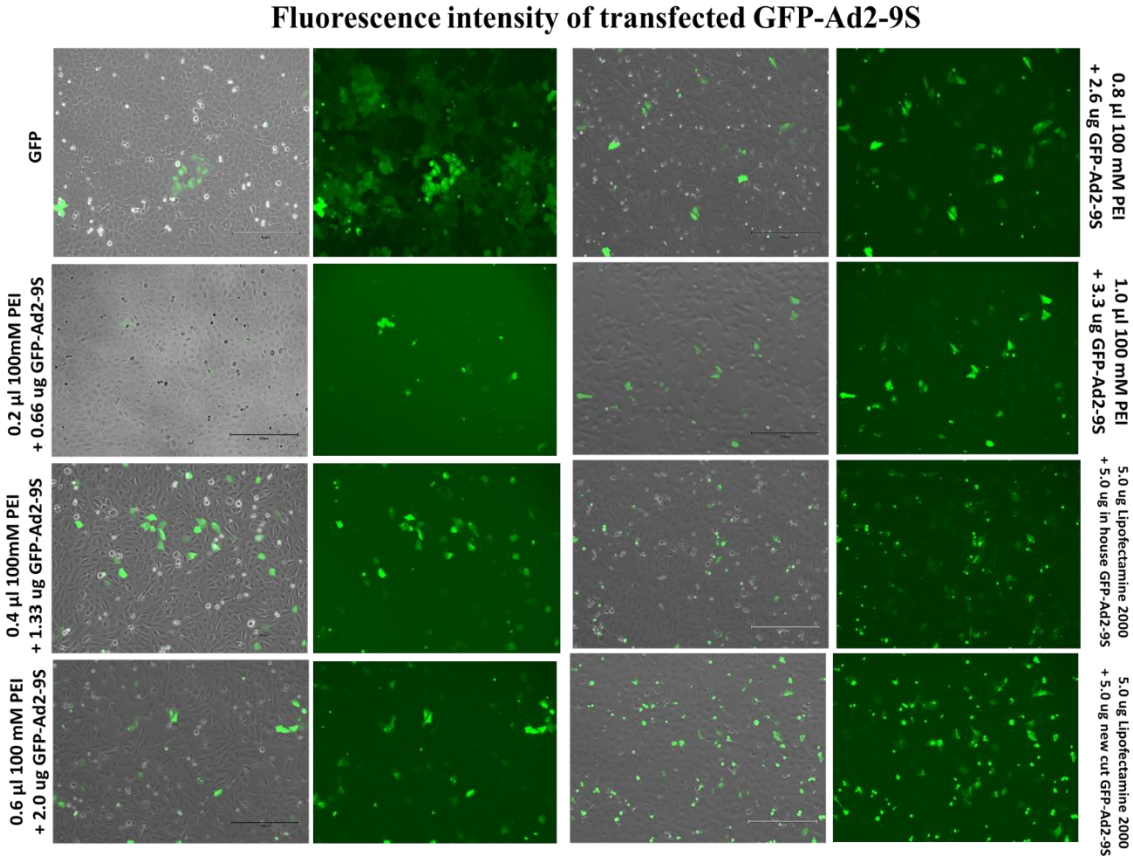

Figure 3.14. Fluorescence microscope images of U2OS cells transfected with GFP-Ad2-9S by PEI or Lipofectamine 2000. GFP-U2OS cells alone are shown for comparison. Images taken at 10X magnification.

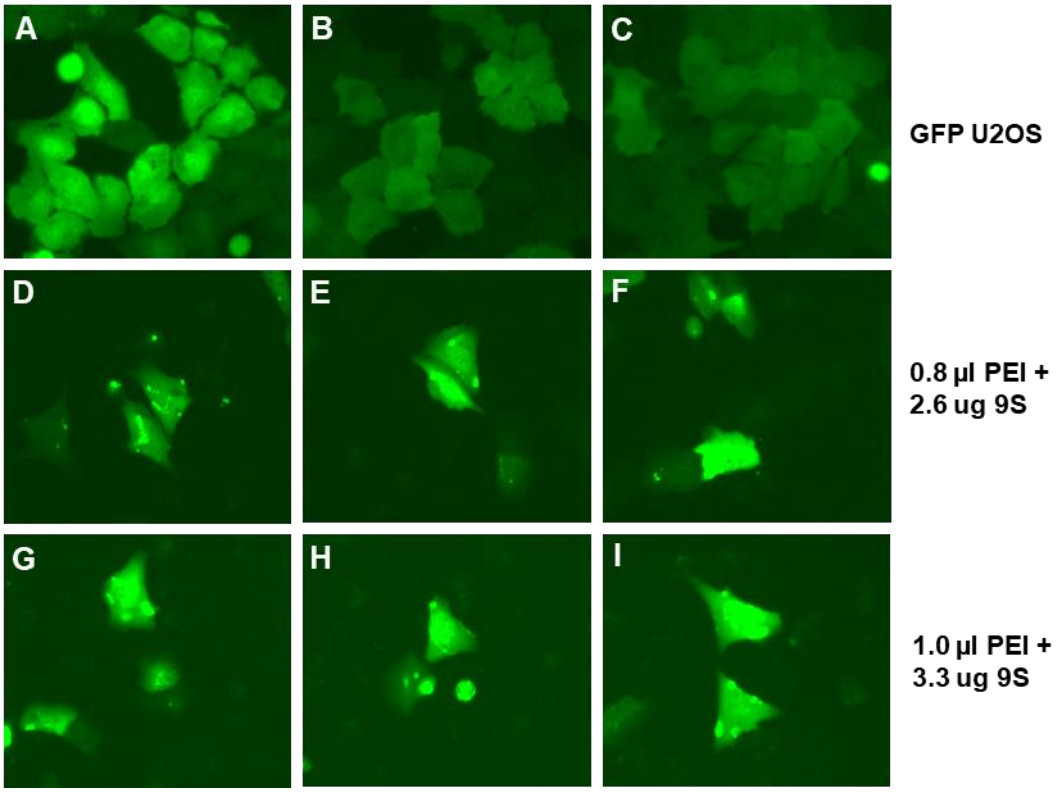

Figure 3.15. Fluorescence microscopy images of distinct images of (A-C) GFP U2OS cells and (B-I) GFP-Ad2-9S staining patterns in transfected U2OS cells. (D-F) 0.8 µl PEI + 2.6 ug 9S (G-I) 1.0 µl PEI + 3.3 ug 9S. Images B-G are enlarged images of GFP-Ad2-9S images taken at 10X magnification.

Next, we investigated the cellular localisation of GFP-tagged Ad12 9S species in the cell. This time however, we optimised transfection with PEI alone. Images revealed that GFP-tagged Ad12 9S was expressed efficiently in U2OS cells transfected with PEI (Figure 3.16). Higher magnification images of transfected cells revealed that GFP-labelled Ad12 9S species were, like Ad2, expressed in both the cytoplasm and nuclei of cells (Figure 3.17). Moreover, GFP-labelled 9S species also formed distinct foci within both the cytoplasm and nucleus suggesting discrete functions for Ad12 9S (Figure 3.17). The dose-dependent expression of the GFP-tagged Ad12 9S was demonstrated in U2OS cells by Western blot analysis, which revealed that Ad12 9S is expressed as one major species in the cell (Figure 3.18).

In summary, the heterogeneous protein expression patterns for both Ad2 and Ad12 9S species, suggests that both Ad 9S species have a number of distinct cellular functions.

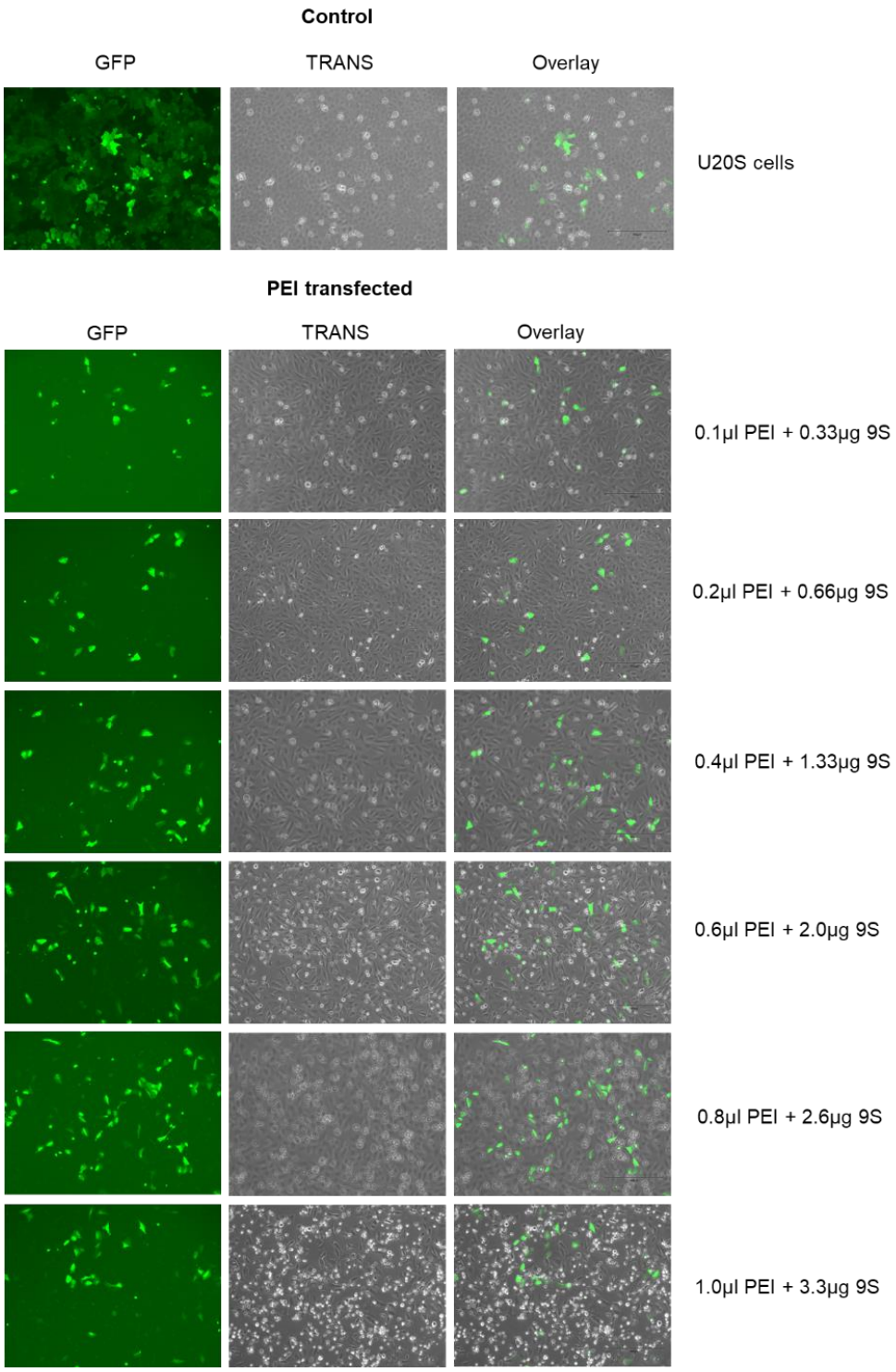

Figure 3.16. Fluorescence microscope images of GFP-Ad12-9S species in U2OS cells 24 h post-transfection with PEI. Images taken at 10X magnification.

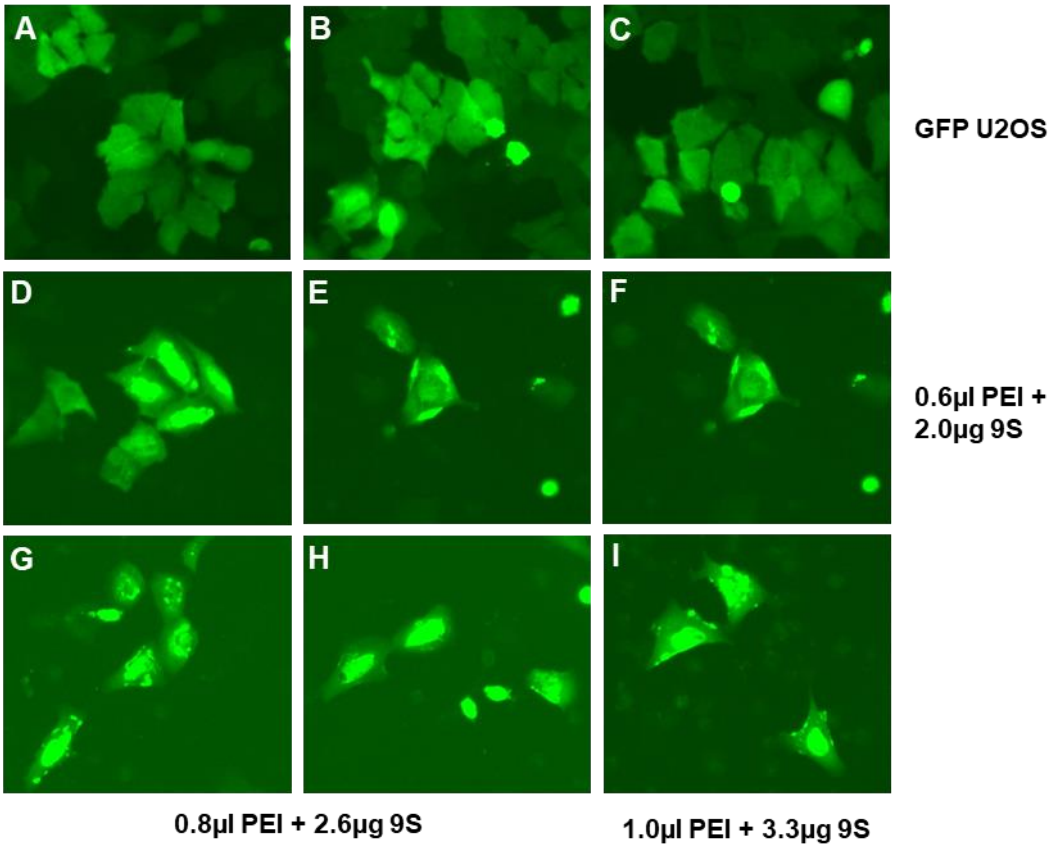

Figure 3.17. Fluorescence microscopy images of (A) GFP U2OS cells and (B-I) GFP-Ad12-9S at 24 h post-transfection with PEI. (D-F) 0.6µl PEI + 2.0µg 9S (G, H) 0.8µl PEI + 2.6µg 9S (I) 1.0µl PEI + 3.3µg 9S. Images B-G are enlarged images of GFP-Ad12-9S images taken at 10X magnification.

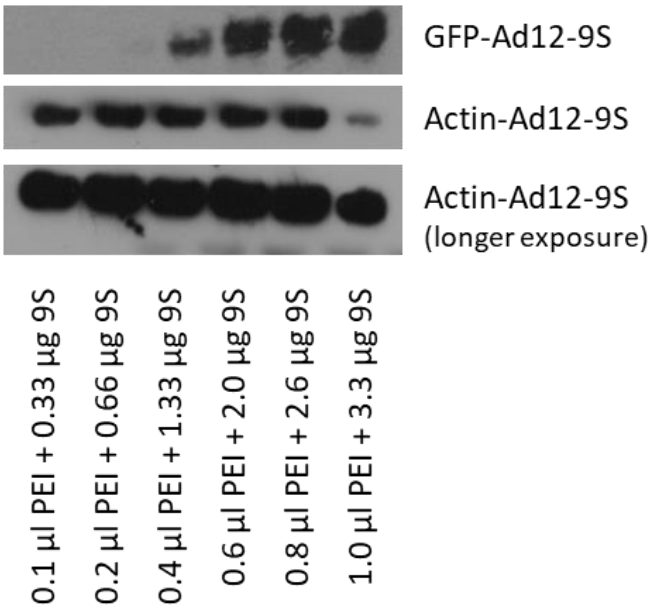

Figure 3.18. Western blot analysis of GFP-Ad12-9S expression following PEI transfection.

3.4. Investigating the Ad12 9S interactome in U2OS cells

To identify cellular proteins that interact with Ad 9S E1A, we decided to perform GFP pulldowns with U2OS cells transiently transfected with the GFP-labelled Ad12 9S E1A construct. As a control, to account for non-specific binding to GFP, we used U2OS cells that stably express GFP alone. Following GFP-pulldown (see section 2.2.7) mass spectrometry was performed (see section 2.2.8) and those proteins identified in the GFP control sample were deleted from the GFP-tagged 9S E1A list to obtain a list of proteins that were found specifically with Ad12 9S E1A (see Appendix 3.1 for the full list of putative cellular Ad12 9S-binding proteins).

The most abundant proteins detected (n=33) were input into String-db (https://string-db.org/) in order to identify known functional relationships between the cellular proteins identified. Figure 3.19 shows proteins colour-coordinated according to the following functions: Negative regulation of mitotic chromosome condensation; Positive regulation of single stranded viral RNA replication via double stranded DNA intermediate; Positive regulation of transcription by RNA polymerase III; Positive regulation of transcription by RNA polymerase I; Histone phosphorylation; Positive regulation of histone acetylation; Chromatin remodeling; DNA replication; Regulation of viral process; mRNA splicing, via spliceosome; rRNA processing; Regulation of DNA metabolic process; RNA processing. The number of proteins in the network (count), strengths (level of enrichment), and the false discovery rates were also recorded (see Appendix 3.2). These data suggest that Ad12 9S E1A primarily functions in transcriptional control and DNA metabolism. Interestingly, some of the proteins identified have previously shown to function in viral processes. It is important to note, that although a number of ribosomal proteins were identified in our mass spectrometry data, they were excluded from this STRING analysis for simplification, as a number of other ribosomal proteins were identified

in the GFP control sample, and could potentially be non-specific contaminants, though this needs to be verified by other methods.

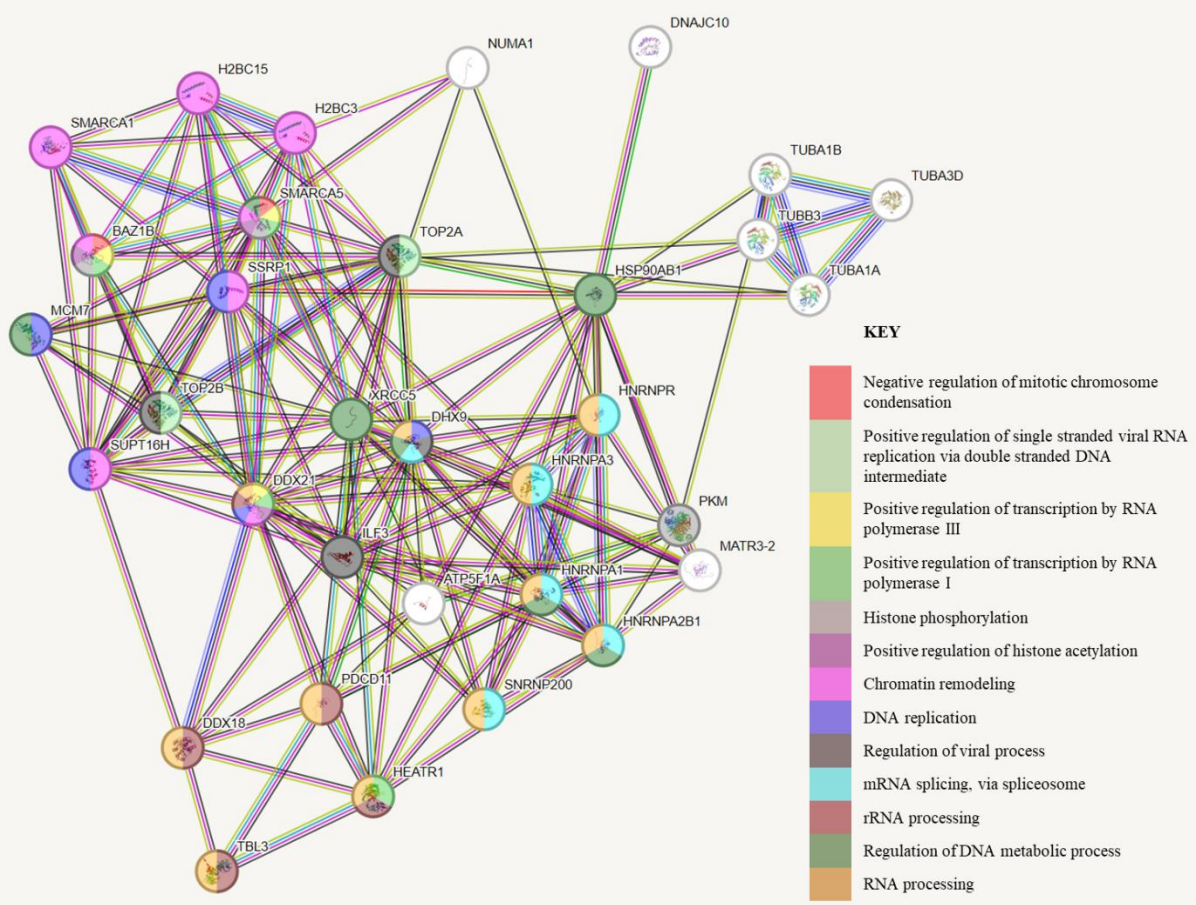

Figure 3.19. STRING analysis of most abundant proteins (n=33) identified by Mass Spectrometric Analysis that associate specifically with Ad12 9S E1A. Proteins are colour-coded according to biological process. Figure was generated using the Protein-Protein Interaction Networks programme, String-db (https://string-db.org/). See Appendix 3.2 for the list of proteins used in this STRING analysis.

Table 3.2. STRING analysis - Most abundant proteins from Mass Spectrometric Analysis that associate specifically with Ad12 9S E1A

UNIPROT_ID	Protein	MW [kDa]	No. of Peptides	Scores
TOP2A_HUMAN	DNA topoisomerase 2-alpha	174.3	55	2724.8
TOP2B_HUMAN	DNA topoisomerase 2-beta	183.2	41	1902.6
DHX9_HUMAN	ATP-dependent RNA helicase A	140.9	34	1561.6
TBB3_HUMAN	Tubulin beta-3 chain	50.4	21	1076.9
SMCA5_HUMAN	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5	121.8	20	861.1
HEAT1_HUMAN	HEAT repeat-containing protein 1	242.2	18	699.5
MATR3_HUMAN	Matrin-3	94.6	18	972.6
TBA1A_HUMAN	Tubulin alpha-1A chain	50.1	18	1277.3
TBA1B_HUMAN	Tubulin alpha-1B chain	50.1	18	1299.9
ILF3_HUMAN	Interleukin enhancer-binding factor 3	95.3	17	786.0
DJC10_HUMAN	DnaJ homolog subfamily C member 10	91	16	770.7
HS90B_HUMAN	Heat shock protein HSP 90-beta	83.2	16	944.3
NUMA1_HUMAN	Nuclear mitotic apparatus protein 1	238.1	16	580.9
RRP5_HUMAN	Protein RRP5 homolog	208.6	16	545.9
TBA3C_HUMAN	Tubulin alpha-3C/D chain	49.9	15	1061.4
DDX21_HUMAN	Nucleolar RNA helicase 2	87.3	14	529.4
HNRPR_HUMAN	Heterogeneous nuclear ribonucleoprotein R	70.9	14	593.0
MCM7_HUMAN	DNA replication licensing factor MCM7	81.3	14	727.4
ROA3_HUMAN	Heterogeneous nuclear ribonucleoprotein A3	39.6	14	856.2
SP16H_HUMAN	FACT complex subunit SPT16	119.8	14	533.4
TBL3_HUMAN	Transducin beta-like protein 3	89	14	646.5
DDX18_HUMAN	ATP-dependent RNA helicase DDX18	75.4	12	596.4
KPYM_HUMAN	Pyruvate kinase isozymes M1/M2	57.9	12	523.5
ROA1_HUMAN	Heterogeneous nuclear ribonucleoprotein A1	38.7	12	755.7
SMCA1_HUMAN	Probable global transcription activator SNF2L1	122.5	12	515.0
XRCC5_HUMAN	X-ray repair cross-complementing protein 5	82.7	12	517.8
SSRP1_HUMAN	FACT complex subunit SSRP1	81	11	464.9
ATPA_HUMAN	ATP synthase subunit alpha, mitochondrial	59.7	10	481.1
BAZ1B_HUMAN	Tyrosine-protein kinase BAZ1B	170.8	10	360.4
H2B1B_HUMAN	Histone H2B type 1-B	13.9	10	604.9
H2B1N_HUMAN	Histone H2B type 1-N	13.9	10	617.4
ROA2_HUMAN	Heterogeneous nuclear ribonucleoproteins A2/B1	37.4	10	409.6

Table 3.2: Mass spectrometric identification of cellular proteins that associate with Ad12-9S. GFP pulldowns were performed upon lysates from GFP-U2OS cells and U2OS cells transfected with a GFP-labelled Ad12 9S E1A construct. After GFP pulldown samples were processed for, and analysed by, mass spectrometry.

3.4.1. Analysis of Ad12 9S E1A binding proteins using the DAVID database.

Mass spectrometric analysis identified 278 cellular proteins that were found specifically within the Ad12 9S E1A samples. To further analyse this dataset to determine which biological pathways 9S E1A might modulate, the UNIPROT IDs (Accession IDs) of proteins identified exclusively in the 9S samples were input into DAVID Bioinformatics Resources - Laboratory of Human Retrovirology and Immunoinformatics (LHRI) (http://david.niaid.nih.gov. The Functional Annotation Tool was then used to generate gene lists according to biological processes. Those identified included: Ribosome biogenesis (n=32); mRNA splicing (n=43); mRNA processing (n=44); rRNA processing (n=27); mRNA transport (n=13); DNA replication (n=11); Translation regulation (n=11); DNA repair (n=10); Host-virus interaction (n=22); Viral RNA replication (n=2); Nonsense-mediated mRNA decay (n=3); Biological rhythms (n=6); Innate immunity (n=13). The results presented in Table 3.3 include: the number and percentages of proteins (from the 278 proteins) that are involved in known biological processes; the p-values represent the probability that the genes are associated with the GO (Gene Ontology) term – with a smaller p-value suggests that the genes are more significantly associated with the GO term and less likely due to chance. The threshold for significance <0.05 with the exception of DNA repair, nonsense-mediated mRNA decay, and biological rhythms where the cut-off was <0.1. Benjamini is the Modified Fisher Exact p-value; the lower the score, the higher the gene enrichment. Taken together, these data suggest that despite its small size Ad12 9S E1A might function in a number of biological pathways, though this clearly needs to be validated experimentally.

Table 3.3: GO analysis of Ad12 9S E1A-binding proteins. The table summarises the number of genes from our list that are associated with the listed GO terms and biological processes.

Term	Count	%	p-value	Benjamini
Ribosome biogenesis	32	11.5	2.30E-33	1.60E-31
mRNA splicing	43	15.5	5.20E-29	1.80E-27
mRNA processing	44	15.8	1.80E-25	3.40E-24
rRNA processing	27	9.7	1.90E-25	3.40E-24
mRNA transport	13	4.7	1.50E-07	2.10E-06
DNA replication	11	4	1.80E-06	2.20E-05
Translation regulation	11	4	2.80E-05	2.80E-04
DNA repair*	10	3.6	8.70E-02* (0.087)	4.70E-01
Host-virus interaction	22	7.9	1.10E-03	1.00E-02
Viral RNA replication	2	0.7	3.00E-02	2.10E-01
Nonsense-mediated mRNA decay*	3	1.1	8.50E-02* (0.085)	4.70E-01
Biological rhythms*	6	2.2	7.50E-02* (0.075)	4.70E-01
Innate immunity	13	4.7	2.30E-02	1.80E-01

3.4.2. KEGG pathway analysis of 9S-interacting proteins

We next analysed the MS interactomic data by KEGG analysis. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway maps are diagrams summarising interactions in protein networks. KEGG pathways group proteins into biological processes, drugs, and diseases. They are generated via the DAVID Bioinformatics Resources – NIAID/NIH (https://david.ncifcrf.gov). Mass spectrometry analysis following GFP pull down of Ad12-9S (see Appendix 3.1) detected 278 proteins that associated specifically with 9S. Results showed that 41 of these proteins (14.7%) were involved in the KEGG pathway for ribosome function with a p-value of 2.90E-31 and Benjamini value of 5.20E-29. The KEGG pathway for the spliceosome determined that 31 proteins functioned in this pathway (11.2%) (p=2.30E-16; Benjamini=1.40E-14), whilst The KEGG pathway for the term ribosome biogenesis in eukaryotes identified 24 9S-interacting proteins (8.6%) (p=7.70E-13; Benjamini=3.40E-11). The KEGG pathway for the term transcriptional misregulation in cancer identified 14 proteins (5%) (p=3.00E-04; Benjamini=5.90E-03), whilst the KEGG pathway for the term mRNA surveillance pathway identified 9 9S-interacting proteins (3.2%) (p=1.20E-03; Benjamini=2.20E-02). The KEGG pathway for the term nucleocytoplasmic transport also

identified 9 proteins (3.2%) (p=2.40E-03; Benjamini=3.90E-02), the KEGG pathway for the term protein processing in endoplasmic reticulum identified 11 proteins (4%) (p=4.80E-03; Benjamini=7.00E-02), and the KEGG pathway for the term DNA replication identified 5 proteins (1.8%) (p=7.50E-03; Benjamini=1.00E-01). Results were highly significant ($p<0.05$) for the terms ribosome, spliceosome, ribosome biogenesis in eukaryotes, transcriptional misregulation in cancer, mRNA surveillance pathway, and nucleocytoplasmic transport. A representative KEGG pathway, for mRNA surveillance, is shown in Figure 3.20. For graphical representation proteins identified from our mass spectrometry data are annotated with a red star on the KEGG pathway maps to show where in the pathway these proteins function. Results for 9S-interacting proteins from protein processing in endoplasmic reticulum and DNA replication were not significant, suggesting that 9S might not function in these pathways, though consideration of the DNA replication pathway suggests that 9S might target the replicative MCM helicase specifically, so some caution, and experimental validation, is needed in establishing the role of 9S E1A in these pathways. Taken together however, these data indicate that multiple proteins were identified in each of the pathways discussed suggesting that 9S E1A targets cellular pathways at multiple levels to perform its functions.

Category	Term	Count	%	p-value	Benjamini
KEGG_PATHWAY	Ribosome	41	14.7	2.90E-31	5.20E-29*
KEGG_PATHWAY	Spliceosome	31	11.2	2.30E-16	1.40E-14*
KEGG_PATHWAY	Ribosome biogenesis in eukaryotes	24	8.6	7.70E-13	3.40E-11*
KEGG_PATHWAY	Transcriptional misregulation in cancer	14	5	3.00E-04	5.90E-03*
KEGG_PATHWAY	mRNA surveillance pathway	9	3.2	1.20E-03	2.20E-02*
KEGG_PATHWAY	Nucleocytoplasmic transport	9	3.2	2.40E-03	3.90E-02*
KEGG_PATHWAY	Protein processing in endoplasmic reticulum	11	4	4.80E-03	7.00E-02
KEGG_PATHWAY	DNA replication	5	1.8	7.50E-03	1.00E-01

Table 3.4. KEGG Pathway Analysis using data from Mass Spectrometric analysis of GFP-Ad12-9S-interacting proteins. The UNIPROT IDs (Accession IDs) were input into DAVID Bioinformatics Resources - Laboratory of Human Retrovirology and Immunoinformatics (LHRI) (http://david niaid.nih.gov). Proteins from the 9S interactomic list are annotated with a red star.

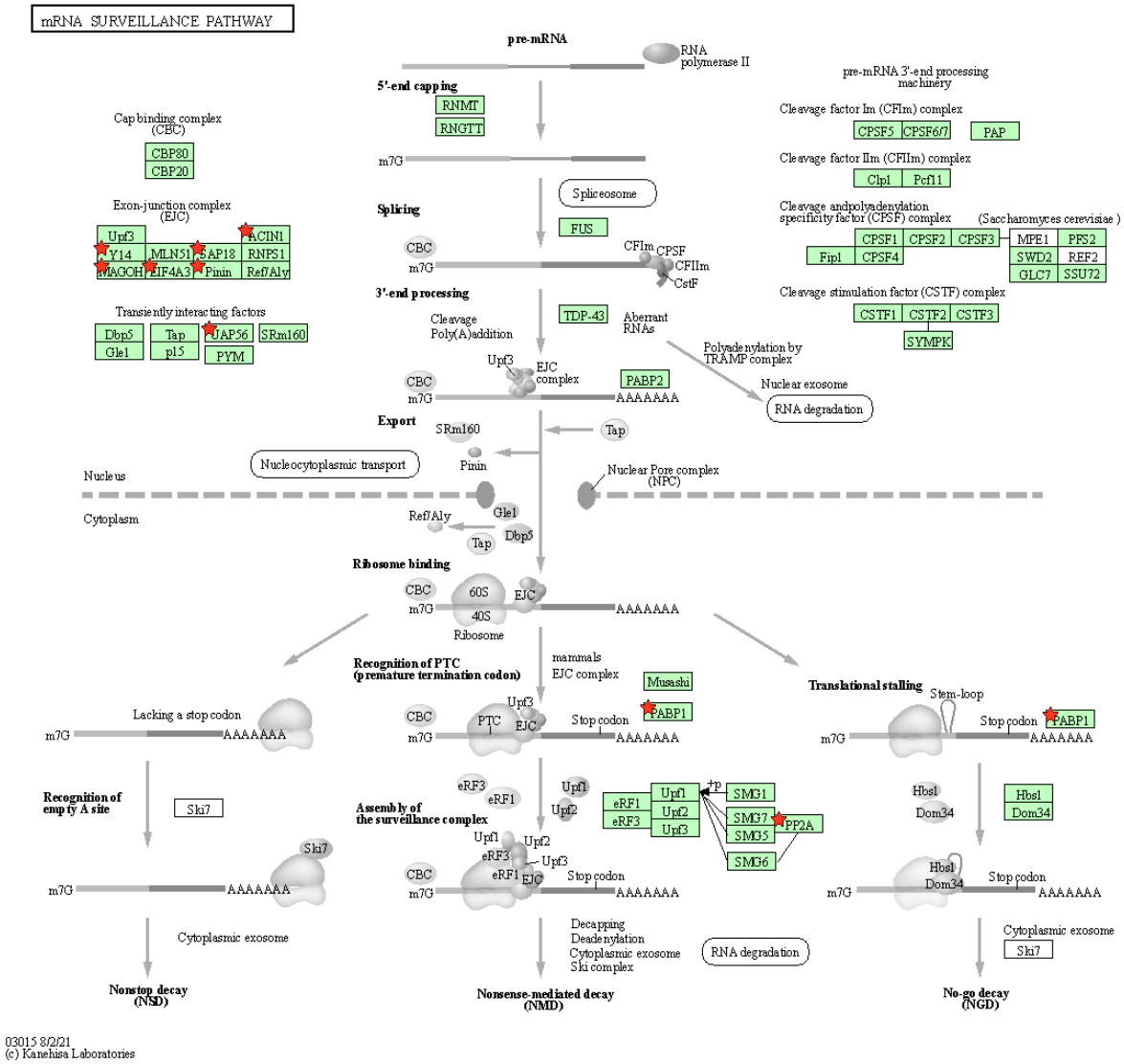

Figure 3.20. KEGG Pathway: mRNA surveillance pathway (n=9). Proteins identified from our mass spectrometry data are annotated with a red star.

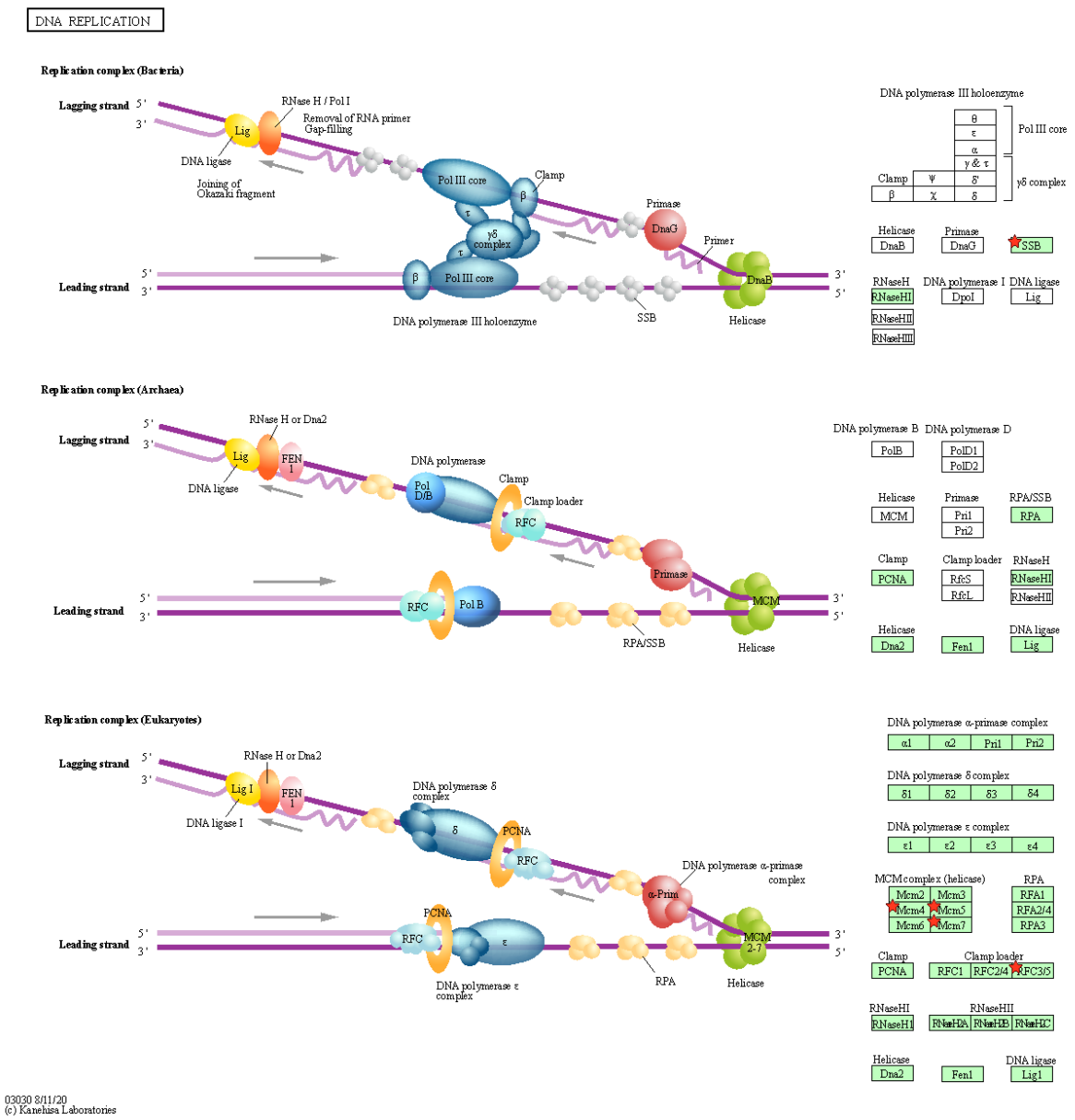

Figure 3.21. KEGG Pathway: DNA replication (n=5). Proteins identified from our mass spectrometry data are annotated with a red star.

3.4.3. Predictive structures of Ad2, Ad5 and Ad12 9S E1A structures

Given the vast number of proteins identified by mass spectrometry that associate specifically with 9S, and that there are no protein structures for 9S E1A, AI software was again utilised to generate predictive structures of 9S E1A for Ad2, Ad5 and Ad12. All 3 proteins were shown

to have high proportions of predicted secondary and tertiary structures, though despite their short size and similarity adopted 3 distinct structures (see Figs. 3.22-3.24).

The 55 residue Ad2-9S protein is proposed to consist of three alpha helices. The first alpha helix is proposed to run from residue Glu-13 to Ser-33, which is followed by a short, unstructured linker that is followed by a second alpha helix from Asn-39 to Asp-44. Following a one amino acid break, a third alpha helix spans Leu-45 to Leu-54 (Fig 3.22). Thus, Ad2 9S E1A is proposed to form a distinctive finger-like structure, with a short unstructured N-terminal region.

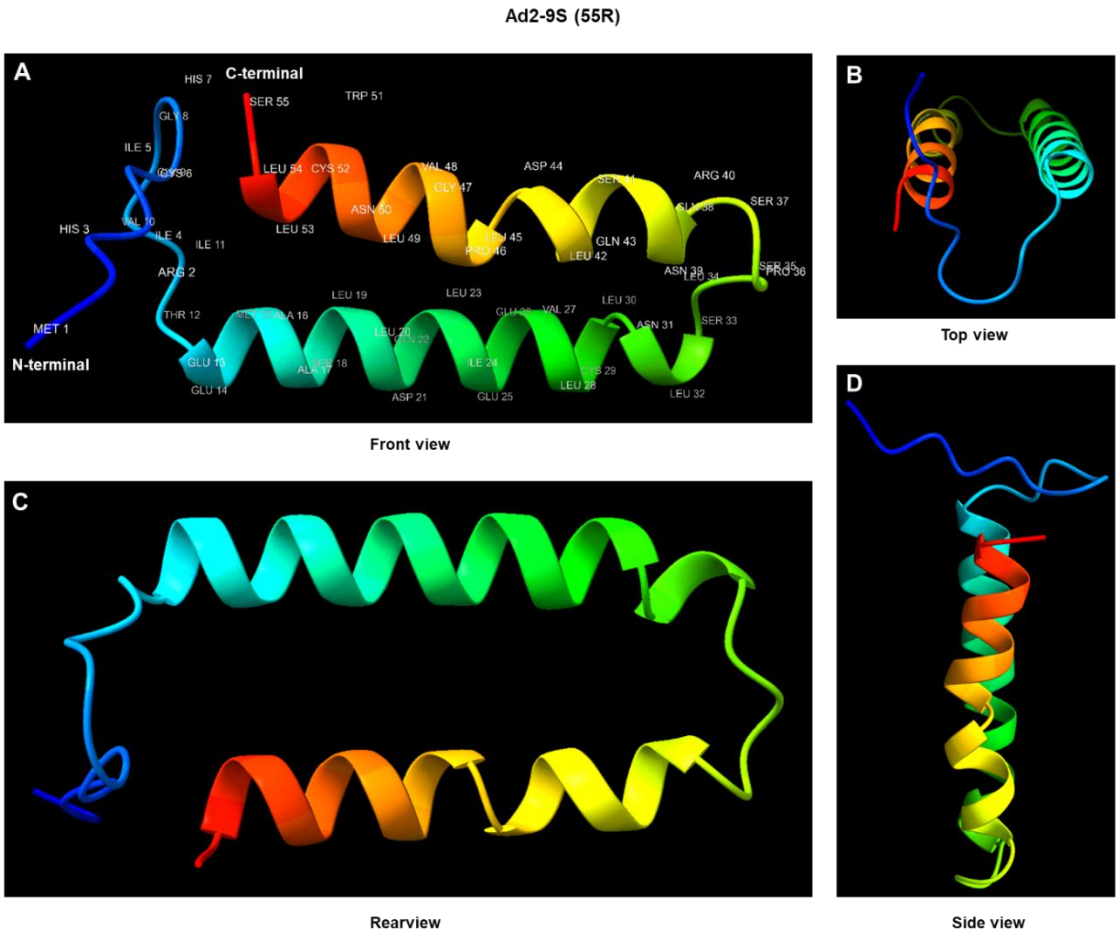

Figure 3.22. 3D predicted structures of Ad2-9S E1A (55R) generated using Colab AlphaFold 2 and ChimeraX software: (A) Front view with N-terminal and C-terminal labelled; (B) Top view; (C) Rear view; (D) Side view. UCSF ChimeraX (http://www.rbvi.ucsf.edu/chimerax/; Pettersen et al., 2021; Goddard et al., 2018).

Ad5-9S E1A (55R) differs from Ad2-9S E1A by 3 residues; however despite the high overall similarity the proposed structure is distinct from Ad2 9S E1A (cf. Figs. 3.22 and 3.23). The N-terminal region again is proposed to be unstructured, which is followed by an alpha-helix from Glu-13 to Leu-32. Following a large unstructured region, the second alpha helix spans Leu-42 to Leu-54. Overall, Ad5 9S E1A consists of 2 alpha helices that lie perpendicular to one another, and are separated by an unstructured loop (Fig. 3.23).

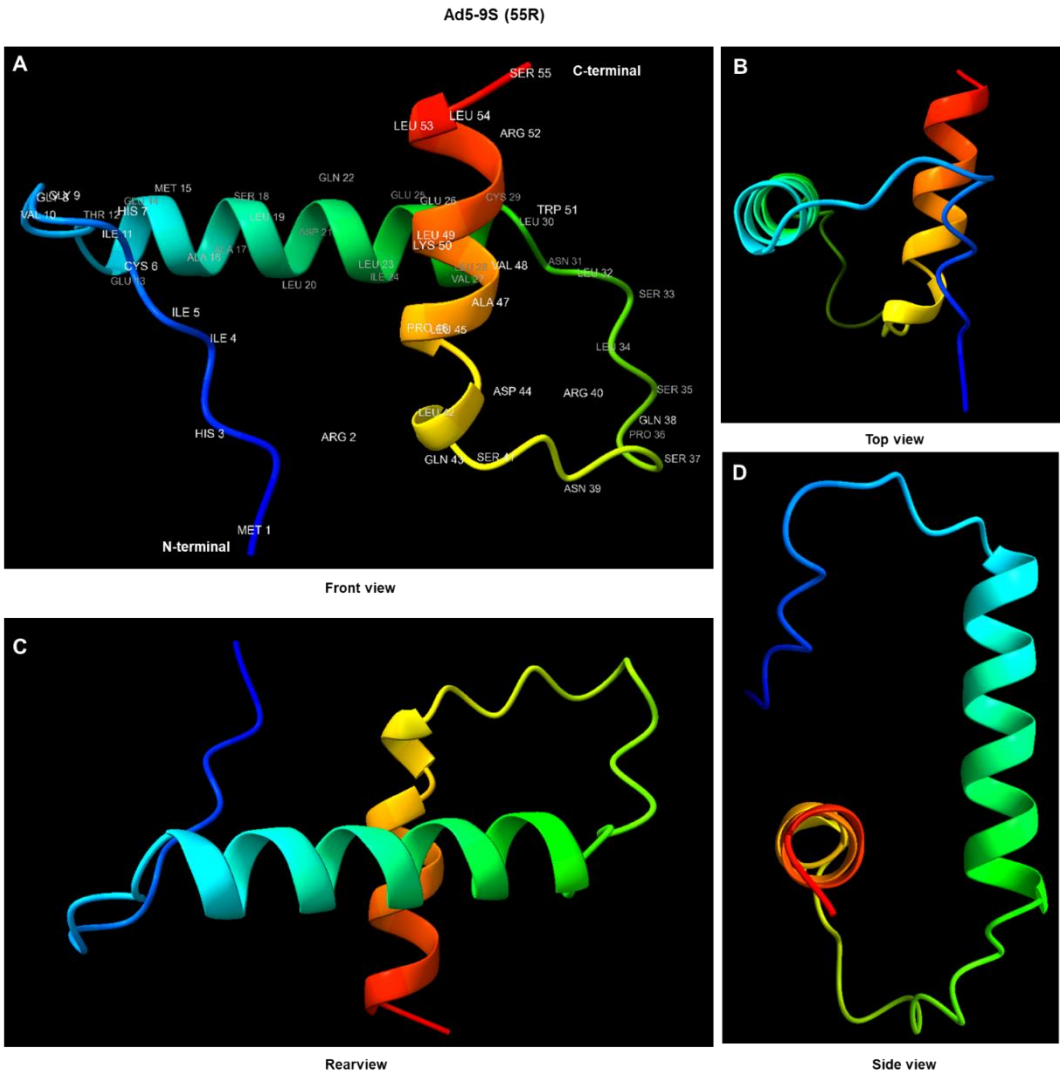

Figure 3.23. 3D predicted structures of Ad5-9S E1A (55R) generated using Colab AlphaFold 2 and ChimeraX software: (A) Front view with N-terminal and C-terminal labelled; (B) Top view; (C) Rear view; (D) Side view. UCSF ChimeraX (http://www.rbvi.ucsf.edu/chimerax/; Pettersen et al., 2021; Goddard et al., 2018).

We next considered the structure of the Ad12 9S E1A species, which we used for the interactomic studies described in section 3.4. Ad12-9S E1A is highly divergent in primary sequence compared to Ad2 and Ad5 9S E1A species. Ad12 9S E1A was proposed to be highly structured following a short unstructured N-terminal region (Fig. 3.24), with two successive tandemly-arranged alpha helices from residue Tyr-12 to Val-30, and Leu-31 to Ser-47 (Fig. 3.24).

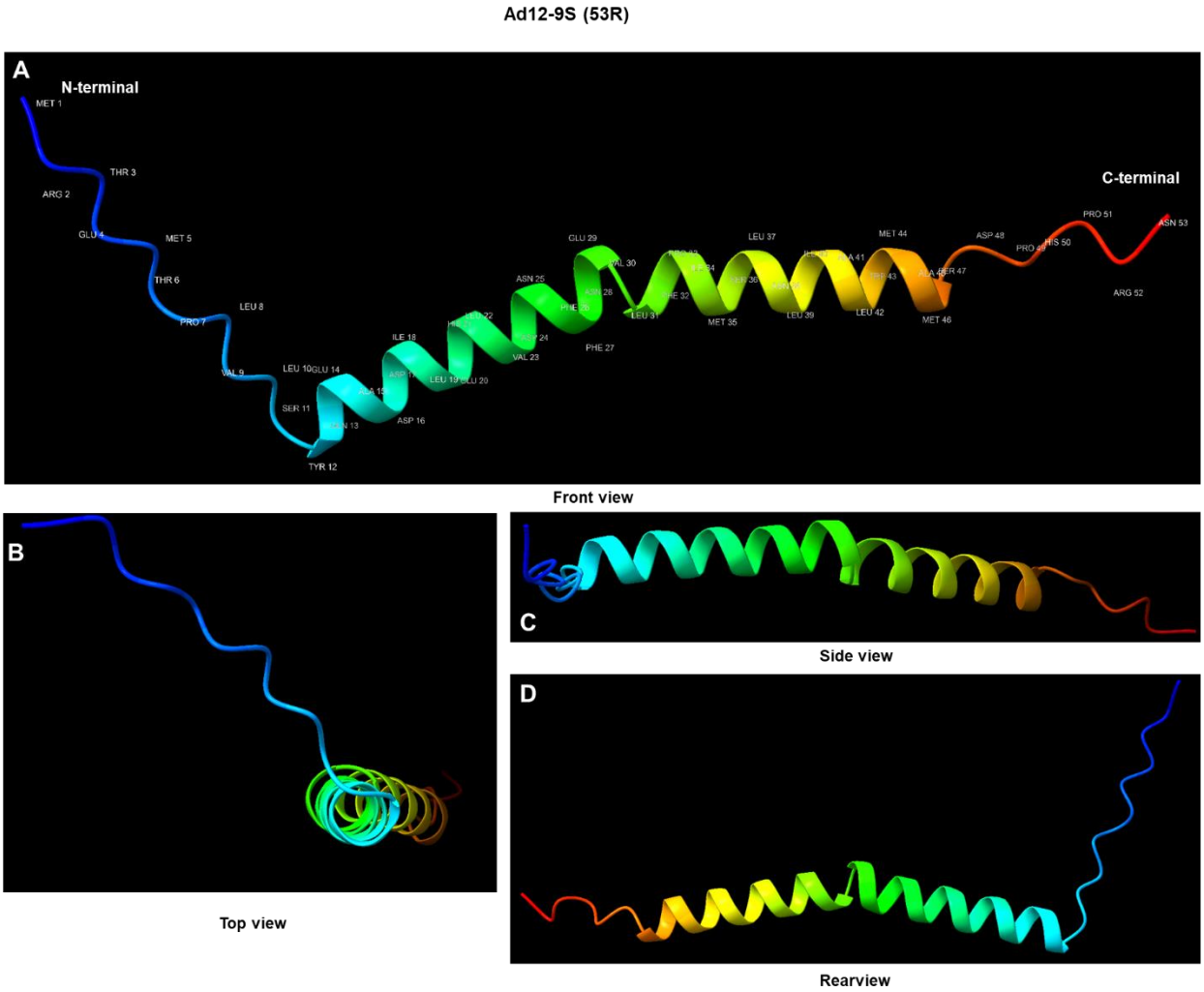

Figure 3.24. 3D predicted structures of Ad12-9S E1A (53R) generated using Colab AlphaFold 2 and ChimeraX software: (A) Front view with N-terminal and C-terminal labelled; (B) Top view; (C) Side view; (D) Rear view. UCSF ChimeraX (http://www.rbvi.ucsf.edu/chimerax/; Pettersen et al., 2021; Goddard et al., 2018).

Given the suggested differences between these 3 9S E1A species it might be suspected that these proteins target distinct sets of cellular proteins to perform its functions. This will be considered in more detail in the discussion.

Chapter 4

Discussion

4. Discussion

4.1. Adenovirus 13S E1A structure and function

This study focussed specifically on investigating the structure and function of adenovirus type 2, 5, and 12 E1A 13S and 9S spliced variants. For Ad 13S E1A we generated and validated clonal TET-inducible Ad5 and Ad12 13S E1A U2OS cell lines, as well Ad5 13S E1A-L1920A and Ad12 13S E1A-RG2 U2OS cell lines (Fig. 3.3). Interestingly, we did not isolate many 13S E1A inducible cell lines compared with other genes of interest studied in the laboratory. Indeed, this type of experiment typically generates 100+ clones but for Ad 13S E1A we only recovered 5 or 6 clones from each condition. We suggest that this is because E1A expression, in isolation, induces apoptosis (Lowe and Ruley, 1993), and leaky expression of E1A in this experiment (due to low concentrations of tetracycline in the growth medium limited the number of clones isolated. However, with the clones generated we were able to study the function of Ad5 and Ad12 CR3 in the absence of viral infection and 12S E1A expression. Indeed, IP-WB analyses revealed that E1A, when induced in these cell lines, was able to interact specifically with known binding partners p107, pRB, CBP, p300 and CtBP (Figs. 3.4 and 3.5), as seen in other studies with E1A (e.g. Dorsman et al., 1997). The paper by Dorsman and colleagues also observed that although different E1A species bind to identical cellular proteins, Ad5 E1A proteins associate with p300 and AP1/ATF family proteins in a different manner to that of Ad12 E1A proteins; this could give rise to the differences in transforming oncogenic properties of Ad5 E1A and Ad12 E1A proteins. We also attempted to use these cell lines to validate novel E1A interactions identified by IP-coupled to mass spectrometry. However, we observed a lot of non-specific binding in the IgG controls as well as the anti-E1A IPs making it difficult to discern any specific binding (Fig. 3.6 A and B). Thus, in summary, although we could not validate any new E1A interactors we have developed an inducible system to investigate 13S

E1A function, in isolation, and the absence of expression of other viral proteins, which could be useful to investigate novel E1A interacting proteins.

Ad 13S E1A is mainly involved in the transactivation of viral early promoters during infection (see section 1.5). In this regard, 13S E1A induces the CBP/p300-dependent acetylation of histone H3 K18 and K27 associated with early viral promoters to induce transcription (Hsu et al., 2018). It would therefore be interesting to see if E1A induction in the inducible cell lines generated similarly induced histone acetylation in the absence of viral infection, or other viral proteins. Indeed, the ability of 13S E1A to transactivate cellular promoters, either during infection or in the absence of infection, is not well studied. Thus, it would be interesting to investigate by RNA-seq whether induction of either Ad5 or Ad12 13S E1A promoted the expression of cellular mRNAs, which would be validated at the gene level by qRT-PCR. This study could also be extended by performing ChIP-Seq to identify cellular promoters to which E1A associates. Genome association studies with 13S E1A could also be related to histone modifications such as those studied by Hsu et al. Additionally, this could be extended to study whether histone methylation is also regulated at promoters targeted by 13S E1A. In this context, we could study the role of specific CR3-binding proteins in this capacity by using CR3 mutants that possess differential binding capacity for CR3-interacting proteins.

AlphaFold has become a powerful technique in which to compare AI-generated structures with experimentally determined structures, as well as studying the proposed structure of proteins for which no physical structure has been determined. In this regard a physical structure for any E1A proteins has been determined. As such, we used AlphaFold II to generate structures for Ad2, Ad5 and Ad12 13S E1A proteins. As anticipated E1A species had limited regions of secondary and tertiary structure and large regions that were proposed to be unstructured. Interestingly however, the Zn finger transactivation domain of CR3 from all 3 Ad types had very similar, proposed, structures for the Zn finger region suggesting that these proteins bind

to a similar, if not identical set of cellular proteins. In this regard, it will be interesting to interrogate the structure of CR3 by determining how mutation might affect structure and protein-protein interactions. Although not considered here, AI platforms can be used to accurately identify and interrogate sites of protein-protein interactions (e.g. Krapp et al., 2023).

4.2. Adenovirus 9S E1A structure and function

In a similar manner to which we studied 13S E1A we also attempted to make clonal TET-inducible Ad2 and Ad12 9S E1A U2OS cell lines. Despite purportedly cloning the construct successfully into the pcDNA5/FRT/TO expression vector we were unable to generate any stable clones (section 3.2). Assuming that the integrity of the construct was correct; indeed transient transfection with this plasmid produced a food GFP signal that resembled staining patterns seen with GFP-tagged 9S seen in Fig 3.16 (data not shown) it is likely that the failure to rescue any colonies relates to leaky expression of Ad 9S during clonal selection and potential effects of 9S E1A expression on cell growth and death pathways. In the future it would be wise to repeat the cloning steps and sequencing validation and perform the clonal selection in serum that has been stripped of tetracycline to ensure that there is no leaky expression of 9S.

Generation of clonal TET-inducible Ad2 and Ad12 GFP-tagged 9S E1A cell lines would greatly facilitate the interactomic GFP pulldown-Western Blot validation assays and studies investigating 9S E1A function. Currently we rely on transient transfections to express GFP-tagged 9S E1A species which is time-consuming and less efficient in that not all cells express the construct. It would be important however, to complement these studies with conventional immunoprecipitation-Western blot and confocal immunofluorescence studies from Ad2, Ad5 or Ad12-infected cells. Given the apparent differences in structure of Ad2, Ad5 and Ad12 9S

E1A species (Figs. 3.8, 3.10 and 3.12) this might necessitate the making of antibodies to all 9S E1A species, rather than to raise an antibody to a region of 100% identity, as structural restraints might limit the accessibility of such an antibody for procedures such as immunoprecipitation and immunofluorescence.

GFP pulldown coupled to mass spectrometric analysis following identified a number of potential cellular Ad12 9S E1A interacting proteins (Table 3.2). Bioinformatic analyses suggested that Ad12 9S E1A functions to modulate a diverse number of pathways (Table 3.3). It will be important in the first instance, therefore, to validate some of these interactions as indicated above. To do this, it would be advisable to perform GFP pulldowns then Western Blot analysis for proteins identified as top hits (i.e. those proteins which had the greatest number of peptides identified) and proteins that function in the same pathways. Ideally, it would be good to identify positive hits from different biological pathways and then perform functional assays to determine the consequence of 9S E1A on these pathways. Given that 9S E1A has previously been implicated in the regulation of transcription (Miller et al., 2012) it is perhaps not surprising that a number of the top hits are involved in DNA regulation or transcriptional control. However, previous studies with 9S E1A were all shown to be reliant on the N-terminal region that is shared with 12S and 13S E1A species it would be important to try and discern the contribution of the unique C-terminal region in protein-protein interactions and the functional regulation of biological pathways. As larger 12S and 13S E1A species also modulate transcription programmes it would be interesting to perform RNA-seq to see the global effects of 9S E1A expression on transcription, which as previously discussed for 13S E1A could be extended to study association with cellular promoters and effects on epigenetic modifications. As we performed the original GFP pulldown-mass spectrometric analyses with Ad12 9S E1A it will also be important to see if these interactions are functionally conserved for Ad2 and Ad5 9S E1A.

If cellular proteins are identified positively as 9S E1A-binding partners, then we would use AlphaFold to try and identify potential binding sites that could inform mutational analyses to identify loss of binding mutants. Using such mutants, it might be possible to dissect the contribution of different cellular proteins to 9S E1A function in different biological pathways. Given the extremely small size of 9S E1A species they might be amenable to conventional structural analysis by NMR or cryo-EM. As such, it would be interesting to validate structures proposed by AlphaFold.

Given the lack of information concerning the function of 9S E1A these follow-up studies to the work presented here would define new, important roles for 9S E1A that might be important during the later stages of adenovirus infection. Moreover, although the studies here have expanded our knowledge of Ad 9S E1A there is certainly more work to do to understand how 9S E1A functions at the molecular level to functionally regulate biological pathways.

4.3. Conclusions

The study presented here has been important towards understanding the functions of different Ad2, Ad5 and Ad12 E1A species and attempting to relate function to protein structure. These studies have been important towards elucidating a highly conserved structure for the Zn finger region of Ad2, Ad5 and Ad12 13S E1A, which might help inform future studies on the function of this region of 13S E1A. Moreover, the structural insights into Ad2, Ad5 and Ad12 9S E1A species and the determination of the Ad12 9S E1A interactome will inform future studies to determine the molecular and biological functions of this viral protein.

5. References

AlphaFold (https://www.alphafold.ebi.ac.uk/).

Arulsundaram, V.D., Webb, P., Yousef, A.F., et al. (2014) The adenovirus 55 residue E1A protein is a transcriptional activator and binds the unliganded thyroid hormone receptor. *The Journal of general virology*, 95 (Pt 1): 142–152. doi:10.1099/VIR.0.056838-0.

Berk, A.J. and Sharp, P.A. (1977) Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. *Cell*, 12 (3): 721–732. doi:10.1016/0092-8674(77)90272-0.

Boyd, J.M., Loewenstein, P.M., Tang, Q., et al. (2002) Adenovirus E1A N-terminal amino acid sequence requirements for repression of transcription in vitro and in vivo correlate with those required for E1A interference with TBP-TATA complex formation. *Journal of virology*, 76 (3): 1461–1474. doi:10.1128/JVI.76.3.1461-1474.2002.

Chang, Y., Moore, P.S. and Weiss, R.A. (2017) Human oncogenic viruses: nature and discovery. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372 (1732). doi:10.1098/RSTB.2016.0264.

Chow, L.T., Broker, T.R. and Lewis, J.B. (1979) Complex splicing patterns of RNAs from the early regions of adenovirus-2. *Journal of molecular biology*, 134 (2): 265–303. doi:10.1016/0022-2836(79)90036-6.

Cohen, M.J., Yousef, A.F., Massimi, P., et al. (2013) Dissection of the C-terminal region of E1A redefines the roles of CtBP and other cellular targets in oncogenic transformation. *Journal of virology*, 87 (18): 10348–10355. doi:10.1128/JVI.00786-13.

Crawford-Miksza, L. and Schnurr, D.P. (1996) Analysis of 15 adenovirus hexon proteins reveals the location and structure of seven hypervariable regions containing serotype-specific residues. *Journal of Virology*, 70 (3): 1836–1844. doi:10.1128/JVI.70.3.1836-1844.1996.

DAVID Bioinformatics Resources – NIAID/NIH (https://david.ncifcrf.gov).

Dhingra, A., Hage, E., Ganzenmueller, T., et al. (2019) Molecular Evolution of Human Adenovirus (HAdV) Species C. *Scientific Reports 2019 9:1*, 9 (1): 1–13. doi:10.1038/s41598-018-37249-4.

Dorsman, J.C., Teunisse, A.F.A.S., Zantema, A., et al. (1997) The adenovirus 12 E1A proteins can bind directly to proteins of the p300 transcription co-activator family, including the CREB-binding protein CBP and p300. *The Journal of general virology*, 78 (Pt 2) (2): 423–426. doi:10.1099/0022-1317-78-2-423.

Farley, D.C., Brown, J.L. and Leppard, K.N. (2004) Activation of the early-late switch in adenovirus type 5 major late transcription unit expression by L4 gene products. *Journal of virology*, 78 (4): 1782–1791. doi:10.1128/JVI.78.4.1782-1791.2004.

Ferrari, R., Gou, D., Jawdekar, G., et al. (2014) Adenovirus Small E1A Employs the Lysine Acetylases p300/CBP and Tumor Suppressor Rb to Repress Select Host Genes and Promote Productive Virus Infection. *Cell host & microbe*, 16 (5): 663. doi:10.1016/J.CHOM.2014.10.004.

Ferrari, R., Pellegrini, M., Horwitz, G.A., et al. (2008) Epigenetic reprogramming by adenovirus e1a. *Science (New York, N.Y.)*, 321 (5892): 1086–1088. doi:10.1126/SCIENCE.1155546.

Fuchs, M., Gerber, J., Drapkin, R., et al. (2001) The p400 complex is an essential E1A transformation target. *Cell*, 106 (3): 297–307. doi:10.1016/S0092-8674(01)00450-0.

Gallimore, P.H. and Turnell, A.S. (2001) Adenovirus E1A: remodelling the host cell, a life or death experience. *Oncogene*, 20 (54): 7824–7835. doi:10.1038/SJ.ONC.1204913.

Geisberg, J. V., Lee, W.S., Berk, A.J., et al. (1994) The zinc finger region of the adenovirus E1A transactivating domain complexes with the TATA box binding protein. *Proceedings of the National Academy of Sciences of the United States of America*, 91 (7): 2488–2492. doi:10.1073/PNAS.91.7.2488.

Glenewinkel, F., Cohen, M.J., King, C.R., et al. (2016) The adaptor protein DCAF7 mediates the interaction of the adenovirus E1A oncoprotein with the protein kinases DYRK1A and HIPK2. *Scientific Reports*, 6. doi:10.1038/SREP28241.

Goddard, T.D., Huang, C.C., Meng, E.C., et al. (2018) UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein science : a publication of the Protein Society*, 27 (1): 14–25. doi:10.1002/PRO.3235.

Gustafsson, B., Huang, W., Bogdanovic, G., et al. (2007) Adenovirus DNA is detected at increased frequency in Guthrie cards from children who develop acute lymphoblastic leukaemia. *British journal of cancer*, 97 (7): 992–994. doi:10.1038/SJ.BJC.6603983.

Hoeben, R.C. and Uil, T.G. (2013) Adenovirus DNA replication. *Cold Spring Harbor perspectives in biology*, 5 (3). doi:10.1101/CSHPERSPECT.A013003.

Horwitz, G.A., Zhang, K., McBrian, M.A., et al. (2008) Adenovirus small e1a alters global patterns of histone modification. *Science (New York, N.Y.)*, 321 (5892): 1084–1085. doi:10.1126/SCIENCE.1155544.

Human adenovirus working group (http://hadvwg.gmu.edu/).

Hsu, E., Pennella, M.A., Zemke, N.R., et al. (2018) Adenovirus E1A Activation Domain Regulates H3 Acetylation Affecting Varied Steps in Transcription at Different Viral Promoters. *Journal of virology*, 92 (18). doi:10.1128/JVI.00805-18.

Ip, W.H. and Dobner, T. (2020) Cell transformation by the adenovirus oncogenes E1 and E4. *FEBS letters*, 594 (12): 1848–1860. doi:10.1002/1873-3468.13717.

Jiang, Y., Han, Q., Zhao, H., et al. (2021) The Mechanisms of HBV-Induced Hepatocellular Carcinoma. *Journal of Hepatocellular Carcinoma*, 8: 435. doi:10.2147/JHC.S307962.

Jumper, J., Evans, R., Pritzel, A., et al. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature 2021 596:7873*, 596 (7873): 583–589. doi:10.1038/s41586-021-03819-2.

Kosulin, K., Haberler, C., Hainfellner, J.A., et al. (2007) Investigation of adenovirus occurrence in pediatric tumor entities. *Journal of virology*, 81 (14): 7629–7635. doi:10.1128/JVI.00355-07.

Krapp, L.F., Abriata, L.A., Cortés Rodriguez, F., et al. (2023) PeSTo: parameter-free geometric deep learning for accurate prediction of protein binding interfaces. *Nature communications*, 14 (1). doi:10.1038/S41467-023-37701-8.

Kulanayake, S. and Tikoo, S.K. (2021) Adenovirus Core Proteins: Structure and Function. *Viruses 2021, Vol. 13, Page 388*, 13 (3): 388. doi:10.3390/V13030388.

Lowe, S.W. and Earl Ruley, H. (1993) Stabilization of the p53 tumor suppressor is induced by adenovirus 5 E1A and accompanies apoptosis. *Genes & development*, 7 (4): 535–545. doi:10.1101/GAD.7.4.535.

Ma, H.-C. and Hearing, P. (2011) Adenovirus Structural Protein IIIa Is Involved in the Serotype Specificity of Viral DNA Packaging. *Journal of Virology*, 85 (15): 7849. doi:10.1128/JVI.00467-11.

MacLennan, S.A. and Marra, M.A. (2023) Oncogenic Viruses and the Epigenome: How Viruses Hijack Epigenetic Mechanisms to Drive Cancer. *International journal of molecular sciences*, 24 (11). doi:10.3390/IJMS24119543.

Marquez-Martinez, S., Vijayan, A., Khan, S., et al. (2023) Cell entry and innate sensing shape adaptive immune responses to adenovirus-based vaccines. *Current Opinion in Immunology*, 80: 102282. doi:10.1016/J.COI.2023.102282.

Miller, M.S., Pelka, P., Fonseca, G.J., et al. (2012) Characterization of the 55-residue protein encoded by the 9S E1A mRNA of species C adenovirus. *Journal of virology*, 86 (8): 4222–4233. doi:10.1128/JVI.06399-11.

NCBI Nucleotide BLAST program (https://blast ncbi nlm nih.gov/Blast.cgi?PAGE=MegaBlast&PROGRAM=blastn&BLAST_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch&BLAST_SPEC=blast2seq&DATABASE=n/a&QUERY=&SUBJECTS=).

Pelka, P., Ablack, J.N.G., Shuen, M., et al. (2009a) Identification of a second independent binding site for the pCAF acetyltransferase in adenovirus E1A. *Virology*, 391 (1): 90–98. doi:10.1016/J.VIROL.2009.05.024.

Pelka, P., Ablack, J.N.G., Torchia, J., et al. (2009b) Transcriptional control by adenovirus E1A conserved region 3 via p300/CBP. *Nucleic acids research*, 37 (4): 1095–1106. doi:10.1093/NAR/GKN1057.

Pérez-Berná, A.J., Mangel, W.F., McGrath, W.J., et al. (2014) Processing of the L1 52/55k Protein by the Adenovirus Protease: a New Substrate and New Insights into Virion Maturation. *Journal of Virology*, 88 (3): 1513. doi:10.1128/JVI.02884-13.

Pettersen, E.F., Goddard, T.D., Huang, C.C., et al. (2021) UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein science : a publication of the Protein Society*, 30 (1): 70–82. doi:10.1002/PRO.3943.

Querido, E., Teodoro, J.G. and Branton, P.E. (1997) Accumulation of p53 induced by the adenovirus E1A protein requires regions involved in the stimulation of DNA synthesis. *Journal of virology*, 71 (5): 3526–3533. doi:10.1128/JVI.71.5.3526-3533.1997.

Rasti, M., Grand, R.J.A., Yousef, A.F., et al. (2006) Roles for APIS and the 20S proteasome in adenovirus E1A-dependent transcription. *The EMBO journal*, 25 (12): 2710–2722. doi:10.1038/SJ.EMBOJ.7601169.

Sharma, P. V., Witteman, M., Sundaravel, S., et al. (2018) A case of HTLV-1 associated adult T-cell lymphoma presenting with cutaneous lesions and tropical spastic paresis. *Intractable & Rare Diseases Research*, 7 (1): 61. doi:10.5582/IRDR.2017.01077.

String-db (https://string-db.org/).

Szymonowicz, K.A. and Chen, J. (2020) Biological and clinical aspects of HPV-related cancers. *Cancer biology & medicine*, 17 (4): 864–878. doi:10.20892/J.ISSN.2095-3941.2020.0370.

Trentin, J.J., Yabe, Y. and Taylor, G. (1962) The quest for human cancer viruses. *Science (New York, N.Y.)*, 137 (3533): 835–841. doi:10.1126/SCIENCE.137.3533.835.

Turnell, A.S., Grand, R.J.A., Gorbea, C., et al. (2000) Regulation of the 26S proteasome by adenovirus E1A. *The EMBO journal*, 19 (17): 4759–4773. doi:10.1093/EMBOJ/19.17.4759.

Ulfendahl, P.J., Linder, S., Kreivi, J.P., et al. (1987) A novel adenovirus-2 E1A mRNA encoding a protein with transcription activation properties. *The EMBO Journal*, 6 (7): 2037. doi:10.1002/J.1460-2075.1987.TB02468.X.

Varadi, M., Anyango, S., Deshpande, M., et al. (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, 50 (D1): D439–D444. doi:10.1093/NAR/GKAB1061.

Wang, H.G., Yaciuk, P., Ricciardi, R.P., et al. (1993) The E1A products of oncogenic adenovirus serotype 12 include amino-terminally modified forms able to bind the retinoblastoma protein but not p300. *Journal of virology*, 67 (8): 4804–4813. doi:10.1128/JVI.67.8.4804-4813.1993.

Xu, W., McDonough, M.C. and Erdman, D.D. (2000) Species-Specific Identification of Human Adenoviruses by a Multiplex PCR Assay. *Journal of Clinical Microbiology*, 38 (11): 4114. doi:10.1128/JCM.38.11.4114-4120.2000.

Yabe, Y., Samper, L., Bryan, E., et al. (1964) ONCOGENIC EFFECT OF HUMAN ADENOVIRUS TYPE 12, IN MICE. *Science (New York, N.Y.)*, 143 (3601): 46–47. doi:10.1126/SCIENCE.143.3601.46.

Zhao, L.J., Subramanian, T. and Chinnadurai, G. (2008) Inhibition of transcriptional activation and cell proliferation activities of adenovirus E1A by the unique N-terminal domain of CtBP2. *Oncogene*, 27 (39): 5214. doi:10.1038/ONC.2008.162.

Appendix

Appendix 3.1. Mass spectrometry data – GFP pull down of Ad12-9S (n=278, excluding contaminants).

Accession	Protein	MW [kDa]	Scores	No. of peptides
TOP2A_HUMAN	DNA topoisomerase 2-alpha	174.3	2724.8 (M:2724.8)	55
TOP2B_HUMAN	DNA topoisomerase 2-beta	183.2	1902.6 (M:1902.6)	41
DHX9_HUMAN	ATP-dependent RNA helicase A	140.9	1561.6 (M:1561.6)	34
TBB3_HUMAN	Tubulin beta-3 chain	50.4	1076 9 (M:1076.9)	21
SMCA5_HUMAN	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5	121.8	861 1 (M:861.1)	20
HEAT1_HUMAN	HEAT repeat-containing protein 1	242.2	699 5 (M:699.5)	18
MATR3_HUMAN	Matrin-3	94.6	972.6 (M:972.6)	18
TBA1A_HUMAN	Tubulin alpha-1A chain	50.1	1277 3 (M:1277.3)	18
TBA1B_HUMAN	Tubulin alpha-1B chain	50.1	1299 9 (M:1299.9)	18
ILF3_HUMAN	Interleukin enhancer-binding factor 3	95.3	786.0 (M:786.0)	17
DJC10_HUMAN	DnaJ homolog subfamily C member 10	91	770.7 (M:770.7)	16
HS90B_HUMAN	Heat shock protein HSP 90-beta	83.2	944 3 (M:944.3)	16
NUMA1_HUMAN	Nuclear mitotic apparatus protein 1	238.1	580 9 (M:580.9)	16
RRP5_HUMAN	Protein RRP5 homolog	208.6	545 9 (M:545.9)	16
TBA3C_HUMAN	Tubulin alpha-3C/D chain	49.9	1061.4 (M:1061.4)	15
DDX21_HUMAN	Nucleolar RNA helicase 2	87.3	529.4 (M:529.4)	14
HNRPR_HUMAN	Heterogeneous nuclear ribonucleoprotein R	70.9	593.0 (M:593.0)	14
MCM7_HUMAN	DNA replication licensing factor MCM7	81.3	727.4 (M:727.4)	14
ROA3_HUMAN	Heterogeneous nuclear ribonucleoprotein A3	39.6	856 2 (M:856.2)	14
SP16H_HUMAN	FACT complex subunit SPT16	119.8	533.4 (M:533.4)	14
TBL3_HUMAN	Transducin beta-like protein 3	89	646 5 (M:646.5)	14
DDX18_HUMAN	ATP-dependent RNA helicase DDX18	75.4	596.4 (M:596.4)	12
KPYM_HUMAN	Pyruvate kinase isozymes M1/M2	57.9	523 5 (M:523.5)	12
ROA1_HUMAN	Heterogeneous nuclear ribonucleoprotein A1	38.7	755.7 (M:755.7)	12
SMCA1_HUMAN	Probable global transcription activator SNF2L1	122.5	515.0 (M:515.0)	12
XRCC5_HUMAN	X-ray repair cross-complementing protein 5	82.7	517.8 (M:517.8)	12
SSRP1_HUMAN	FACT complex subunit SSRP1	81	464 9 (M:464.9)	11
ATPA_HUMAN	ATP synthase subunit alpha, mitochondrial	59.7	481 1 (M:481.1)	10
BAZ1B_HUMAN	Tyrosine-protein kinase BAZ1B	170.8	360.4 (M:360.4)	10
H2B1B_HUMAN	Histone H2B type 1-B	13.9	604 9 (M:604.9)	10
H2B1N_HUMAN	Histone H2B type 1-N	13.9	617.4 (M:617.4)	10
ROA2_HUMAN	Heterogeneous nuclear ribonucleoproteins A2/B1	37.4	409.6 (M:409.6)	10
U520_HUMAN	U5 small nuclear ribonucleoprotein 200 kDa helicase	244.4	326 1 (M:326.1)	10
CBX3_HUMAN	Chromobox protein homolog 3	20.8	611 2 (M:611.2)	9
ECHA_HUMAN	Trifunctional enzyme subunit alpha, mitochondrial	82.9	410 1 (M:410.1)	9
H32_HUMAN	Histone H3.2	15.4	501 2 (M:501.2)	9
HNRPC_HUMAN	Heterogeneous nuclear ribonucleoproteins C1/C2	33.6	544 2 (M:544.2)	9
U5S1_HUMAN	116 kDa U5 small nuclear ribonucleoprotein component	109.4	445.4 (M:445.4)	9
CBX1_HUMAN	Chromobox protein homolog 1	21.4	537.0 (M:537.0)	8

DHX15_HUMAN	Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15	90.9	319.0 (M:319.0)	8
H31_HUMAN	Histone H3.1	15.4	439 2 (M:439.2)	8
HNRL2_HUMAN	Heterogeneous nuclear ribonucleoprotein U-like protein 2	85.1	345 3 (M:345.3)	8
NOP58_HUMAN	Nucleolar protein 58	59.5	379 1 (M:379.1)	8
RALY_HUMAN	RNA-binding protein Raly	32.4	357.8 (M:357.8)	8
RBM14_HUMAN	RNA-binding protein 14	69.4	328.6 (M:328.6)	8
RL1D1_HUMAN	Ribosomal L1 domain-containing protein 1	54.9	306 3 (M:306.3)	8
CBX5_HUMAN	Chromobox protein homolog 5	22.2	357 3 (M:357.3)	7
CKAP4_HUMAN	Cytoskeleton-associated protein 4	66	303.4 (M:303.4)	7
DDX27_HUMAN	Probable ATP-dependent RNA helicase DDX27	89.8	315 2 (M:315.2)	7
HNRH3_HUMAN	Heterogeneous nuclear ribonucleoprotein H3	36.9	423 3 (M:423.3)	7
HNRPL_HUMAN	Heterogeneous nuclear ribonucleoprotein L	64.1	214.6 (M:214.6)	7
NOG1_HUMAN	Nucleolar GTP-binding protein 1	73.9	231.0 (M:231.0)	7
NOL11_HUMAN	Nucleolar protein 11	81.1	276.8 (M:276.8)	7
NOP2_HUMAN	Putative ribosomal RNA methyltransferase NOP2	89.2	376 3 (M:376.3)	7
NOP56_HUMAN	Nucleolar protein 56	66	332 3 (M:332.3)	7
NPM_HUMAN	Nucleophosmin	32.6	382 5 (M:382.5)	7
PRP8_HUMAN	Pre-mRNA-processing-splicing factor 8	273.4	218 5 (M:218.5)	7
SF3B3_HUMAN	Splicing factor 3B subunit 3	135.5	339 5 (M:339.5)	7
XRCC6_HUMAN	X-ray repair cross-complementing protein 6	69.8	273 2 (M:273.2)	7
BOP1_HUMAN	Ribosome biogenesis protein BOP1	83.6	273.4 (M:273.4)	6
DDX3X_HUMAN	ATP-dependent RNA helicase DDX3X	73.2	169.4 (M:169.4)	6
NUCL_HUMAN	Nucleolin	76.6	338 9 (M:338.9)	6
RL10_HUMAN	60S ribosomal protein L10	24.6	281.7 (M:281.7)	6
RS15A_HUMAN	40S ribosomal protein S15a	14.8	225 5 (M:225.5)	6
SRSF1_HUMAN	Serine/arginine-rich splicing factor 1	27.7	269.6 (M:269.6)	6
TDIF2_HUMAN	Deoxynucleotidyltransferase terminal-interacting protein 2	84.4	269 1 (M:269.1)	6
WDR36_HUMAN	WD repeat-containing protein 36	105.3	227 9 (M:227.9)	6
WDR43_HUMAN	WD repeat-containing protein 43	74.8	240 1 (M:240.1)	6
ABC3B_HUMAN	Probable DNA dC->dU-editing enzyme APOBEC-3B	45.9	244 1 (M:244.1)	5
GBLP_HUMAN	Guanine nucleotide-binding protein subunit beta-2-like 1	35.1	296.6 (M:296.6)	5
HNRDL_HUMAN	Heterogeneous nuclear ribonucleoprotein D-like	46.4	248.8 (M:248.8)	5
ILF2_HUMAN	Interleukin enhancer-binding factor 2	43	223.0 (M:223.0)	5
RL18_HUMAN	60S ribosomal protein L18	21.6	301.7 (M:301.7)	5
RS13_HUMAN	40S ribosomal protein S13	17.2	213 1 (M:213.1)	5
RS2_HUMAN	40S ribosomal protein S2	31.3	275.4 (M:275.4)	5
RS27A_HUMAN	Ubiquitin-40S ribosomal protein S27a	18	250.0 (M:250.0)	5
RS9_HUMAN	40S ribosomal protein S9	22.6	179 9 (M:179.9)	5
RSF1_HUMAN	Remodeling and spacing factor 1	163.7	126.0 (M:126.0)	5
SF3B1_HUMAN	Splicing factor 3B subunit 1	145.7	151 2 (M:151.2)	5
SPB1_HUMAN	pre-rRNA processing protein FTSJ3	96.5	201.0 (M:201.0)	5
SRSF9_HUMAN	Serine/arginine-rich splicing factor 9	25.5	207 5 (M:207.5)	5
WDR12_HUMAN	Ribosome biogenesis protein WDR12	47.7	161 5 (M:161.5)	5

WDR75_HUMAN	WD repeat-containing protein 75	94.4	163 2 (M:163.2)	5
ZN326_HUMAN	DBIRD complex subunit ZNF326	65.6	228.0 (M:228.0)	5
ARF4_HUMAN	ADP-ribosylation factor 4	20.5	186 5 (M:186.5)	4
BRX1_HUMAN	Ribosome biogenesis protein BRX1 homolog	41.4	186.4 (M:186.4)	4
DHRS2_HUMAN	Dehydrogenase/reductase SDR family member 2	27.4	163 2 (M:163.2)	4
FBRL_HUMAN	rRNA 2'-O-methyltransferase fibrillarin	33.8	160 5 (M:160.5)	4
G3P_HUMAN	Glyceraldehyde-3-phosphate dehydrogenase	36	197 1 (M:197.1)	4
GPC1_HUMAN	Glypican-1	61.6	120.4 (M:120.4)	4
H2A1A_HUMAN	Histone H2A type 1-A	14.2	193 2 (M:193.2)	4
H2A1C_HUMAN	Histone H2A type 1-C	14.1	221 9 (M:221.9)	4
HNRPQ_HUMAN	Heterogeneous nuclear ribonucleoprotein Q	69.6	199 9 (M:199.9)	4
IF6_HUMAN	Eukaryotic translation initiation factor 6	26.6	187 2 (M:187.2)	4
KT33B_HUMAN	Keratin, type I cuticular Ha3-II	46.2	124 2 (M:124.2)	4
MK67I_HUMAN	MKI67 FHA domain-interacting nucleolar phosphoprotein	34.2	198.4 (M:198.4)	4
MPCP_HUMAN	Phosphate carrier protein, mitochondrial	40.1	152.8 (M:152.8)	4
PCBP2_HUMAN	Poly(rC)-binding protein 2	38.6	104 3 (M:104.3)	4
PININ_HUMAN	Pinin	81.6	190 3 (M:190.3)	4
PP2AA_HUMAN	Serine/threonine-protein phosphatase 2A catalytic subunit alpha isoform	35.6	235 5 (M:235.5)	4
PWP2_HUMAN	Periodic tryptophan protein 2 homolog	102.4	117.0 (M:117.0)	4
RBMX_HUMAN	RNA-binding motif protein, X chromosome	42.3	226 1 (M:226.1)	4
RL15_HUMAN	60S ribosomal protein L15	24.1	223 5 (M:223.5)	4
RL23_HUMAN	60S ribosomal protein L23	14.9	214.4 (M:214.4)	4
RL3_HUMAN	60S ribosomal protein L3	46.1	176.8 (M:176.8)	4
RL4_HUMAN	60S ribosomal protein L4	47.7	142 9 (M:142.9)	4
RL6_HUMAN	60S ribosomal protein L6	32.7	211 5 (M:211.5)	4
RMXL1_HUMAN	RNA binding motif protein, X-linked-like-1	42.1	193 2 (M:193.2)	4
RS16_HUMAN	40S ribosomal protein S16	16.4	186.7 (M:186.7)	4
SRSF3_HUMAN	Serine/arginine-rich splicing factor 3	19.3	174.8 (M:174.8)	4
TRA2B_HUMAN	Transformer-2 protein homolog beta	33.6	150.8 (M:150.8)	4
ACINU_HUMAN	Apoptotic chromatin condensation inducer in the nucleus	151.8	116.7 (M:116.7)	3
ATAD2_HUMAN	ATPase family AAA domain-containing protein 2	158.5	80.6 (M:80.6)	3
BAF_HUMAN	Barrier-to-autointegration factor	10.1	134 1 (M:134.1)	3
CIR1A_HUMAN	Cirhin	76.8	125.6 (M:125.6)	3
CO8A1_HUMAN	Collagen alpha-1(VIII) chain	73.3	72.9 (M:72 9)	3
CRYAB_HUMAN	Alpha-crystallin B chain	20.1	118.4 (M:118.4)	3
DCD_HUMAN	Dermcidin	11.3	110.7 (M:110.7)	3
DDB1_HUMAN	DNA damage-binding protein 1	126.9	81.0 (M:81.0)	3
ERH_HUMAN	Enhancer of rudimentary homolog	12.3	130.6 (M:130.6)	3
H14_HUMAN	Histone H1.4	21.9	110 3 (M:110.3)	3
IF4A3_HUMAN	Eukaryotic initiation factor 4A-III	46.8	101.8 (M:101.8)	3
KAT7_HUMAN	Histone acetyltransferase KAT7	70.6	131.0 (M:131.0)	3
KI67_HUMAN	Antigen KI-67	358.5	71.4 (M:71.4)	3
NAT10_HUMAN	N-acetyltransferase 10	115.7	104.8 (M:104.8)	3

NH2L1_HUMAN	NHP2-like protein 1	14.2	159 9 (M:159.9)	3
NOC2L_HUMAN	Nucleolar complex protein 2 homolog	84.9	99.2 (M:99 2)	3
PESC_HUMAN	Pescadillo homolog	68	71.0 (M:71.0)	3
RL12_HUMAN	60S ribosomal protein L12	17.8	164 1 (M:164.1)	3
RL14_HUMAN	60S ribosomal protein L14	23.4	138.4 (M:138.4)	3
RL18A_HUMAN	60S ribosomal protein L18a	20.7	146 5 (M:146.5)	3
RL24_HUMAN	60S ribosomal protein L24	17.8	165.0 (M:165.0)	3
RL30_HUMAN	60S ribosomal protein L30	12.8	166 9 (M:166.9)	3
RS12_HUMAN	40S ribosomal protein S12	14.5	132.7 (M:132.7)	3
SRSF7_HUMAN	Serine/arginine-rich splicing factor 7	27.4	152.0 (M:152.0)	3
SSRD_HUMAN	Translocon-associated protein subunit delta	19	149.6 (M:149.6)	3
SYG_HUMAN	Glycine--tRNA ligase	83.1	137.0 (M:137.0)	3
U3IP2_HUMAN	U3 small nucleolar RNA-interacting protein 2	51.8	100 5 (M:100.5)	3
WDR3_HUMAN	WD repeat-containing protein 3	106	171.0 (M:171.0)	3
4F2_HUMAN	4F2 cell-surface antigen heavy chain	68	83.4 (M:83.4)	2
C1TC_HUMAN	C-1-tetrahydrofolate synthase, cytoplasmic	101.5	107.7 (M:107.7)	2
DX39B_HUMAN	Spliceosome RNA helicase DDX39B	49	68.6 (M:68.6)	2
EF2_HUMAN	Elongation factor 2	95.3	78.0 (M:78.0)	2
FBX22_HUMAN	F-box only protein 22	44.5	120.8 (M:120.8)	2
FBX3_HUMAN	F-box only protein 3	54.5	73.7 (M:73.7)	2
G6PD_HUMAN	Glucose-6-phosphate 1-dehydrogenase	59.2	63.1 (M:63 1)	2
GFPT1_HUMAN	Glutamine--fructose-6-phosphate aminotransferase [isomerizing] 1	78.8	67.4 (M:67.4)	2
GNAO_HUMAN	Guanine nucleotide-binding protein G(o) subunit alpha	40	63.7 (M:63.7)	2
HNRPD_HUMAN	Heterogeneous nuclear ribonucleoprotein D0	38.4	99.2 (M:99 2)	2
HNRPK_HUMAN	Heterogeneous nuclear ribonucleoprotein K	50.9	109 9 (M:109.9)	2
IMP3_HUMAN	U3 small nucleolar ribonucleoprotein protein IMP3	21.8	97.3 (M:97 3)	2
MGN_HUMAN	Protein mago nashi homolog	17.2	69.7 (M:69.7)	2
MGST3_HUMAN	Microsomal glutathione S-transferase 3	16.5	136 3 (M:136.3)	2
MPP10_HUMAN	U3 small nucleolar ribonucleoprotein protein MPP10	78.8	74.3 (M:74 3)	2
MTCH2_HUMAN	Mitochondrial carrier homolog 2	33.3	66.1 (M:66 1)	2
NIP7_HUMAN	60S ribosome subunit biogenesis protein NIP7 homolog	20.4	63.7 (M:63.7)	2
NUP93_HUMAN	Nuclear pore complex protein Nup93	93.4	55.8 (M:55.8)	2
OCAD2_HUMAN	OCIA domain-containing protein 2	16.9	91.5 (M:91 5)	2
PABP3_HUMAN	Polyadenylate-binding protein 3	70	109.7 (M:109.7)	2
PCBP1_HUMAN	Poly(rC)-binding protein 1	37.5	57.4 (M:57.4)	2
PHF14_HUMAN	PHD finger protein 14	100	126 1 (M:126.1)	2
PIMT_HUMAN	Protein-L-isoaspartate(D-aspartate) O-methyltransferase	24.6	80.2 (M:80 2)	2
PPIA_HUMAN	Peptidyl-prolyl cis-trans isomerase A	18	76.0 (M:76.0)	2
PRP19_HUMAN	Pre-mRNA-processing factor 19	55.1	71.0 (M:71.0)	2
RBBP4_HUMAN	Histone-binding protein RBBP4	47.6	76.8 (M:76.8)	2
RBM28_HUMAN	RNA-binding protein 28	85.7	78.3 (M:78 3)	2
RFC3_HUMAN	Replication factor C subunit 3	40.5	114 1 (M:114.1)	2
RL10A_HUMAN	60S ribosomal protein L10a	24.8	85.8 (M:85.8)	2

RL13_HUMAN	60S ribosomal protein L13	24.2	111.7 (M:111.7)	2
RL13A_HUMAN	60S ribosomal protein L13a	23.6	85.9 (M:85 9)	2
RL21_HUMAN	60S ribosomal protein L21	18.6	66.1 (M:66 1)	2
RL22_HUMAN	60S ribosomal protein L22	14.8	97.2 (M:97 2)	2
RL31_HUMAN	60S ribosomal protein L31	14.5	110.7 (M:110.7)	2
RL32_HUMAN	60S ribosomal protein L32	15.8	84.8 (M:84.8)	2
RL35_HUMAN	60S ribosomal protein L35	14.5	98.2 (M:98 2)	2
RL5_HUMAN	60S ribosomal protein L5	34.3	78.6 (M:78.6)	2
RL7A_HUMAN	60S ribosomal protein L7a	30	59.8 (M:59.8)	2
RL7L_HUMAN	60S ribosomal protein L7-like 1	28.6	101 9 (M:101.9)	2
ROA0_HUMAN	Heterogeneous nuclear ribonucleoprotein A0	30.8	99.8 (M:99.8)	2
RPF2_HUMAN	Ribosome production factor 2 homolog	35.6	46.1 (M:46 1)	2
RS17L_HUMAN	40S ribosomal protein S17-like	15.5	70.1 (M:70 1)	2
RS24_HUMAN	40S ribosomal protein S24	15.4	95.4 (M:95.4)	2
RS25_HUMAN	40S ribosomal protein S25	13.7	95.6 (M:95.6)	2
RS27_HUMAN	40S ribosomal protein S27	9.5	112.0 (M:112.0)	2
RS28_HUMAN	40S ribosomal protein S28	7.8	92.5 (M:92 5)	2
SAFB2_HUMAN	Scaffold attachment factor B2	107.4	92.0 (M:92.0)	2
SKP1_HUMAN	S-phase kinase-associated protein 1	18.6	85.4 (M:85.4)	2
SLTM_HUMAN	SAFB-like transcription modulator	117.1	48.5 (M:48 5)	2
SMD1_HUMAN	Small nuclear ribonucleoprotein Sm D1	13.3	48.4 (M:48.4)	2
SRS10_HUMAN	Serine/arginine-rich splicing factor 10	31.3	57.9 (M:57 9)	2
SSF1_HUMAN	Suppressor of SWI4 1 homolog	53.2	54.5 (M:54 5)	2
SUGP2_HUMAN	SURP and G-patch domain-containing protein 2	120.1	79.1 (M:79 1)	2
TGM2_HUMAN	Protein-glutamine gamma-glutamyltransferase 2	77.3	69.9 (M:69 9)	2
UTP18_HUMAN	U3 small nucleolar RNA-associated protein 18 homolog	62	56.0 (M:56.0)	2
UTP6_HUMAN	U3 small nucleolar RNA-associated protein 6 homolog	70.1	69.7 (M:69.7)	2
WDR46_HUMAN	WD repeat-containing protein 46	68	85.4 (M:85.4)	2
WDR74_HUMAN	WD repeat-containing protein 74	42.4	66.8 (M:66.8)	2
YBOX1_HUMAN	Nuclease-sensitive element-binding protein 1	35.9	64.7 (M:64.7)	2
ZN384_HUMAN	Zinc finger protein 384	63.2	84.1 (M:84 1)	2
AATF_HUMAN	Protein AATF	63.1	28.8 (M:28.8)	1
AT1A3_HUMAN	Sodium/potassium-transporting ATPase subunit alpha-3	111.7	31.7 (M:31.7)	1
AT2A3_HUMAN	Sarcoplasmic/endoplasmic reticulum calcium ATPase 3	113.9	30.8 (M:30.8)	1
ATPG_HUMAN	ATP synthase subunit gamma, mitochondrial	33	42.5 (M:42 5)	1
BAG2_HUMAN	BAG family molecular chaperone regulator 2	23.8	42.8 (M:42.8)	1
BAZ2A_HUMAN	Bromodomain adjacent to zinc finger domain protein 2A	211.1	38.6 (M:38.6)	1
BMS1_HUMAN	Ribosome biogenesis protein BMS1 homolog	145.7	51.8 (M:51.8)	1
BOREA_HUMAN	Borealin	31.3	29.6 (M:29.6)	1
CAF1A_HUMAN	Chromatin assembly factor 1 subunit A	106.9	27.1 (M:27 1)	1
CEBPB_HUMAN	CCAAT/enhancer-binding protein beta	36.1	29.2 (M:29 2)	1
CEBPZ_HUMAN	CCAAT/enhancer-binding protein zeta	120.9	32.0 (M:32.0)	1
CENPB_HUMAN	Major centromere autoantigen B	65.1	64.0 (M:64.0)	1

CHTOP_HUMAN	Chromatin target of PRMT1 protein	26.4	41.5 (M:41 5)	1
CRNL1_HUMAN	Crooked neck-like protein 1	100.4	44.6 (M:44.6)	1
CUL1_HUMAN	Cullin-1	89.6	31.4 (M:31.4)	1
DCA13_HUMAN	DDB1- and CUL4-associated factor 13	51.4	29.2 (M:29 2)	1
DDX56_HUMAN	Probable ATP-dependent RNA helicase DDX56	61.6	39.5 (M:39 5)	1
DHCR7_HUMAN	7-dehydrocholesterol reductase	54.5	38.9 (M:38 9)	1
DHSA_HUMAN	Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial	72.6	25.2 (M:25 2)	1
DHX30_HUMAN	Putative ATP-dependent RNA helicase DHX30	133.9	42.7 (M:42.7)	1
DNJA1_HUMAN	DnaJ homolog subfamily A member 1	44.8	41.4 (M:41.4)	1
DNJC9_HUMAN	DnaJ homolog subfamily C member 9	29.9	44.7 (M:44.7)	1
DSC1_HUMAN	Desmocollin-1	99.9	28.7 (M:28.7)	1
EBP2_HUMAN	Probable rRNA-processing protein EBP2	34.8	34.0 (M:34.0)	1
EF1G_HUMAN	Elongation factor 1-gamma	50.1	26.0 (M:26.0)	1
EFTU_HUMAN	Elongation factor Tu, mitochondrial	49.5	76.6 (M:76.6)	1
ELOC_HUMAN	Transcription elongation factor B polypeptide 1	12.5	33.5 (M:33 5)	1
FCF1_HUMAN	rRNA-processing protein FCF1 homolog	23.4	29.8 (M:29.8)	1
FILA_HUMAN	Filaggrin	434.9	29.0 (M:29.0)	1
FXR1_HUMAN	Fragile X mental retardation syndrome-related protein 1	69.7	28.4 (M:28.4)	1
H10_HUMAN	Histone H1.0	20.9	63.6 (M:63.6)	1
HM20A_HUMAN	High mobility group protein 20A	40.1	30.6 (M:30.6)	1
HNRL1_HUMAN	Heterogeneous nuclear ribonucleoprotein U-like protein 1	95.7	44.5 (M:44 5)	1
IF2B3_HUMAN	Insulin-like growth factor 2 mRNA-binding protein 3	63.7	38.4 (M:38.4)	1
IMP4_HUMAN	U3 small nucleolar ribonucleoprotein protein IMP4	33.7	58.2 (M:58 2)	1
K0020_HUMAN	Pumilio domain-containing protein KIAA0020	73.5	40.8 (M:40.8)	1
KRA22_HUMAN	Keratin-associated protein 2-2	12.9	34.2 (M:34 2)	1
KRR1_HUMAN	KRR1 small subunit processome component homolog	43.6	34.3 (M:34 3)	1
MCM4_HUMAN	DNA replication licensing factor MCM4	96.5	33.1 (M:33 1)	1
MCM5_HUMAN	DNA replication licensing factor MCM5	82.2	29.9 (M:29 9)	1
METK2_HUMAN	S-adenosylmethionine synthase isoform type-2	43.6	34.2 (M:34 2)	1
MP2K3_HUMAN	Dual specificity mitogen-activated protein kinase kinase 3	39.3	50.4 (M:50.4)	1
NEP1_HUMAN	Ribosomal RNA small subunit methyltransferase NEP1	26.7	39.4 (M:39.4)	1
NOL6_HUMAN	Nucleolar protein 6	127.5	27.1 (M:27 1)	1
PARP1_HUMAN	Poly [ADP-ribose] polymerase 1	113	40.3 (M:40 3)	1
PM14_HUMAN	Pre-mRNA branch site protein p14	14.6	25.7 (M:25.7)	1
RBM8A_HUMAN	RNA-binding protein 8A	19.9	27.4 (M:27.4)	1
RCL1_HUMAN	RNA 3'-terminal phosphate cyclase-like protein	40.8	24.9 (M:24 9)	1
RL27_HUMAN	60S ribosomal protein L27	15.8	44.3 (M:44 3)	1
RL36_HUMAN	60S ribosomal protein L36	12.2	32.8 (M:32.8)	1
RL36A_HUMAN	60S ribosomal protein L36a	12.4	39.9 (M:39 9)	1
RLA1_HUMAN	60S acidic ribosomal protein P1	11.5	52.3 (M:52 3)	1
RLA2_HUMAN	60S acidic ribosomal protein P2	11.7	59.5 (M:59 5)	1
RM11_HUMAN	39S ribosomal protein L11, mitochondrial	20.7	37.7 (M:37.7)	1
RPN1_HUMAN	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1	68.5	45.9 (M:45 9)	1

RRS1_HUMAN	Ribosome biogenesis regulatory protein homolog	41.2	34.4 (M:34.4)	1
RS26_HUMAN	40S ribosomal protein S26	13	43.9 (M:43 9)	1
RS7_HUMAN	40S ribosomal protein S7	22.1	33.0 (M:33.0)	1
S61A1_HUMAN	Protein transport protein Sec61 subunit alpha isoform 1	52.2	35.9 (M:35 9)	1
SAP18_HUMAN	Histone deacetylase complex subunit SAP18	17.5	30.9 (M:30 9)	1
SAS10_HUMAN	Something about silencing protein 10	54.5	27.2 (M:27 2)	1
SDC4_HUMAN	Syndecan-4	21.6	32.2 (M:32 2)	1
SERA_HUMAN	D-3-phosphoglycerate dehydrogenase	56.6	67.9 (M:67 9)	1
SF3A1_HUMAN	Splicing factor 3A subunit 1	88.8	55.8 (M:55.8)	1
SMC1A_HUMAN	Structural maintenance of chromosomes protein 1A	143.1	39.1 (M:39 1)	1
SMD2_HUMAN	Small nuclear ribonucleoprotein Sm D2	13.5	32.6 (M:32.6)	1
SMD3_HUMAN	Small nuclear ribonucleoprotein Sm D3	13.9	50.2 (M:50 2)	1
SON_HUMAN	Protein SON	263.7	28.6 (M:28.6)	1
SRBP1_HUMAN	Sterol regulatory element-binding protein 1	121.6	30.8 (M:30.8)	1
SRP14_HUMAN	Signal recognition particle 14 kDa protein	14.6	66.2 (M:66 2)	1
SSBP_HUMAN	Single-stranded DNA-binding protein, mitochondrial	17.2	56.5 (M:56 5)	1
SURF4_HUMAN	Surfeit locus protein 4	30.4	24.8 (M:24.8)	1
THIO_HUMAN	Thioredoxin	11.7	28.8 (M:28.8)	1
TMM33_HUMAN	Transmembrane protein 33	28	28.4 (M:28.4)	1
TRA2A_HUMAN	Transformer-2 protein homolog alpha	32.7	71.5 (M:71 5)	1
USMG5_HUMAN	Up-regulated during skeletal muscle growth protein 5	6.5	25.3 (M:25 3)	1

Appendix 3.1. Full list of putative cellular Ad12 9S-binding proteins identified using Mass spectrometric analysis data of GFP-Ad12-9S following GFP-pulldown and SDS-PAGE.

Appendix 3.2. STRING Functional enrichment table – Biological Process (Gene Ontology).

GO-TERM	DESCRIPTION	COUNT	STRENGTH	FDR
GO:1905213	Negative regulation of mitotic chromosome condensation	2 of 2	2.78	0.0067
GO:0045870	Positive regulation of single stranded viral RNA replication v...	2 of 2	2.78	0.0067
GO:1990164	Histone H2A phosphorylation	2 of 6	2.3	0.0194
GO:0044806	G-quadruplex DNA unwinding	2 of 7	2.23	0.0238
GO:0006265	DNA topological change	2 of 10	2.08	0.0397
GO:0045945	Positive regulation of transcription by RNA polymerase Ill	3 of 19	1.97	0.0037
GO:0045943	Positive regulation of transcription by RNA polymerase I	4 of 33	1.86	0.00043
GO:0016572	Histone phosphorylation	3 of 33	1.73	0.0096
GO:1904358	Positive regulation of telomere maintenance via telomere le...	3 of 37	1.68	0.0122
GO:0035066	Positive regulation of histone acetylation	3 of 41	1.64	0.0151
GO:0051972	Regulation of telomerase activity	3 of 51	1.55	0.0241
GO:0071103	DNA conformation change	6 of 104	1.54	5.02E-05
GO:0048524	Positive regulation of viral process	3 of 62	1.46	0.0394
GO:2001251	Negative regulation of chromosome organization	4 of 84	1.45	0.006
GO:0032392	DNA geometric change	4 of 95	1.4	0.008
GO:0006334	Nucleosome assembly	5 of 124	1.38	0.0017
GO:2001252	Positive regulation of chromosome organization	4 of 107	1.35	0.0106
GO:2000278	Regulation of DNA biosynthetic process	4 of 124	1.28	0.0162
GO:0033044	Regulation of chromosome organization	7 of 253	1.22	0.00023
GO:0006338	Chromatin remodeling	8 of 303	1.2	5.38E-05
GO:0006260	DNA replication	5 of 203	1.17	0.0081
GO:0050792	Regulation of viral process	4 of 161	1.17	0.0368
GO:0000398	mRNA splicing, via spliceosome	6 of 245	1.16	0.0023
GO:0006364	rRNA processing	5 of 220	1.13	0.0106
GO:0022613	Ribonucleoprotein complex biogenesis	9 of 449	1.08	6.06E-05
GO:0042254	Ribosome biogenesis	6 of 299	1.08	0.0051
GO:0051054	Positive regulation of DNA metabolic process	6 of 304	1.07	0.0053
GO:0051276	Chromosome organization	15 of 968	0.97	1.82E-07
GO:0034470	ncRNA processing	6 of 409	0.94	0.0157
GO:0051052	Regulation of DNA metabolic process	7 of 541	0.89	0.0093
GO:0006396	RNA processing	11 of 868	0.88	0.00012
GO:0006259	DNA metabolic process	10 of 785	0.88	0.00043
GO:0006974	Cellular response to DNA damage stimulus	7 of 744	0.75	0.0434
GO:0090304	Nucleic acid metabolic process	19 of 2203	0.71	7.95E-07
GO:0016070	RNA metabolic process	13 of 1550	0.7	0.00054
GO:0006139	Nucleobase-containing compound metabolic process	21 of 2722	0.66	4.83E-07
GO:0065003	Protein-containing complex assembly	9 of 1303	0.62	0.0458
GO:0045935	Positive regulation of nucleobase-containing compound me...	14 of 2056	0.61	0.0017
GO:0010557	Positive regulation of macromolecule biosynthetic process	13 of 1935	0.6	0.0038
GO:0031328	Positive regulation of cellular biosynthetic process	13 of 2041	0.58	0.0058
GO:0033554	Cellular response to stress	10 of 1572	0.58	0.0397
GO:0010467	Gene expression	13 of 2101	0.57	0.0067
GO:0006996	Organelle organisation	20 of 3470	0.54	6.06E-05

GO:0071840	Cellular component organisation or biogenesis	25 of 5639	0.42	5.02E-05
GO:0016043	Cellular component organisation	21 of 5436	0.36	0.0067
GO:0071704	Organic substance metabolic process	24 of 7522	0.28	0.0162

Appendix 3.2. STRING Functional enrichment table – Biological Process (Gene Ontology). Count in network represents the number of proteins in the gene list input into STRING (n=33) that are involved in the biological processes. The second number in the count is the total number of proteins that are associated with the biological process; this includes proteins not in the network. The strength indicates the level of enrichment. The False Discovery Rate (FDR) indicates the significance of the enrichment.

Appendix 3.3. STRING Functional enrichment table – Ribosome biogenesis (Gene Ontology)

UNIPROT_ID	GENE NAME	MW [kDa]	Score	No. of peptides
BMS1_HUMAN	BMS1 ribosome biogenesis factor(BMS1)	145.7	51.8	1
BOP1_HUMAN	BOP1 ribosomal biogenesis factor(BOP1)	83.6	273.4	6
DCA13_HUMAN	DDB1 and CUL4 associated factor 13(DCAF13)	51.4	29.2	1
DDX27_HUMAN	DEAD-box helicase 27(DDX27)	89.8	315.2	7
DDX3X_HUMAN	DEAD-box helicase 3 X-linked(DDX3X)	73.2	169.4	6
DDX56_HUMAN	DEAD-box helicase 56(DDX56)	61.6	39.5	1
DHX30_HUMAN	DExH-box helicase 30(DHX30)	133.9	42.7	1
EBP2_HUMAN	EBNA1 binding protein 2(EBNA1BP2)	34.8	34.0	1
NEP1_HUMAN	EMG1 N1-specific pseudouridine methyltransferase(EMG1)	26.7	39.4	1
FCF1_HUMAN	FCF1 rRNA-processing protein(FCF1)	23.4	29.8	1
SPB1_HUMAN	FtsJ RNA 2'-O-methyltransferase 3(FTSJ3)	96.5	201.0	5
HEAT1_HUMAN	HEAT repeat containing 1(HEATR1)	242.2	699.5	18
IMP3_HUMAN	IMP U3 small nucleolar ribonucleoprotein 3(IMP3)	21.8	97.3	2
IMP4_HUMAN	IMP U3 small nucleolar ribonucleoprotein 4(IMP4)	33.7	58.2	1
KRR1_HUMAN	KRR1 small subunit processome component homolog(KRR1)	43.6	34.3	1
MPP10_HUMAN	M-phase phosphoprotein 10(MPHOSPH10)	78.8	74.3	2
NOP2_HUMAN	NOP2 nucleolar protein(NOP2)	89.2	376.3	7
NOP56_HUMAN	NOP56 ribonucleoprotein(NOP56)	66	332.3	7
NOP58_HUMAN	NOP58 ribonucleoprotein(NOP58)	59.5	379.1	8
RCL1_HUMAN	RNA terminal phosphate cyclase like 1(RCL1)	40.8	24.9	1
WDR12_HUMAN	WD repeat domain 12(WDR12)	47.7	161.5	5
WDR36_HUMAN	WD repeat domain 36(WDR36)	105.3	227.9	6
WDR43_HUMAN	WD repeat domain 43(WDR43)	74.8	240.1	6
WDR75_HUMAN	WD repeat domain 75(WDR75)	94.4	163.2	5
XRCC5_HUMAN	X-ray repair cross complementing 5(XRCC5)	82.7	517.8	12
BRX1_HUMAN	biogenesis of ribosomes BRX1(BRIX1)	41.4	186.4	4
IF6_HUMAN	eukaryotic translation initiation factor 6(EIF6)	26.6	187.2	4
NIP7_HUMAN	nucleolar pre-rRNA processing protein NIP7(NIP7)	20.4	63.7	2
NOL11_HUMAN	nucleolar protein 11(NOL11)	81.1	276.8	7
PESC_HUMAN	pescadillo ribosomal biogenesis factor 1(PES1)	68	71.0	3
RRS1_HUMAN	ribosome biogenesis regulator 1 homolog(RRS1)	41.2	34.4	1
RPF2_HUMAN	ribosome production factor 2 homolog(RPF2)	35.6	46.1	2

Appendix 3.3. STRING Functional enrichment table – Ribosome biogenesis (Gene Ontology). STRING analysis of Mass spectrometric analysis data of GFP-Ad12-9S following GFP-pulldown and SDS-PAGE. The

table summarises the genes from our list that are associated ribosome biogenesis (n=32). Listed are the molecular weights (kDa), number of peptides, and the scores of each gene.

Appendix 3.4. STRING Functional Annotation Table - mRNA splicing

UNIPROT_ID	GENE NAME	MW [kDa]	Score	No. of peptides
DHX15_HUMAN	DEAH-box helicase 15(DHX15)	90.9	319.0	8
DX39B_HUMAN	DExD-box helicase 39B(DDX39B)	49	68.6	2
DHX9_HUMAN	DExH-box helicase 9(DHX9)	140.9	1561.6	34
RALY_HUMAN	RALY heterogeneous nuclear ribonucleoprotein(RALY)	32.4	357.8	8
RMXL1_HUMAN	RBMX like 1(RBMXL1)	42.1	193.2	4
RBM28_HUMAN	RNA binding motif protein 28(RBM28)	85.7	78.3	2
RBM8A_HUMAN	RNA binding motif protein 8A(RBM8A)	19.9	27.4	1
RBMX_HUMAN	RNA binding motif protein X-linked(RBMX)	42.3	226.1	4
SON_HUMAN	SON DNA and RNA binding protein(SON)	263.7	28.6	1
SUGP2_HUMAN	SURP and G-patch domain containing 2(SUGP2)	120.1	79.1	2
SAP18_HUMAN	Sin3A associated protein 18(SAP18)	17.5	30.9	1
YBOX1_HUMAN	Y-box binding protein 1(YBX1)	35.9	64.7	2
ACINU_HUMAN	apoptotic chromatin condensation inducer 1(ACIN1)	151.8	116.7	3
CRNL1_HUMAN	crooked neck pre-mRNA splicing factor 1(CRNKL1)	100.4	44.6	1
U5S1_HUMAN	elongation factor Tu GTP binding domain containing 2(EFTUD2)	109.4	445.4	9
IF4A3_HUMAN	eukaryotic translation initiation factor 4A3(EIF4A3)	46.8	101.8	3
ROA1_HUMAN	heterogeneous nuclear ribonucleoprotein A1(HNRNPA1)	38.7	755.7	12
ROA2_HUMAN	heterogeneous nuclear ribonucleoprotein A2/B1(HNRNPA2B1)	37.4	409.6	10
ROA3_HUMAN	heterogeneous nuclear ribonucleoprotein A3(HNRNPA3)	39.6	856.2	14
HNRPC_HUMAN	heterogeneous nuclear ribonucleoprotein C(HNRNPC)	33.6	544.2	9
HNRPK_HUMAN	heterogeneous nuclear ribonucleoprotein K(HNRNPK)	50.9	109.9	2
HNRPR_HUMAN	heterogeneous nuclear ribonucleoprotein R(HNRNPR)	70.9	593.0	14
MGN_HUMAN	mago homolog, exon junction complex subunit(MAGOH)	17.2	69.7	2
PININ_HUMAN	pinin, desmosome associated protein(PNN)	81.6	190.3	4
PRP19_HUMAN	pre-mRNA processing factor 19(PRPF19)	55.1	71.0	2
PRP8_HUMAN	pre-mRNA processing factor 8(PRPF8)	273.4	218.5	7
SRSF1_HUMAN	serine and arginine rich splicing factor 1(SRSF1)	27.7	269.6	6
SRS10_HUMAN	serine and arginine rich splicing factor 10(SRSF10)	31.3	57.9	2
SRSF3_HUMAN	serine and arginine rich splicing factor 3(SRSF3)	19.3	174.8	4
SRSF7_HUMAN	serine and arginine rich splicing factor 7(SRSF7)	27.4	152.0	3
SRSF9_HUMAN	serine and arginine rich splicing factor 9(SRSF9)	25.5	207.5	5
NH2L1_HUMAN	small nuclear ribonucleoprotein 13(SNU13)	14.2	159.9	3
SMD1_HUMAN	small nuclear ribonucleoprotein D1 polypeptide(SNRPD1)	13.3	48.4	2
SMD2_HUMAN	small nuclear ribonucleoprotein D2 polypeptide(SNRPD2)	13.5	32.6	1
SMD3_HUMAN	small nuclear ribonucleoprotein D3 polypeptide(SNRPD3)	13.9	50.2	1
U520_HUMAN	small nuclear ribonucleoprotein U5 subunit 200(SNRNP200)	244.4	326.1	10
SF3A1_HUMAN	splicing factor 3a subunit 1(SF3A1)	88.8	55.8	1
SF3B1_HUMAN	splicing factor 3b subunit 1(SF3B1)	145.7	151.2	5
SF3B3_HUMAN	splicing factor 3b subunit 3(SF3B3)	135.5	339.5	7

HNRPQ_HUMAN	synaptotagmin binding cytoplasmic RNA interacting protein(SYNCRIP)	69.6	199.9	4
TRA2A_HUMAN	transformer 2 alpha homolog(TRA2A)	32.7	71.5	1
TRA2B_HUMAN	transformer 2 beta homolog(TRA2B)	33.6	150.8	4
ZN326_HUMAN	zinc finger protein 326(ZNF326)	65.6	228.0	5

Appendix 3.4. STRING Functional enrichment table – mRNA splicing (Gene Ontology). STRING analysis of Mass spectrometric analysis data of GFP-Ad12-9S following GFP-pulldown and SDS-PAGE. The table summarises the genes from our list that are associated mRNA splicing (n=43). Listed are the molecular weights (kDa), number of peptides, and the scores of each gene.

Appendix 3.5. STRING Functional Annotation Table - mRNA processing

UNIPROT_ID	GENE NAME	MW [kDa]	Scores	No. of peptides
DHX15_HUMAN	DEAH-box helicase 15(DHX15)	90.9	319.0	8
DX39B_HUMAN	DExD-box helicase 39B(DDX39B)	49	68.6	2
DHX9_HUMAN	DExH-box helicase 9(DHX9)	140.9	1561.6	34
RALY_HUMAN	RALY heterogeneous nuclear ribonucleoprotein(RALY)	32.4	357.8	8
RMXL1_HUMAN	RBMX like 1(RBMXL1)	42.1	193.2	4
RBM28_HUMAN	RNA binding motif protein 28(RBM28)	85.7	78.3	2
RBM8A_HUMAN	RNA binding motif protein 8A(RBM8A)	19.9	27.4	1
RBMX_HUMAN	RNA binding motif protein X-linked(RBMX)	42.3	226.1	4
SON_HUMAN	SON DNA and RNA binding protein(SON)	263.7	28.6	1
SUGP2_HUMAN	SURP and G-patch domain containing 2(SUGP2)	120.1	79.1	2
SAP18_HUMAN	Sin3A associated protein 18(SAP18)	17.5	30.9	1
YBOX1_HUMAN	Y-box binding protein 1(YBX1)	35.9	64.7	2
ACINU_HUMAN	apoptotic chromatin condensation inducer 1(ACIN1)	151.8	116.7	3
CRNL1_HUMAN	crooked neck pre-mRNA splicing factor 1(CRNKL1)	100.4	44.6	1
U5S1_HUMAN	elongation factor Tu GTP binding domain containing 2(EFTUD2)	109.4	445.4	9
IF4A3_HUMAN	eukaryotic translation initiation factor 4A3(EIF4A3)	46.8	101.8	3
ROA1_HUMAN	heterogeneous nuclear ribonucleoprotein A1(HNRNPA1)	38.7	755.7	12
ROA2_HUMAN	heterogeneous nuclear ribonucleoprotein A2/B1(HNRNPA2B1)	37.4	409.6	10
ROA3_HUMAN	heterogeneous nuclear ribonucleoprotein A3(HNRNPA3)	39.6	856.2	14
HNRPC_HUMAN	heterogeneous nuclear ribonucleoprotein C(HNRNPC)	33.6	544.2	9
HNRH3_HUMAN	heterogeneous nuclear ribonucleoprotein H3(HNRNPH3)	36.9	423.3	7
HNRPK_HUMAN	heterogeneous nuclear ribonucleoprotein K(HNRNPK)	50.9	109.9	2
HNRPR_HUMAN	heterogeneous nuclear ribonucleoprotein R(HNRNPR)	70.9	593.0	14
MGN_HUMAN	mago homolog, exon junction complex subunit(MAGOH)	17.2	69.7	2
PININ_HUMAN	pinin, desmosome associated protein(PNN)	81.6	190.3	4
PRP19_HUMAN	pre-mRNA processing factor 19(PRPF19)	55.1	71.0	2
PRP8_HUMAN	pre-mRNA processing factor 8(PRPF8)	273.4	218.5	7
SRSF1_HUMAN	serine and arginine rich splicing factor 1(SRSF1)	27.7	269.6	6

SRS10_HUMAN	serine and arginine rich splicing factor 10(SRSF10)	31.3	57.9	2
SRSF3_HUMAN	serine and arginine rich splicing factor 3(SRSF3)	19.3	174.8	4
SRSF7_HUMAN	serine and arginine rich splicing factor 7(SRSF7)	27.4	152.0	3
SRSF9_HUMAN	serine and arginine rich splicing factor 9(SRSF9)	25.5	207.5	5
NH2L1_HUMAN	small nuclear ribonucleoprotein 13(SNU13)	14.2	159.9	3
SMD1_HUMAN	small nuclear ribonucleoprotein D1 polypeptide(SNRPD1)	13.3	48.4	2
SMD2_HUMAN	small nuclear ribonucleoprotein D2 polypeptide(SNRPD2)	13.5	32.6	1
SMD3_HUMAN	small nuclear ribonucleoprotein D3 polypeptide(SNRPD3)	13.9	50.2	1
U520_HUMAN	small nuclear ribonucleoprotein U5 subunit 200(SNRNP200)	244.4	326.1	10
SF3A1_HUMAN	splicing factor 3a subunit 1(SF3A1)	88.8	55.8	1
SF3B1_HUMAN	splicing factor 3b subunit 1(SF3B1)	145.7	151.2	5
SF3B3_HUMAN	splicing factor 3b subunit 3(SF3B3)	135.5	339.5	7
HNRPQ_HUMAN	synaptotagmin binding cytoplasmic RNA interacting protein(SYNCRIP)	69.6	199.9	4
TRA2A_HUMAN	transformer 2 alpha homolog(TRA2A)	32.7	71.5	1
TRA2B_HUMAN	transformer 2 beta homolog(TRA2B)	33.6	150.8	4
ZN326_HUMAN	zinc finger protein 326(ZNF326)	65.6	228.0	5

Appendix 3.5. STRING Functional enrichment table – mRNA processing (Gene Ontology). STRING analysis of Mass spectrometric analysis data of GFP-Ad12-9S following GFP-pulldown and SDS-PAGE. The table summarises the genes from our list that are associated mRNA processing (n=44). Listed are the molecular weights (kDa), number of peptides, and the scores of each gene.

Appendix 3.6. STRING Functional Annotation Table - rRNA processing

UNIPROT_ID	GENE NAME	MW [kDa]	Scores	No. of peptides
BOP1_HUMAN	BOP1 ribosomal biogenesis factor(BOP1)	83.6	273.4	6
DCA13_HUMAN	DDB1 and CUL4 associated factor 13(DCAF13)	51.4	29.2	1
DDX27_HUMAN	DEAD-box helicase 27(DDX27)	89.8	315.2	7
DDX56_HUMAN	DEAD-box helicase 56(DDX56)	61.6	39.5	1
DDX21_HUMAN	DExD-box helicase 21(DDX21)	87.3	529.4	14
NEP1_HUMAN	EMG1 N1-specific pseudouridine methyltransferase(EMG1)	26.7	39.4	1
FCF1_HUMAN	FCF1 rRNA-processing protein(FCF1)	23.4	29.8	1
SPB1_HUMAN	FtsJ RNA 2'-O-methyltransferase 3(FTSJ3)	96.5	201.0	5
HEAT1_HUMAN	HEAT repeat containing 1(HEATR1)	242.2	699.5	18
IMP3_HUMAN	IMP U3 small nucleolar ribonucleoprotein 3(IMP3)	21.8	97.3	2
IMP4_HUMAN	IMP U3 small nucleolar ribonucleoprotein 4(IMP4)	33.7	58.2	1
KRR1_HUMAN	KRR1 small subunit processome component homolog(KRR1)	43.6	34.3	1
MPP10_HUMAN	M-phase phosphoprotein 10(MPHOSPH10)	78.8	74.3	2
NAT10_HUMAN	N-acetyltransferase 10(NAT10)	115.7	104.8	3
NOP2_HUMAN	NOP2 nucleolar protein(NOP2)	89.2	376.3	7
UTP18_HUMAN	UTP18 small subunit processome component(UTP18)	62	56.0	2
UTP6_HUMAN	UTP6 small subunit processome component(UTP6)	70.1	69.7	2
WDR12_HUMAN	WD repeat domain 12(WDR12)	47.7	161.5	5
WDR36_HUMAN	WD repeat domain 36(WDR36)	105.3	227.9	6
WDR43_HUMAN	WD repeat domain 43(WDR43)	74.8	240.1	6
WDR75_HUMAN	WD repeat domain 75(WDR75)	94.4	163.2	5

IF4A3_HUMAN	eukaryotic translation initiation factor 4A3(EIF4A3)	46.8	101.8	3
FBRL_HUMAN	fibrillarin(FBL)	33.8	160.5	4
NOL11_HUMAN	nucleolar protein 11(NOL11)	81.1	276.8	7
PESC_HUMAN	pescadillo ribosomal biogenesis factor 1(PES1)	68	71.0	3
RRP5_HUMAN	programmed cell death 11(PDCD11)	208.6	545.9	16
U3IP2_HUMAN	ribosomal RNA processing 9, U3 small nucleolar RNA binding protein(RRP9)	51.8	100.5	3

Appendix 3.6. STRING Functional enrichment table – rRNA processing (Gene Ontology). STRING analysis of Mass spectrometric analysis data of GFP-Ad12-9S following GFP-pulldown and SDS-PAGE. The table summarises the genes from our list that are associated rRNA processing (n=27). Listed are the molecular weights (kDa), number of peptides, and the scores of each gene.

Appendix 3.7. STRING Functional Annotation Table - mRNA transport

UNIPROT_ID	GENE NAME	MW [kDa]	Scores	No. of peptides
DX39B_HUMAN	DExD-box helicase 39B(DDX39B)	49	68.6	2
DHX9_HUMAN	DExH-box helicase 9(DHX9)	140.9	1561.6	34
RBM8A_HUMAN	RNA binding motif protein 8A(RBM8A)	19.9	27.4	1
CHTOP_HUMAN	chromatin target of PRMT1(CHTOP)	26.4	41.5	1
IF4A3_HUMAN	eukaryotic translation initiation factor 4A3(EIF4A3)	46.8	101.8	3
ROA1_HUMAN	heterogeneous nuclear ribonucleoprotein A1(HNRNPA1)	38.7	755.7	12
ROA2_HUMAN	heterogeneous nuclear ribonucleoprotein A2/B1(HNRNPA2B1)	37.4	409.6	10
IF2B3_HUMAN	insulin like growth factor 2 mRNA binding protein 3(IGF2BP3)	63.7	38.4	1
MGN_HUMAN	mago homolog, exon junction complex subunit(MAGOH)	17.2	69.7	2
NUP93_HUMAN	nucleoporin 93(NUP93)	93.4	55.8	2
SRSF1_HUMAN	serine and arginine rich splicing factor 1(SRSF1)	27.7	269.6	6
SRSF3_HUMAN	serine and arginine rich splicing factor 3(SRSF3)	19.3	174.8	4
SRSF7_HUMAN	serine and arginine rich splicing factor 7(SRSF7)	27.4	152.0	3

Appendix 3.7. STRING Functional enrichment table – mRNA transport (Gene Ontology). STRING analysis of Mass spectrometric analysis data of GFP-Ad12-9S following GFP-pulldown and SDS-PAGE. The table summarises the genes from our list that are associated mRNA transport (n=13). Listed are the molecular weights (kDa), number of peptides, and the scores of each gene.

Appendix 3.8. STRING Functional Annotation Table - DNA replication

UNIPROT_ID	GENE NAME	MW [kDa]	Scores	No. of peptides
DHX9_HUMAN	DExH-box helicase 9(DHX9)	140.9	1561.6	34
RBBP4_HUMAN	RB binding protein 4, chromatin remodeling factor(RBBP4)	47.6	76.8	2
SP16H_HUMAN	SPT16 homolog, facilitates chromatin remodeling subunit(SUPT16H)	119.8	533.4	14
CAF1A_HUMAN	chromatin assembly factor 1 subunit A(CHAF1A)	106.9	27.1	1
KAT7_HUMAN	lysine acetyltransferase 7(KAT7)	70.6	131.0	3
MCM4_HUMAN	minichromosome maintenance complex component 4(MCM4)	96.5	33.1	1

MCM5_HUMAN	minichromosome maintenance complex component 5(MCM5)	82.2	29.9	1
MCM7_HUMAN	minichromosome maintenance complex component 7(MCM7)	81.3	727.4	14
RFC3_HUMAN	replication factor C subunit 3(RFC3)	40.5	114.1	2
SSBP_HUMAN	single stranded DNA binding protein 1(SSBP1)	17.2	56.5	1
SSRP1_HUMAN	structure specific recognition protein 1(SSRP1)	81	464.9	11

Appendix 3.8. STRING Functional enrichment table – DNA replication (Gene Ontology). STRING analysis of Mass spectrometric analysis data of GFP-Ad12-9S following GFP-pulldown and SDS-PAGE. The table summarises the genes from our list that are associated DNA replication (n=11). Listed are the molecular weights (kDa), number of peptides, and the scores of each gene.

Appendix 3.9. STRING Functional Annotation Table - Translation regulation

UNIPROT_ID	GENE NAME	MW [kDa]	Scores	No. of peptides
DDX3X_HUMAN	DEAD-box helicase 3 X-linked(DDX3X)	73.2	169.4	6
DHX9_HUMAN	DExH-box helicase 9(DHX9)	140.9	1561.6	34
RBM8A_HUMAN	RNA binding motif protein 8A(RBM8A)	19.9	27.4	1
IF4A3_HUMAN	eukaryotic translation initiation factor 4A3(EIF4A3)	46.8	101.8	3
G3P_HUMAN	glyceraldehyde-3-phosphate dehydrogenase(GAPDH)	36	197.1	4
IF2B3_HUMAN	insulin like growth factor 2 mRNA binding protein 3(IGF2BP3)	63.7	38.4	1
MGN_HUMAN	mago homolog, exon junction complex subunit(MAGOH)	17.2	69.7	2
KPYM_HUMAN	pyruvate kinase M1/2(PKM)	57.9	523.5	12
RL10_HUMAN	ribosomal protein L10(RPL10)	24.6	281.7	6
RL13A_HUMAN	ribosomal protein L13a(RPL13A)	23.6	85.9	2
HNRPQ_HUMAN	synaptotagmin binding cytoplasmic RNA interacting protein(SYNCRIP)	69.6	199.9	4

Appendix 3.9. STRING Functional enrichment table – Translation regulation (Gene Ontology). STRING analysis of Mass spectrometric analysis data of GFP-Ad12-9S following GFP-pulldown and SDS-PAGE. The table summarises the genes from our list that are associated translation regulation (n=11). Listed are the molecular weights (kDa), number of peptides, and the scores of each gene.

Appendix 3.10. STRING Functional Annotation Table - Host-virus interaction

UNIPROT_ID	GENE NAME	MW [kDa]	Scores	No. of peptides
BAF_HUMAN	BAF nuclear assembly factor 1(BANF1)	10.1	134.1	3
DDX3X_HUMAN	DEAD-box helicase 3 X-linked(DDX3X)	73.2	169.4	6
DHX9_HUMAN	DExH-box helicase 9(DHX9)	140.9	1561.6	34
FBX3_HUMAN	F-box protein 3(FBXO3)	54.5	73.7	2
SKP1_HUMAN	S-phase kinase associated protein 1(SKP1)	18.6	85.4	2
SP16H_HUMAN	SPT16 homolog, facilitates chromatin remodeling subunit(SUPT16H)	119.8	533.4	14
SMCA5_HUMAN	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 5(SMARCA5)	121.8	861.1	20
EFTU_HUMAN	Tu translation elongation factor, mitochondrial(TUFM)	49.5	76.6	1

XRCC5_HUMAN	X-ray repair cross complementing 5(XRCC5)	82.7	517.8	12
XRCC6_HUMAN	X-ray repair cross complementing 6(XRCC6)	69.8	273.2	7
CBX5_HUMAN	chromobox 5(CBX5)	22.2	357.3	7
CUL1_HUMAN	cullin 1(CUL1)	89.6	31.4	1
DDB1_HUMAN	damage specific DNA binding protein 1(DDB1)	126.9	81.0	3
ELOC_HUMAN	elongin C(ELOC)	12.5	33.5	1
ROA1_HUMAN	heterogeneous nuclear ribonucleoprotein A1(HNRNPA1)	38.7	755.7	12
ROA2_HUMAN	heterogeneous nuclear ribonucleoprotein A2/B1(HNRNPA2B1)	37.4	409.6	10
HNRPK_HUMAN	heterogeneous nuclear ribonucleoprotein K(HNRNPK)	50.9	109.9	2
NPM_HUMAN	nucleophosmin 1(NPM1)	32.6	382.5	7
PPIA_HUMAN	peptidylprolyl isomerase A(PPIA)	18	76.0	2
PCBP2_HUMAN	poly(rC) binding protein 2(PCBP2)	38.6	104.3	4
SSRP1_HUMAN	structure specific recognition protein 1(SSRP1)	81	464.9	11
HNRPQ_HUMAN	synaptotagmin binding cytoplasmic RNA interacting protein(SYNCRIP)	69.6	199.9	4

Appendix 3.10. STRING Functional enrichment table – Host-virus interaction (Gene Ontology). STRING analysis of Mass spectrometric analysis data of GFP-Ad12-9S following GFP-pulldown and SDS-PAGE. The table summarises the genes from our list that are associated host-virus interaction (n=22). Listed are the molecular weights (kDa), number of peptides, and the scores of each gene.

Appendix 3.11. STRING Functional Annotation Table – DNA repair

UNIPROT_ID	GENE NAME	MW [kDa]	Scores	No. of peptides
SP16H_HUMAN	SPT16 homolog, facilitates chromatin remodeling subunit(SUPT16H)	119.8	533.4	14
XRCC5_HUMAN	X-ray repair cross complementing 5(XRCC5)	82.7	517.8	12
XRCC6_HUMAN	X-ray repair cross complementing 6(XRCC6)	69.8	273.2	7
CAF1A_HUMAN	chromatin assembly factor 1 subunit A(CHAF1A)	106.9	27.1	1
DDB1_HUMAN	damage specific DNA binding protein 1(DDB1)	126.9	81.0	3
KAT7_HUMAN	lysine acetyltransferase 7(KAT7)	70.6	131.0	3
PARP1_HUMAN	poly(ADP-ribose) polymerase 1(PARP1)	113	40.3	1
PRP19_HUMAN	pre-mRNA processing factor 19(PRPF19)	55.1	71.0	2
SMC1A_HUMAN	structural maintenance of chromosomes 1A(SMC1A)	143.1	39.1	1
SSRP1_HUMAN	structure specific recognition protein 1(SSRP1)	81	464.9	11

Appendix 3.11. STRING Functional enrichment table – DNA repair (Gene Ontology). STRING analysis of Mass spectrometric analysis data of GFP-Ad12-9S following GFP-pulldown and SDS-PAGE. The table summarises the genes from our list that are associated DNA repair (n=10). Listed are the molecular weights (kDa), number of peptides, and the scores of each gene.

Appendix 3.12. STRING Functional Annotation Table - Other biological processes

UNIPROT_ID	GENE NAME	MW [kDa]	Score	No. of peptides	Process
PCBP1_HUMAN	poly(rC) binding protein 1(PCBP1)	37.5	57.4	2	Viral RNA replication
PCBP2_HUMAN	poly(rC) binding protein 2(PCBP2)	38.6	104.3	4	
RBM8A_HUMAN	RNA binding motif protein 8A(RBM8A)	19.9	27.4	1	Nonsense-mediated mRNA decay
IF4A3_HUMAN	eukaryotic translation initiation factor 4A3(EIF4A3)	46.8	101.8	3	
MGN_HUMAN	mago homolog, exon junction complex subunit(MAGOH)	17.2	69.7	2	
DHX9_HUMAN	DExH-box helicase 9(DHX9)	140.9	1561.6	34	Biological rhythms
TOP2A_HUMAN	DNA topoisomerase II alpha(TOP2A)	174.3	2724.8	55	
CBX3_HUMAN	chromobox 3(CBX3)	20.8	611.2	9	
DDB1_HUMAN	damage specific DNA binding protein 1(DDB1)	126.9	81.0	3	
GFPT1_HUMAN	glutamine--fructose-6-phosphate transaminase 1(GFPT1)	78.8	67.4	2	
HNRPD_HUMAN	heterogeneous nuclear ribonucleoprotein D(HNRNPD)	38.4	99.2	2	
DDX3X_HUMAN	DEAD-box helicase 3 X-linked(DDX3X)	73.2	169.4	6	Innate immunity
DHX15_HUMAN	DEAH-box helicase 15(DHX15)	90.9	319.0	8	
DDX21_HUMAN	DExD-box helicase 21(DDX21)	87.3	529.4	14	
DHX9_HUMAN	DExH-box helicase 9(DHX9)	140.9	1561.6	8	
RBM14_HUMAN	RNA binding motif protein 14(RBM14)	69.4	328.6	12	
XRCC5_HUMAN	X-ray repair cross complementing 5(XRCC5)	82.7	517.8	7	
XRCC6_HUMAN	X-ray repair cross complementing 6(XRCC6)	69.8	273.2	5	
ABC3B_HUMAN	apolipoprotein B mRNA editing enzyme catalytic subunit 3B(APOBEC3B)	45.9	244.1	4	
G3P_HUMAN	glyceraldehyde-3-phosphate dehydrogenase(GAPDH)	36	197.1	18	
MATR3_HUMAN	matrin 3(MATR3)	94.6	972.6	1	
PARP1_HUMAN	poly(ADP-ribose) polymerase 1(PARP1)	113	40.3	1	
PCBP2_HUMAN	poly(rC) binding protein 2(PCBP2)	38.6	104.3	2	
TMM33_HUMAN	transmembrane protein 33(TMEM33)	28	28.4	4	

Appendix 3.12. STRING Functional enrichment table – Other biological processes (Gene Ontology). STRING analysis of Mass spectrometric analysis data of GFP-Ad12-9S following GFP-pulldown and SDS-PAGE. The table summarises the genes from our list that are associated viral RNA replication (n=2), nonsense-mediated mRNA decay (n=3), biological rhythms (n=6), and innate immunity (n=13). Listed are the molecular weights (kDa), number of peptides, and the scores of each gene.