

**Characterization of the Host Transcriptional Response to Adenovirus E1A Deletion
Mutants**

Émilie Charette

Supervisor: Dr. Gregory Fonseca

Department of Medicine, Division of Experimental Medicine

McGill University, Montreal

February 2023

A thesis submitted to McGill University in partial fulfillment of the requirements of the
degree of Master of Science

© Émilie Charette 2023

Table des matières

ABSTRACT.....	IV
RESUME.....	VI
AUTHOR CONTRIBUTIONS.....	VIII
ACKNOWLEDGEMENTS	IX
LIST OF ABBREVIATIONS.....	X
LIST OF FIGURES.....	XIV
1 – INTRODUCTION.....	1
1.1 WHY STUDY ADENOVIRUS?	1
1.1.2 Adenovirus discovery and characterization	1
1.1.3 Human Adenovirus 5 (HAdV-5) – Characteristics and Disease	2
1.1.4 Adenovirus life cycle.....	3
1.1.5 Adenovirus Genome.....	5
1.1.6 The role of E1A in early infection	7
1.2 ONCOPROTEINS	9
1.2.1 The E1A gene.....	10
1.2.2 E1A as a hub protein	12
1.3 BAYLEY DELETION MUTANTS	18
1.4 TRANSCRIPTIONAL REGULATORY NETWORKS (TRNs)	21
1.4.1 Computational methods for reconstruction of TRNs	24
1.5 HYPOTHESIS AND AIMS	26
2 - METHODS	28
2.1 CELL LINES + VIRUS USED	28
2.2 INFECTION.....	29
2.3 RNA-SEQUENCING LIBRARY PREPARATION.....	29
2.4 PIPELINE FOR PROCESSING SEQUENCING DATA	36
2.5 BUILDING TRNs.....	36
3 – RESULTS	39
3.1 E1A DELETION MUTANTS HAVE DIFFERENT RNA EXPRESSION PROFILES	39
3.2 HIERARCHICAL CLUSTERING REVEALS ASSOCIATIONS BETWEEN MUTANTS IN SIMILAR REGIONS OF THE E1A GENE.....	45
3.3 CONSTRUCTING TRNs.....	48
3.4 EXPLORING THE DOWNREGULATED PHENOTYPE OF <i>DL1101</i> AND <i>DL1132</i>	53
3.5 EXPLORING THE UPREGULATED PHENOTYPE OF <i>DL1113-DL1116</i>	56
3.6 MOTIF ANALYSIS REVEALS POTENTIALLY IMPORTANT MOTIFS IN A GLOBAL MANNER	59
4 - DISCUSSION	61
4.1 E1A AS A MOLECULAR HUB PROTEIN.....	61

4.2 E1A CONTROL OF CELL CYCLE AS A TUMOR SUPPRESSOR AND ONCOGENE	62
4.3 VIRAL GENE EXPRESSION	64
4.4 TRN CONSTRUCTION	65
4.4.1 PA2G4.....	65
4.4.2 PITX1	66
4.4.3 EGR1.....	66
4.4.4 dl1101 vs. dl1132 – <i>The downregulated phenotype</i>	67
4.4.5 dl1113-dl1116 – <i>The upregulated phenotype</i>	69
4.5 TF BINDING TO DNA MOTIFS	70
4.6 FUTURE DIRECTIONS	71
5 – CONCLUSION	73
6 - BIBLIOGRAPHY	74

Abstract

Members of the Adenoviridae family are non-enveloped viruses which infect mammals and birds. In humans, more than 70 adenoviruses (HAdVs) of distinct serotype have been identified, and can infect a wide range of tissues and cause illness. Infections will most often settle within the upper or lower respiratory tracts, and are usually self-resolving. However, in certain cases, infection dissemination or pneumonia development can become fatal. As with other obligate intracellular pathogens, HAdVs rely on the host machinery for continued replication and infection. Early Region 1A (E1A) was identified as one of two oncogenes present in the adenovirus genome. This protein is expressed early in the viral life cycle and is implicated in the process of transformation by adenoviruses. Upon infection, E1A remodels the cellular environment to favour viral replication, modifying pathways involved in cell cycle regulation and transcription. Transcriptional regulatory networks (TRNs) are the networks which encode the instructions for animal development and physiological responses. These networks are highly complex, and have been shown to respond to stimuli in highly dynamic and unexpected ways. We are interested in the ways in which cells will adapt due to physiological dysregulation, and by which their TRNs may change. This can help us to learn more about the mechanisms behind disease processes, and allow us to therapeutically target specific regulatory steps. The E1A protein itself is known as a molecular hub protein, meaning it interacts with many structurally diverse targets, thus defining a specific pathway or phenotype. As E1A is unable to bind to DNA, it interacts with many different regulatory proteins and transcription factors to instill transcriptional changes benefitting viral replication. Although studies have been done regarding the effects of these mutations on viral replication and infection, little is known

about how these affect host cell transcriptional regulatory networks. Our objective is to reconstruct host cell TRNs, following infection with various adenovirus E1A deletion mutants. This will allow us to further characterize the effects of each of these deleted regions on the cellular transcriptional landscape. To carry this out, we used the Bayley *d11100* series of deletion mutants, a series of small in-frame deletion mutants of the adenovirus 5 E1A gene. We infected growth-arrested IMR90 cells with each mutant and isolated RNA for bulk RNA-Sequencing. Reads were then mapped to the human genome using the STAR genome alignment tool. Read counting was performed using the HOMER library, and differential gene expression analysis was done using the EdgeR package in R. TRNs were reconstructed using an elastic net regression model. We used a hierarchical clustering algorithm to sort mutants based on their gene expression profiles. Interestingly, mutants which had deletions in similar regions of the genome tended to cluster more closely together. We ran a KEGG pathway enrichment analysis to better understand the functional changes induced by the different mutants. We saw a glaring heterogeneity across all mutants, in terms of which pathways are differentially regulated and to what degree. Following selection of transcription factors and target genes, TRNs were reconstructed for each mutant. Comparison of these TRNs, and the contribution of transcription factors to each one, demonstrates, once again, a heterogeneity in the E1A gene and its functions. Using large scale bioinformatics approaches, we begun to elucidate downstream transcriptional regulation associated with regions of E1A. This will help us to more precisely differentiate E1A activities, facilitating a better understanding of host pathogen interactions.

Résumé

Les membres de la famille Adenoviridae sont des virus non-enveloppés qui infectent les mammifères et les oiseaux. Chez les humains, il existe plus de 70 adénovirus de différents sérotypes, et ceux-ci peuvent infecter une large gamme de tissus. Comme plusieurs autres pathogènes intracellulaires, les adénovirus se fient sur la machinerie cellulaire de leur hôte pour continuer leur réplication et infection. La protéine « Early Region 1 A » (E1A) fut identifiée comme étant un des deux oncogènes présents dans le génome adénovirus. Cette protéine est exprimée tôt dans le cycle de vie viral et est impliquée dans le processus de transformation. Suivant l'infection, E1A remodèle l'environnement cellulaire afin de favoriser la réplication virale, modifiant ainsi des voies de signalisation impliquée dans la régulation du cycle cellulaire et la transcription. Les réseaux de régulation transcriptionnelle (TRN) sont les réseaux qui codent les instructions pour le développement animal et les réponses physiologiques. Ces réseaux sont très complexes et il a été démontré qu'ils répondent aux stimuli de manière très dynamique et inattendue. Nous nous intéressons aux façons dont les cellules vont s'adapter en raison de la dérégulation physiologique et par lesquelles leurs TRNs peuvent changer. Cela peut nous aider à en savoir plus sur les mécanismes à l'origine des processus pathologiques et nous permettre de cibler de façon thérapeutique plusieurs étapes de régulation spécifiques. La protéine E1A elle-même est connue sous le nom de protéine de « hub moléculaire », ce qui signifie qu'elle interagit avec de nombreuses cibles avec diverses structures, définissant ainsi une voie ou un phénotype spécifique. Bien que des études aient été réalisées concernant les effets de ces mutations sur la réplication virale et l'infection, on sait peu de choses sur la façon dont elles affectent les réseaux de régulation transcriptionnelle des cellules hôtes. Notre objectif est

de reconstruire les TRNs chez les cellules hôtes, après infection par divers mutants de délétion E1A de l'adénovirus. Cela nous permettra de caractériser davantage les effets de chacune de ces régions supprimées sur le paysage transcriptionnel cellulaire. Pour ce faire, nous avons utilisé la série Bayley *d11100* de mutants de délétion, une série de petits mutants de délétion dans le cadre du gène E1A de l'adénovirus 5. Nous avons infecté des cellules IMR90 à croissance arrêtée avec chaque mutant et isolé l'ARN afin de l'envoyer pour le séquençage. Les lectures ont ensuite été alignées sur le génome humain à l'aide de l'outil d'alignement STAR. Le comptage des lectures a été effectué à l'aide de la bibliothèque HOMER et l'analyse différentielle de l'expression génique a été effectuée à l'aide de l'outil EdgeR. Les TRNs ont été reconstruits à l'aide d'un modèle de régression élastique net. Nous avons utilisé un algorithme de regroupement hiérarchique pour trier les mutants en fonction de leurs profils d'expression génique. En effet, les mutants qui avaient des délétions dans des régions similaires du génome avaient tendance à se regrouper plus étroitement. Nous avons effectué une analyse d'enrichissement des voies KEGG pour mieux comprendre les changements fonctionnels induits par les différents mutants. Nous avons constaté une hétérogénéité flagrante entre tous les mutants, en termes de régulation différentielle des voies et dans quelle mesure. Suivant la sélection des facteurs de transcription et des gènes cibles, les TRNs ont été reconstruits pour chaque mutant. La comparaison de ces TRN, et la contribution des facteurs de transcription à chacun, démontre, une fois de plus, une hétérogénéité du gène E1A et de ses fonctions. En utilisant des approches bio-informatiques à grande échelle, nous avons commencé à élucider la régulation transcriptionnelle associée aux régions de E1A.

Author Contributions

G. F. and E.C. designed the study.

G.F. performed the wet-lab experiments.

E.C. performed the computational analysis, with help from O.L. and G.F.

E.C. wrote the thesis with help from G.F.

Acknowledgements

First, I would like to thank my supervisor, Dr. Gregory Fonseca, for all the help and support he has shown me throughout the process of finishing this project. Coming into my graduate studies with a tight timeline, and high expectations for myself, Greg helped me to develop and achieve my goals. Working with him has made me a more astute, critical and hardworking scientist, and I am grateful to him for this.

Second, I would like to thank my fellow lab mates, Orsi and Mikal, for all their help with this project. Coming into this degree with very little computational knowledge, I relied on both heavily for their expertise. Additionally, the countless hours they both contributed in helping me reason through my ideas, and talk about my project were essential in allowing me to complete it.

Third, I would also like to thank my undergraduate project supervisor, Dr. Simon Rousseau, who was instrumental in igniting in me a scientific curiosity, and encouraging me to pursue graduate studies.

Lastly, on a more personal note, I would like to thank my parents and siblings, Pierre, Christine, Marie-Ève, Audrée-Anne and Pierre-Olivier, and my dogs, Mya and Billie, for their unyielding long distance love and support. I also would like to thank my closest friend, Iszy. Without you, I don't think I would have gotten through this degree. You've been there for me in the worst of times, and have helped me to believe in myself, both as a scientist and a person.

List of Abbreviations

3D	3 Dimensions
AA	Amino Acid
ADP	Adenovirus Death Protein
AMP	Adenosine Monophosphate
ARDS	Acute Respiratory Distress Syndrome
ATAC	Assay for Transposase Accessible Chromatin
ATF	Activating Transcription Factor
cAMP	Cyclic Adenosine Monophosphate
CAR	Coxsackie and Adenovirus Receptor
CBP	CREB Binding Protein
CDK	Cyclin Dependent Kinase
cGAS	Cyclic GMP-AMP Synthase
ChIP	Chromatin Immuno-Precipitation
CHR	Cell Cycle Gene Homology Region
CR	Conserved Region
CREB3L2	Cyclic AMP Responsive Element Binding Protein 3 Like 2
CtBP	C-Terminal Binding Protein
CUX1	Cut-like Homeobox 1
CYTH2	Cytohesin 2
DCAF7	DDB1 And CUL4 Associated Factor 7
DDR	DNA Damage Response
DEG	Differentially Expressed Gene
DGEA	Differential Gene Expression
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic Acid

dNTPs	Deoxynucleoside Triphosphohydrolase
dsDNA	Double Stranded Deoxyribonucleic Acid
dUTP	Deoxyuridine Phosphate
DYRK	Dual Specificity Tyrosine Regulated Kinases
E1A	Early Region 1 A
E1B	Early Region 1 B
ECM	Extracellular Matrix
EGR1	Early Growth Response protein 1
FBS	Fetal Bovine Serum
FDR	False Discovery Rate
FOXK	Forkhead Transcription Factor
GED	Graph Edit Distance
GMP	Guanosine Monophosphate
GO	Gene Ontology
HAdV	Human Adenovirus
HAT	Histone Acetyltransferase
hBre1	Histone H2B Specific Ubiquitin Ligase
HDAC	Histone Deacetylases
HEK	Human Embryonic Kidney Cells
hPaf1	Human RNA Polymerase II-Associated Factor 1
IFN	Interferon
IL	Interleukin
IRF3	Interferon Regulatory Factor-3
ISG	Interferon Stimulated Genes
ITR	Inverted Terminal Repeats
K/O	Knock-Out
KB	Kilobases

kDa	Kilo Dalton
KEGG	Kyoto Encyclopedia of Genes and Genomes
KLF10	Krueppel-Like Factor 10
MCP	Membrane Cofactor Protein
MDM2	Mouse Double Minute 2 Homolog
MED23	Mediator of RNA Polymerase II Transcription Subunit 23
miRNA	Micro Ribonucleic Acid
MLP	Major Late Promoter
MOI	Multiplicity of Infection
MoRF	Molecular Recognition Features
mRNA	Messenger Ribonucleic Acid
NADH	Nicotinamide Adenine Dinucleotide Hydrogen
NF1	Nuclear Factor 1
NFE2L2	Nuclear Factor Erythroid 2-related Factor 2
NFY	Nuclear transcription Factor Y
NLS	Nuclear Localization Signal
NPC	Nuclear Pore Complex
NuA4	Nucleosome Acetyltransferase of H4
OCT-1	Organic Cation Transporter 1
ODE	Ordinary Differential Equations
P.I.	Post-Infection
P/S	Penicillin/Streptomycin
PA2G4	Proliferation Associated protein 2G4
PCA	Principal Component Analysis
PCAF	p300/CBP-Associated Factor
PD	Packaging Domain
PITX1	Paired-like Homeodomain 1

PRZ	Peripheral Replicative Zone
pTP	Adenovirus Terminal Protein
RAS	Rat Sarcoma
RB1	Retinoblastoma Protein
RCAd	Replication Competent Adenovirus
RDAd	Replication Deficient Adenovirus
RNA	Ribonucleic Acid
RPKM	Reads per Kilobase of exon per Milion Reads Mapped
RuvBL1	RuvB-like 1
SMAD3	Suppressor of Mothers Against Decapentaplegic homolog 3
STAT1	Signal Transducer and Activator of Transcription 1
STING	Stimulator of Interferon Genes
SWI/SNF	SWItch/Sucrose Non-Fermentable
TBP	TATA Box Binding Protein
TCF21	Transcription Factor 21
TERT	Telomerase Reverse Transcriptase
TF	Transcription Factor
TGF	Tumour Growth Factor
TLR	Toll-like Receptor
TNF	Tumour Necrosis Factor
TRN	Transcriptional Regulatory Networks
TRRAP	Transformation/Transcription Domain Associated Protein
UBC9	Ubiquitin Carrier Protein 9
Ube2b	Ubiquitin-Conjugating Enzyme 2B
VAI/VAII	Virus Associated RNA I/II
WT	Wild Type
ZBTB20	Zinc Finger And BTB Domain containing 20

List of Figures

Figure 1: Adenovirus transcription

Figure 2: Adenovirus binding partners

Table 1: Known Information about the dl1100 series of Deletion Mutants

Figure 3: Expression profiles of E1A mutants

Figure 4: Clustering E1A deletion mutants

Figure 5: TRN reconstruction and comparison

Figure 6: Exploring *dl1101* and *dl1132* TRNs

Figure 7: Exploring *dl1113-dl1116* TRNs

Figure 8: Motif analysis for all mutants

1 – Introduction

1.1 Why study adenovirus?

Within human cells, most proteins are responsible for interactions with one or two other proteins, being enmeshed in complex processes. However, there are a few proteins which can interact with tens or hundreds of other proteins, thus having control over a module of interaction¹. This is a feature ubiquitous to many other small oncogenic DNA viruses, including adenovirus' Early Region 1A protein². Due to the small packaging capacity of many viruses, their encoded genomes are very small. As such, it becomes important for the few genes expressed to hold multiple functions. Some of these proteins can interact with several host factors, which can carry out functions essential to viral survival³. Hub proteins within cells are prime targets during viral infection as they are responsible for rewriting entire cellular processes through the targeting of cellular hub proteins. Therefore, by studying adenovirus' E1A protein, we can begin to identify global transcriptional phenotypes and cellular molecular hub proteins.

1.1.2 Adenovirus discovery and characterization

It was in 1953 that adenoviruses were first isolated as novel infectious agents in samples from military recruits. Rowe and colleagues were studying the growth of polioviruses in adenoidal tissue, hence the name eventually awarded to these viruses⁴. Belonging to the Mastadenovirus genus and *Adenoviridae* family, adenoviruses are large non-enveloped viruses with icosahedral capsids⁵. They possess a linear double-stranded (ds)DNA genome of length 34-43 kilobases (kb). This genome is enclosed in an inner core, along with many histone-like proteins^{4,6,7}. Adenoviruses infect hosts among all vertebrates, including birds and mammals⁷. Over 50 human serotypes have been identified⁶. Human

Adenoviruses (HAdV) have been classified into 7 different species (A-G)⁵. Infection is prevalent worldwide and can occur during any season, with particularly susceptible populations being young children and immunocompromised individuals, as well as military training camps and crowded communities⁴. Tropism and different manifestations of the infection are associated with varying serotypes⁶. Adenovirus is transmissible by aerosol droplets, the fecal-oral route and contaminated fomites⁴. It can survive for extended periods of time on environmental surfaces and is resistant to lipid disinfectants. However, it can be inactivated by heat, formaldehyde and bleach⁴. One of the findings which has propelled adenoviruses into the public eye is that serotype 12 (HAdV-12) can transform rodent cells⁸. While some serotypes can cause oncogenesis and transform rodent cells in culture, several adenovirus serotypes were found to lack tumorigenic properties. This includes the most widely studied serotypes, HAdV-2 and HAdV-5. There are many factors which affect the tumorigenic ability of the virus, namely virus dose, host genetics and immune status, as well as age at exposure⁹. For instance, it was shown that cells transformed by non-oncogenic adenoviruses had tumorigenic potential, but only in immunocompromised hosts¹⁰. Importantly, oncogenesis in human cells has yet to be demonstrated¹¹.

1.1.3 Human Adenovirus 5 (HAdV-5) – Characteristics and Disease

Most adenovirus infections are self-limiting, and will rarely cause serious complications in adults and healthy children. On the other hand, in immunocompromised hosts, dissemination is more likely and disease can become severe⁵. Risk factors for severe infection include the following: age under 7 years, a history of chronic disease and a recent organ transplant¹². Fatality rates for untreated severe disseminated HAdV infection are

estimated to exceed 50%⁵. This is especially worrying in underdeveloped countries, in which acute respiratory infections remain a major cause of morbidity and mortality¹³. In the last few years, several countries in Asia have seen outbreak numbers increase, with presentation of acute respiratory distress syndrome (ARDS) becoming a major worry for patients infected with adenovirus⁴.

There are a series of antiviral drugs which can be used to combat adenovirus infections. Cidofovir is the most commonly used in this context. It is a nucleotide analog of cytosine which selectively inhibits viral DNA polymerase, thus halting viral replication⁴. While Cidofovir increases overall survival, it is only available intravenously and dosing regimens are highly variable, making it difficult to standardize clinical practice⁵. More research is needed to identify agents which have more specific and targeted action. In this work, the serotype of interest is HAdV-5, which is typically associated with infection in immunocompromised patients, specifically those that have undergone either hematopoietic stem cell, or solid organ transplants⁵. The study of the HAdV-C species, including HAdV-5 has allowed for the discovery of virus host interactions as well as host cell molecular processes. More specifically, this has led to a deeper understanding of RNA splicing, DNA replication processes within human cells, protein folding with a molecular folding chaperone (100K protein), and more^{7,14}.

1.1.4 Adenovirus life cycle

The infectious cycle of type C human adenoviruses in permissive cells in culture is characterized by a time-dependent program of viral gene expression, leading to viral DNA replication, and structural protein translation⁶. In HEK 293 cells and at an MOI of 50, this takes approximately 36 hours, with progeny virion appearing around the 24-hour mark

post-infection (p.i.)¹⁵. Adenovirus infection begins with the attachment of elongated fibers to one of several host cell receptors. At the end of each of these protruding fibers are pentameric structures which form globular knob domains. It is this structure which is directly involved in receptor binding¹⁶. The most common receptor targets are the CAR (coxsackie and adenovirus receptor) and MCP (membrane cofactor protein)^{4,17}. These receptors are not used simultaneously by most adenovirus groups. Species C adenovirus particles typically enter the cells through receptor-mediated endocytosis via clathrin-coated pits. After endocytosis, acidification of the endosome leads to activation of viral proteases, resulting in partial disassembly of the capsid. A quick escape from endosomes into the cytoplasm ensures virus particles will not be degraded^{18,19}. The mature viral particle is then translocated in a dynein-dependent manner along microtubules throughout the cytoplasm. As it makes its way to the nuclear pore complex (NPC), the virus undergoes slow uncoating¹⁸. It is necessary for viral genomes to be associated with a protein containing a nuclear localization signal (NLS) in order for them to be properly targeted to the NPC. This is the case with core protein VII, the most abundant of the core proteins which contains three separate NLS structures, allowing for efficient translocation into the nucleus²⁰. Protein VII is tightly associated with the viral genome in a sequence independent way, keeping the DNA packaged in nucleosome like structures. The protein VII-wrapped adenoviral genome will then enter the nucleus⁴. Following this translocation, adenovirus gene expression and genome replication will take place. These processes will be further explored in the *Adenovirus Genome* section of this work. Viral particle assembly will occur in the nucleus, within the previously recognized peripheral replicative zone (PRZ). In the concurrent model for adenovirus packaging, there is one pathway by which capsid proteins

are synthesized and assembled and a second pathway for core proteins. These pathways are coupled to one another to ensure viral particle production, in which capsids are observed to grow around cores. Packaging of the viral genome starts at its left end, where the packaging sequence (Ψ) is located. Packaging proteins will aggregate and bind to Ψ , and, by the action of the core proteins, the genome will condense¹⁵. After assembly, virion progeny are released through cell lysis. This is facilitated by the adenovirus death protein (ADP), which localizes to the inner nuclear membrane, where it will destabilize the nuclear envelope and promote membrane rupture and, as a consequence, cell lysis²¹.

1.1.5 Adenovirus Genome

As mentioned previously, adenovirus virions are characterized by a nucleoprotein core which contains a linear dsDNA genome of length between ~34-43 kb. Despite large differences in host and tissue tropism among different adenovirus family members, there is little variation in terms of genomic parameters. HAdV-5 contains a genome of ~36kb in length. There are over 40 proteins encoded in the open reading frames of this genome. The open reading frames are labelled as early (E1-E4) and late (L1-L5) depending on whether they are expressed before or after DNA replication²², as shown in Figure 1.

Both ends of the genome contain inverted terminal repeats (ITR) of ~100bp, which act as the origin of replication. Between the left ITR and the E1A transcriptional start site, there is a packaging domain (PD), which is comprised of 7 AT-rich repeats. This PD is important to encapsulation of the viral DNA²³. The adenovirus terminal protein (pTP) is covalently linked to both 5' ends of the genome and plays an important role in the initiation of viral DNA replication. The capsid proteins are composed of 3 major proteins and 5

minor proteins. Of these, proteins V, VII and X/mu are responsible for condensing the adenoviral DNA and mediating interactions between the core and capsid²⁴.

Like most obligate intracellular parasites, adenoviruses must efficiently use their

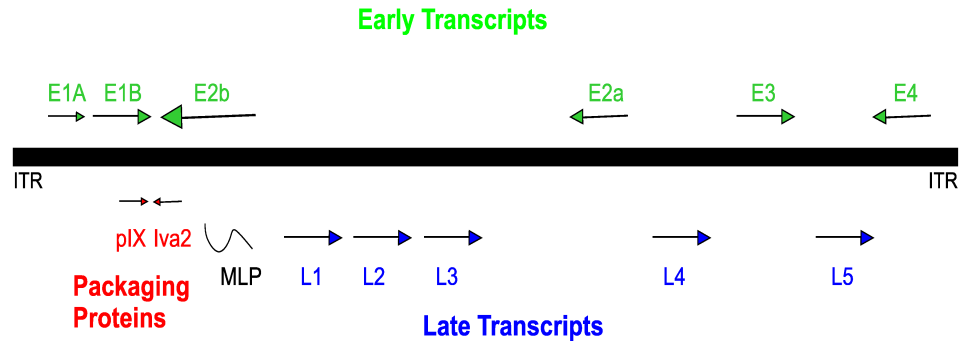


Figure 1: Adenovirus transcription. Map of adenovirus genome with early (green) and late (dark blue) transcripts. Packaging proteins identified in red.

limited genetic space. For this reason, expression of viral genes is tightly regulated to ensure temporal progression of viral genes in a manner which allows progression through the viral life cycle²⁵. The early regions E1A, E1B, E2, E3 and E4 are the first to be transcribed. Their functions mainly revolve around transactivation of other viral regions and modification of the cellular environment to promote viral production. After the start of viral DNA synthesis, the major late promoter (MLP) region is activated, allowing expression of the late genes. These late regions (L1-L5) are transcribed from an alternatively spliced transcript and primarily encode structural proteins which will form the replicated virion. There are also four small proteins which are produced in the later stages of infection. These include the structural protein IX (pIX) and the IVa2 protein, which help package viral DNA into immature virions, as well as the late virus-associated RNAs I and II (VA RNA), which are hypothesized to dampen the host immune response and influence host gene expression^{26,27}.

Adenovirus genome replication is important to viral fitness and propagation and although this is facilitated by several host factors, the virus itself also encodes some necessary proteins. Early genes encode the pTP, DNA polymerase and DNA-binding protein, which together form the viral replication machinery²⁵. The ITRs at both ends of the genome possess the origin of replication. This is separated into the core origin sequence, containing a binding site for pTP and the adenovirus polymerase, as well as the auxiliary origin which contains binding sites for the cellular TFs nuclear factor 1 (NF1) and Oct-1²⁸. Experiments have demonstrated that while viral replication machinery proteins are necessary for genome replication, both NF1 and Oct-1 are not required but will greatly enhance the efficiency of transcription. The replication itself occurs through a mechanism of strand displacement, which is common in viral genome replication²⁹.

1.1.6 The role of E1A in early infection

There are several features of human adenoviruses which are recognized by nuclear and cytoplasmic sensing, leading to innate immune responses. To ensure continued infection, viruses must have mechanisms by which to counteract the host immune response. These will vary depending on the cell type which is infected. In several human cell lines, adenovirus infection will lead to activation of the cytoplasmic cyclic GMP-AMP stimulator of interferon genes³⁰. This will result in IRF3 activation and IFN- β gene expression³⁰. Sensing of viral DNA through the endosomal sensor TLR9 can lead to adenovirus induced IL-1 β release³¹. In primary lung fibroblasts, a DNA sensing TLR mechanism leads to the stimulation of IRF3-mediated IFN and a proinflammatory response³². The adenovirus genome contains two main features which are important in immune sensing. First, its linear genome has open ends, which resemble dsDNA breaks, which will be sensed by the DNA

damage response (DDR) sensor complex³³. Second, the pTP which is covalently attached to each 5' end of the adenovirus genome is read by the cell as a DNA-protein adduct and is also be targeted by the DDR repair mechanisms³⁴. As viral DNA is coated with protein VII during the early stages of infection, this acts as a protective mechanism to avoid DDR signaling. Another mechanism by which cells can target DNA viruses is the cGAS-STING pathway, in which cGAS detects the presence of cytosolic DNA³⁵. Signalling through the adapter STING protein initiates the antiviral response. The LXCXE motif of E1A (also found on several other viral oncogenes) can bind to STING, hereby antagonizing viral DNA sensing³⁵. Viral gene expression within the first, and early, stage of viral replication is set to accomplish several functions, one of which being to produce proteins which can counteract cellular and host immune responses to the virus.

E1A, being the first protein which is produced upon viral infection, plays an important role in modulating host immune functions to allow for viral survival and continued replication¹⁷. It has been shown to use a variety of mechanisms to block IFN signaling pathways and their induction of ISG signaling³⁶. For instance, E1A can inhibit ISG gene expression through modulation of the RuvBL1 protein³⁷. It binds this protein through its C-terminus³⁷. In several cancer cell lines, E1A prolongs STAT1 phosphorylation, and promotes its recruitment to viral DNA replication centers to sequester it³⁸. However, this appears to contrast with observations in primary human airway epithelial cells, in which E1A appears to lead to STAT1 phosphorylation downregulation in later infection timepoints³⁹. In a similar way, E1A's N-terminal region has also been shown to bind to the cellular protein hBre1, blocking it from interacting with the catalytic subunit of the ubiquitin-conjugating enzyme Ube2b⁴⁰. Because of this, H2B

monoubiquitination at the *ISG* locus will not occur. ISG gene expression is subsequently blocked⁴¹. In a demonstration of its duality and efficiency, E1A also uses its binding to hBrel to recruit the RNA Polymerase II transcription elongation factor, hPaf1, stimulating early viral gene expression⁴². E1A CR4 binding proteins also hold important roles in immune evasion. E1A needs to bind all three proteins: FOXK, DCAF7 and CtBP in order to properly suppress activation of ISGs⁴³. It is suggested that E1A regulates the stabilities of several key proteins in this pathway, such as IRF3⁴³. Overall, E1A proteins have been shown to, first block cytoplasmic IFN signaling pathways, and second, downregulate cellular mechanisms meant to promote ISG expression in cells. Although its main function has always been assumed to be the activation of early gene expression, E1A is a major player in immune evasion.

1.2 Oncoproteins

The field of oncology is centered around understanding the cellular mechanisms which lead to the transformation of a normal cell into a cancer cell⁴⁴. The genes which hijack these cellular pathways and cause cancer are known as oncogenes⁴⁵. Throughout history, oncogenes have been defined by their ability to transform a monolayer of cells in culture⁴⁶. They significantly contribute to many if not all steps of cancer formation⁴⁴. The original discovery and characterization of these oncogenes stems from the study of tumor viruses, such as Rous Sarcoma Virus⁴⁷. It was long believed that for each transformation inducing oncogene, there was a homolog proto-oncogene; a normal cellular gene involved in healthy cell processes. It is a mutation of this normally functioning proto-oncogene which will lead to dysfunctional behavior within cancerous cells⁴⁸. However, what is most striking through the study of these proteins is that they don't cause cancer in all tissues,

and may not transform all cell types in the same way, or to the same degree⁴⁹. There are both intrinsic and extrinsic factors which affect an oncogene's ability to induce a cancerous phenotype. Intrinsic factors can include gene dosage, loss of the WT allele, and cell lineage⁵⁰. Extrinsic factors can include the tissue architecture and microenvironment as well as cell-cell competition⁵¹⁻⁵³. It is hypothesized that the ability of an oncoprotein to cause transformation in cells must have something to do with the pre-existing transcriptional and proteomic landscapes. The baseline epigenetic programming of cells can thus inform how the oncoproteins will act within different tissues and cell types⁴⁹. Oncoproteins will typically affect signaling related to the cell cycle and mitotic pathways, in order to induce a favorable state of continuous replication. However, not all mutations to these pathways will lead to cancerous transformations⁵⁴. This underlines the importance of recognizing the versatility of oncogenes and their roles within cells.

1.2.1 The E1A gene

The first viral protein-coding gene to be expressed following adenovirus infection is Early Region 1A (E1A)^{55,56}. This gene is composed of 2 different exons which splice differentially in order to give rise to the 5 different protein isoforms⁵⁷. The two most highly abundant during early infection are the largest 13S and 12S, of 289 and 243 amino acid residues in HAdV5, respectively⁵⁸. However, in later infection, it is one of the smaller isoforms, the 10S of length 171 amino acid residues, which dominates⁵⁷. The main role of the proteins encoded by this transcript is to activate viral transcription and reprogram the host gene expression to provide an ideal environment for virion production⁵⁵. The major E1A isoform is shown in Figure 2²⁴.

The four conserved regions are illustrated in blue (conserved region 1 – CR1), pink (conserved region 2 – CR2), yellow (conserved region 3 – CR3) and orange (conserved region 4 – CR4). The first exon leads from the first amino acid to the start of CR3, while the second exon leads from the end of CR3 to the last amino acid in the sequence. These conserved regions were identified based on high amounts of sequence conservation among primate adenovirus subgroups and species⁵⁹. CR1, CR2 and CR4 are common to both isoforms, while CR3 is unique to the larger E1A isoform. Several independent studies have begun to reveal the functionalities of each of these regions, although there is still much which remains to be discovered. Furthermore, E1A doesn't have the ability to bind DNA, due to its lack of a DNA binding domain. As such, it needs to interact directly with cellular factors and proteins to enact broad changes within cells⁶⁰. Some of these binding partners are illustrated in Figure 2. They will be discussed in later sections of this work. However, it should be noted that the binding illustrated in Figure 2 is a simplified visual

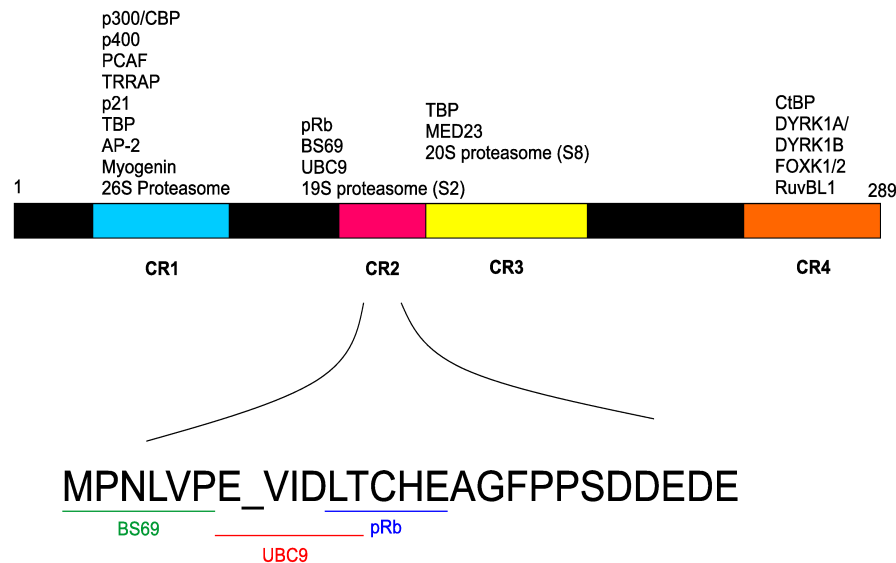


Figure 2: Adenovirus Binding Partners. 13S isoform of the E1A protein is pictured, with the 4 conserved regions identified by color. Known binding partners of E1A are listed above the specified region. The AA sequence of CR2 is blown up to illustrate binding sites for BS69, UBC9 and pRb.

representation of the molecular interactions which occur. Several of these proteins require interactions with multiple regions of E1A for optimal function⁶¹.

1.2.2 E1A as a hub protein

An important component of hub proteins of which E1A is no exception, is intrinsic structural disorder. To enable the high connectivity which is associated with these proteins, the concept of a fixed 3-dimensional structure with high affinity to each of its targets is not sensible. In fact, proteins may be completely unstructured, a concept called intrinsically disordered, or may combine both the structure of globular domains with interspersed disordered regions^{62,63}. This means that these proteins lack a rigid 3D structure under physiological conditions and only transiently form local structures upon protein interaction⁶⁴⁻⁶⁶. These interactions are mediated by short linear segments, known as molecular recognition features (MoRFs)⁶⁴. This lack of a rigid structure is evidenced in Figure 2, in that several proteins have overlapping binding sites on E1A. This type of regulation would not be possible without a malleable protein structure.

Looking at the functions of E1A under the lens of a molecular hub protein, it becomes evident that this protein regulates several signaling pathways within cells, mostly through the modulation of transcription. E1A is responsible for activating viral gene replication, forcing quiescent cells to re-enter the cell cycle and blocking cell differentiation⁶⁷. Additionally, it is responsible for suppressing the host immune response to allow for continued infectivity⁶⁸. E1A mediates completely different infection cycles in different hosts. While in human hosts, the infection is lytic, resulting in death of the host cell and release of the viral progeny, infection in rodent cells is non-productive and does not lead to cell death. Instead, it can cause tumor formation⁹. E1A's ability to promote

vastly different regulatory programmes, depending on the molecular context highlights its versatility. There are currently over 50 proteins, many of which are hub proteins themselves, which have been reported to interact with E1A, demonstrating its clear role as a molecular hub protein⁶⁹.

Shedding light on more of these important interactions, specifically with TFs, which themselves control gene expression, can reveal important cellular mechanisms driving molecular pathways surrounding cell growth, differentiation, immunity and transcription.

1.2.2.1 Conserved Region 1

CR1 and CR2 have functions related to induction of the cell cycle and entry into S phase. CR1 is known to bind several proteins involved in chromatin modification, including p300 and its analog CBP. This binding is bipartite, which means that two sections of the E1A gene concurrently bind to p300/CBP, namely the non-conserved N-terminus, and the FPDSVML sequence in CR1⁷⁰. This interaction likely inhibits and redirects the histone acetyltransferase (HAT) activity of p300/CBP through acetylation of histone tails and lysines. E1A has been found to promote p300 acetylation of RB1 K873/K874, locking it into a repressive conformation which interacts with repressive chromatin-modifying enzymes⁷¹. This p300-E1A-RB1 complex interacts with specific pathways such as the TGF- β , TNF and interleukin signaling pathways, repressing signaling which would normally hinder viral production⁷¹. It is hypothesized that the repression of selective HAT activity is the event which leads to entry into S phase, although the exact mechanism by which this takes place is lesser known. Parallels can be drawn between this mechanism as well as several human non-viral cancers, which start with somatic p300/CBP mutations⁷².

CR1 is also shown to interact with several other chromatin modifying complexes. This includes p400, similar to the SWI/SNF family of chromatin-remodelling complexes, the HAT PCAF as well as TRRAP, a member of several human HAT complexes. TRRAP bridges the interactions between HATs and other transcriptional regulators. E1A's association with TRRAP allows for the assembly of acetyltransferase complexes, such as the NuA4 complex⁷³. Moreover, E1A's interaction with p400 leads to a stabilization of the interaction between p400 and Myc at the site of Myc target genes, upregulating their expression. This is important to E1A's ability to induce apoptosis, helping to maintain normal cell growth in the presence of overwhelming positive growth signals⁷⁴. Additionally, it is thought that this interaction could also promote the displacement of p54 from p400 and the subsequent loss of p54's helicase activity⁷⁵. CR1's binding to the PCAF histone acetyltransferase inhibits its HAT activity, leading to repression of PCAF-mediated activation of transcription. This interaction blocks the induction of the myogenic pathway⁷⁶. Additionally, CR1 binds the CDK inhibitors p21 and p27, releasing inhibition of the cyclins necessary for entry into the cell cycle. Specifically, p21 binds to and stabilizes cyclin D expression. Releasing this inhibition promotes a favorable environment for viral replication⁷⁷. Additionally, E1A allows continued cell proliferation in an environment which is extremely high in p27 by targeting specific effectors of its target, CDK2⁷⁸. It can also bind to and inhibit several components of the 26S proteasome, which in turn contributes to p53 stabilization⁷⁹. Although several of the important binding partners for CR1 have been mentioned, to date over 15 different partners have been identified. These also include AP-2⁸⁰, myogenin⁸¹, TR⁸² and TATA-box binding protein⁸³, among others.

1.2.2.2 Conserved Region 2

CR1 is essential to the driving of the cell cycle, however it requires the action of CR2 to initiate this signaling⁸⁴. This second conserved region contains an LxCxE motif, allowing it to bind to RB as well as its related proteins p107 and p130⁸⁵. This ability to sequester pRb allows for the release of E2F, thus forcing G₀ cells into S phase⁸⁶. It has been found that the mechanism by which this occurs is that, in binding pRb to the CR2 region, E1A increases the local concentration of pRb. This effectively means that the lower affinity CR1 pRb binding domain can bind to pRb and replace E2F, leading to maximal E2F displacement⁸⁷. The CR1 domain binds to the same region of pRb as does E2F, thus competitively inhibiting its binding. Given that HAdV-5 has evolved to infect mainly growth-arrested epithelial cells in the respiratory tract, it has developed mechanisms by which to overcome stable growth arrest. In normal growth arrested cells, RB proteins bind multisubunit histone deacetylase complexes in order to block E2F activation⁸⁸. E2Fs are responsible for activating the transcription of several important cell cycle related genes, including late G₁ and S phase cyclin-dependent kinase CDK2 and cyclins E and A⁸⁸. The p300 protein, which also binds E1A can acetylate pRb as well. This acetylation has been shown to promote its interaction with the MDM2 protein, an interaction partner of p53. This association inhibits the p53 degradation through the proteasome, and allows for transcriptional repression and induction of apoptosis⁸⁹. Other than this important activation function, CR2 is predicted to be largely unstructured⁹⁰. It has three other known targets: BS69, UBC9 and the S2 subunit of the 19S proteasome. CR2's interaction with the transcriptional repressor BS69 is less well characterized. It is reported to stabilize E1A by blocking ubiquitination, while also inhibiting transcriptional activation by E1A CR3. This may serve as the means by which E1A can control viral gene expression in the process of

infection⁹¹. Moreover, CR2 can also bind to the small ubiquitin-like moiety (SUMO) conjugase UBC9, which is likely a way by which E1A can control polySUMOylation of many important cellular factors, such as its own interacting partners, CtBP, p300/CBP and pRb⁹². Lastly, CR2 can bind the proteasomal regulatory S2 subunit, leading to proteasome inhibition as well as increased p53 expression, as it is not being degraded⁹³.

As many of the functions of both the CR1 and CR2 regions revolve around the activation of the cell cycle, it is important to understand that a such sustained expression of E2F proteins will lead to apoptotic mechanisms within the cell. Adenovirus requires expression of the E1B protein to maintain cell cycling. E1B has functions involving sequestering of p53, thus snuffing out the pro-apoptotic signaling⁹⁴.

1.2.2.3 Conserved Region 3

Due to alternative splicing, the third E1A conserved region, CR3, is only present in the largest 13S isoform of the protein. This region is slightly more structurally ordered than the others, as it encodes a conserved zinc finger. Its transcriptional activation domain is crucial to the activation of early viral gene expression⁹⁵. Unlike the other conserved regions, most deletion mutants created in CR3 result in a protein which is unable to transactivate⁹⁶. Its main role is to promote the assembly of a transcriptional initiation complex by recruiting E1A to the template DNA using DNA-binding TFs. It can then recruit multiple important transcriptional regulators⁹⁷. Two of its main interaction partners are TATA-box binding protein (TBP) and MED23, which is part of the mediator adapter complex. Both proteins interact with the zinc finger subdomain of this protein, and retain functions important in allowing E1A to nucleate the transcriptional preinitiation complex⁹⁸. Furthermore, as E1A has no DNA binding regions of its own, it relies on cellular sequence-

specific DNA binding TFs in order to be recruited to promoter regions. These can include TFs from the cAMP response as well as from the ATF family⁹⁹. CR3's interactions with two major subunits of the proteasome, S8 and 20S are also important and are posited to have an impact on transcriptional initiation and elongation. It is likely that the efficiency of transcription is dependent on E1A, more specifically CR3's stability¹⁰⁰.

1.2.2.4 Conserved Region 4

Much less is known about the last conserved region, CR4, which occupies the C terminus region of the E1A peptide, and contains a bipartite nuclear localization signal¹⁰¹. This region is required for oncogenic transformation of rodent cells with E1B¹⁰², while also being responsible for suppression of transformation when in the presence of activated *ras*¹⁰³. CR4's interaction with CtBPs, cellular transcriptional corepressors, are the most highly documented. CtBP will interact with E1A through its conserved PLDLS domain, and is highly important in *ras* dependent transformation, the exact mechanism for which is not yet well known¹⁰⁴. The CtBP protein itself dimerizes upon NADH binding and will direct the formation of a silencing complex at specific promoters, after being recruited by a TF. This repressive phenotype is largely mediated through the recruitment of histone deacetylases (HDACs) or the heterochromatin-interacting protein Polycomb-2. It is primarily responsible for the repression of pro-apoptotic and epithelial gene expression pathways. In fact, the interaction of E1A with CtBP is said to sensitize cells to anoikis, which is the triggering of apoptosis in cells after losing contact with the ECM¹⁰⁵. By competing with other cellular factors containing the PLDLS conserved binding region, E1A can effectively block CtBP targeting and repression¹⁰⁶. The identity of the specific genes which are expressed upon this interaction are currently not known, however, studies

have found that CtBP's interaction with E1A is important to blocking the expression of IFN-stimulated genes (ISGs)⁴³. Additionally, it is proposed that other CtBP targets play important roles in suppressing the immortalization of primary tumor cells, and enhancing virus replication during productive infection¹⁰⁷.

CR4 has been reported to interact with the DYRK1A and DYRK1B kinases, and to stimulate their activity. These are involved in regulation of cell proliferation, survival and differentiation. Recent studies have also demonstrated the importance of the DCAF7 protein in mediating E1A's interactions with these proteins, particularly DYRK1A¹⁰⁸. Interactome studies have shown that all major E1A binding partners are also interacting with DYRK1A and DYRK1B⁹⁷. As negative regulators of the cell cycle, interaction with these kinases is reported to have a suppressive effect on the proliferation and transformation of adenovirus-infected cells¹⁰⁹. This interaction, and the phosphorylation of E1A by DYRK1A is proposed to lead to its interaction with the FOXK1/2 TFs, which work to inhibit oncogenic transformation of cells¹⁰⁸. Lastly, CR4 is reported to interact with the RuvBL1 protein as a mechanism by which it can suppress interferon-stimulated gene expression, thus dampening the immune system³⁷.

1.3 Bayley Deletion mutants

In an effort to better systematically examine the E1A protein domains, a series of in-frame deletions and missense point mutations have been created in the E1A gene⁹⁶. Together, these mutations cover the entire gene coding region⁹⁶. These mutants were generated in the 1990s and have allowed researchers to identify a region which is important to transactivation, that is the region unique to the 13S E1A isoform^{59,96}. These mutants were created in a *dl309* background. The *dl309* mutant expresses the WT E1A gene, but has a

deletion within the E3 gene, meaning it loses the RID and 14.7K proteins¹¹⁰. Creation of the mutants *dl 1101-1107*, *1109*, *1112-1116* and *1119* was carried out through oligonucleotide mutagenesis using the two primer method elaborated by Zoller and Smith¹¹¹. Construction of mutant *dl1110* was carried out through deletion loop mutagenesis, which removed 63 amino acids, including the 5' splice site for the 12S isoform, which cannot form the corresponding product¹¹². Mutant *dl1108* was later constructed to remove the residues between *dl1107* and *dl1109*. This was an attempt to eliminate binding to the p107 and pRb proteins. The mutant *dl520* produces the E1A 12S and not the larger 13S isoform¹¹³. The mutant *dl975* synthesizes a protein which is translated only from the E1A 12S mRNA. It is missing the first 36 amino acids encoded by the N-terminal second exon¹¹⁴. The deletion mutants *dl1132-1136* were prepared similarly to the *dl1101-dl1107*, but in the laboratory of Boyd *et al.* and aimed to investigate the CR4 and C-terminus region¹¹⁵. Lastly, mutants *dl1141-1142* were created similarly to *dl520*⁸⁴.

Previous studies have looked at the viral kinetics of these mutants. In one such study, deletion mutants in the second exon (or C-terminus) of the protein were shown to be deficient in growth, induction of S-phase, viral protein and gene expression as well as genome replication¹¹⁶. This applied not only to deletions found within the last conserved region (CR4), but also to those which were found between conserved regions¹¹⁶. In another study, mutations found in the first exon, or N-terminus region, of the protein were studied⁵⁶. In this case, almost all deletion mutants were found to have a negative impact on viral replication, except for *dl1106*, which is a deletion of amino acids 90-105⁵⁶.

Table 1: Known Information about the dl1100 series of Deletion Mutants. This table illustrates known information about viral kinetics and phenotype of the mutants used in this study. In cases in which information is unknown, cells are marked N/A. The “+” designation indicates the strength of a given effect, with “++++” indicating a very strong effect.

Deletion Mutant	Mutation	Viral Growth ⁵⁶	Viral Gene Expression ⁵⁶	Other Notes ^{79,96,115-117}
<i>dl309</i>	WT E1A	+++	++++	Has deletions in the E3 gene, namely the 10.4K, 14.5K and 14.7K genes are deleted, while the 6.7K gene has a two-AA deletion. All subsequent mutations are done in a <i>dl309</i> background.
<i>dl1101</i>	Δ4-25	++	+	Does not transform with <i>ras</i> , induce mitosis, repress SV40 transcription, block myogenic differentiation, induce differentiation in P19 cells, induce cytogenetic damage, induce <i>c-fos</i> expression in synergy with cAMP, block IFN stimulated gene expression or cytolytic resistance, induce a G1 growth arrest in yeast or bind p400 or p300.
<i>dl1102</i>	Δ26-35	+++	++++	Does not bind p400, but appears otherwise WT.
<i>dl1103</i>	Δ30-49	+	+	Does not transform with <i>ras</i> , induce mitosis, repress SV40 transcription, or block myogenic differentiation. Has a reduced ability to induce <i>c-fos</i> expression in synergy with cAMP and to induce G1 growth arrest in yeast. Does not bind to p300 or p400
<i>dl1104</i>	Δ48-60	++	++	Does not transform with <i>ras</i> , induce mitosis, repress SV40 transcription, block myogenic differentiation or induce differentiation in P19 cells. Does not induce <i>c-fos</i> expression in synergy with cAMP. Inability to block IFN stimulated gene expression or cytolytic resistance. No induction of growth arrest in yeast. No binding to p300.
<i>dl1105</i>	Δ70-81	++	+++	Appears WT, with the exception of a reduced ability to induce <i>c-fos</i> expression in synergy with cAMP.
<i>dl1106</i>	Δ90-105	++++	+++	Appears WT.
<i>dl1107</i>	Δ111-123	+++	+++	Defective in transformation, mitosis and binding to pRb and p130.
<i>dl1108</i>	Δ124-127	+++	++	Defective in transformation, and mitosis. Induces DNA over-replication. Doesn't bind pRb, p107 or p130.
<i>dl1109</i>	Δ128-138	+	++++	Defective in mitosis. Induces DNA over-replication. Poor binding to pRb. Less stable than the WT.
<i>dl1110</i>	Δ140-160	N/A	N/A	Defective in CR2 transactivation. No 12S product. No induction of cytogenic damage.
<i>dl1112</i>	Δ161-168	N/A	N/A	Defective in CR2 transactivation. No induction of cytogenic damage.

<i>dl1113</i>	Δ169-177	N/A	N/A	Defective in CR2 transactivation. No 12S product. No induction of cytogenic damage.
<i>dl1114</i>	Δ178-184	N/A	N/A	Defective in CR2 transactivation. No 12S product. No induction of cytogenic damage.
<i>dl1115</i>	Δ188-204	N/A	N/A	Defective in CR2 transactivation. No 12S product. No induction of cytogenic damage.
<i>dl1116</i>	Δ205-221	+++	+	Impaired induction of cytogenic damage.
<i>dl1119</i>	Δ1-139	+	+	Deletion of the entirety of Exon 1.
<i>dl1132</i>	Δ224-238	+++	++	Deficient in induction of S phase related genes.
<i>dl1133</i>	Δ239-254	+++	+++	Poor immortalization. Enhanced tumorigenicity and metastasis.
<i>dl1134</i>	Δ255-270	++	+	Poor immortalization. Enhanced tumorigenicity and metastasis.
<i>dl1135</i>	Δ271-284	+++	++	No binding to CtBP and poor immortalization. Enhanced tumorigenicity and metastasis. Not recognized by M73.
<i>dl1136</i>	Δ285-289	+	++	Deletion of nuclear localization signal KRPRP. Poor immortalization, enhanced tumorigenicity and metastasis. Not recognized by M73.
<i>dl1141</i>	Δ61-69	N/A	N/A	No repression of SV40 enhancer. Blocks myogenic differentiation and induces differentiation in P19 cells. Induces <i>c-fos</i> expression in synergy with cAMP, No binding to p300
<i>dl1142</i>	Δ82-92	N/A	N/A	Reduced ability to bind to Sug1.
<i>dl975</i>	Δ140-186	N/A	N/A	Does not transform with <i>ras</i> .
<i>dl520</i>	Δ139-221	N/A	N/A	No 13S mRNA.

1.4 Transcriptional Regulatory Networks (TRNs)

It is a well-known fact that cells must continuously adapt to changing micro environmental conditions and signaling, and that they do so by modulating their gene expression patterns. Gene expression is directly controlled by a precise interplay of transcription factors (TFs), with each TF and its target set of genes called a Transcriptional regulatory network (TRN). TRNs are computationally represented as bipartite network containing nodes connected by

edges. Here, nodes represent proteins, or genes. Edges represent the direct regulatory interactions that a node will have with another – such as a TF binding to specific DNA motifs within the promoter region of its target gene and modulating transcription^{118,119}. The interplay within these networks is responsible for the cell type or cell state specific changes. However, they are robust enough to allow for dynamic changes in the case of important signaling cues. They can help to maintain homeostasis and can both drive the response to infection, and promote disease through their dysregulation¹²⁰.

One of the main characteristics of TRNs is that they are sparse in nature. The number of nodes which have a high number of edges, also known as “molecular hubs” are much fewer than those which have few connected edges. This is a phenomenon which is widely observed in both ecological and biological networks. This is thought to reflect the important function of hubs in regulating specific transcriptional programs directed towards specific stimuli. This ensures biological networks are less sensitive, and will remain stable in response to perturbations. It has been found that the percentage of active interactions (which can also be thought of as “connectivity”) is inversely proportional to the size of the studied system. In other words, as the network size increases, the percentage of active interactions decreases, to maintain sparsity. Another advantage gained in sparser systems is that the system can maximize its ability to adapt to newly intervening changes, also known as its “explorability”¹²¹.

Overall, TRNs adopt a “scale-free” structure, meaning they include the presence of large regulatory hubs. On a basic level, these networks can be thought of as containing 4 levels of detailed structures¹²². First, there is the interactions between TFs and their target genes, through the intermediary of the binding sites within their DNA promoter regions.

Second, these interactions can be organized in the form of repeated similar interactions and interconnections. These are called networks motifs, and they will appear repeatedly within a set network motif. They are not independent and cannot be functionally separated from the rest of the network. However, they have been shown to have specific kinetic properties which determine the temporal expression program of its regulated genes¹²³. Third, these motifs can be found to cluster semi-independently within transcriptional units which are known as network modules. In yeast, it has been reported that the overall regulatory system is fragmented into these independent module hubs, each responsible for its specific function¹²⁴. However, with the collection of more data it has been found that these modules are generally still highly interconnected within networks, and can rarely be fully separated from the network and retain their functions. These can be identified using gene expression data in which clustered gene groups have been identified as belonging to the same modules¹²⁵. Finally, it is the interconnection of these modules which builds up the overall TRN. The incoming connectivity is defined as the number of TFs regulating a target gene and these quantify the combinatorial effect of gene regulation. A recent study has shown that the fraction of target genes with a certain number of connections decreases exponentially, indicating that most targets interact with similar low numbers of TFs¹²⁴. On the other hand, the outgoing connectivity is defined as the number of target genes regulated by each TF. In this case, there are a few disproportionately highly connected TFs, responsible for global hub regulation, while most TFs have fewer connections and are mainly responsible for fine tuning the global mechanisms in play. When considering entire TRNs, studies have shown that the response to different kinds of stress will affect the complexity of the TRN's change. For instance, in the case of a cell responding rapidly to

external changes (such as DNA damage, cellular stress, infection), the response will be simple and composed of few cascades of responding TFs. However, in the case of multistage processes (include cell cycle, or sporulation), sequential complex cascading interactions are necessary¹²⁶. Most of the work which has so far been done on TRNs has been through the bacterium *Escherichia coli* and the yeast *Saccharomyces cerevisiae*, as sequencing data has been the most abundant. However, principles have been found to apply to many other living systems¹²⁷⁻¹²⁹.

There are several computational methods by which these networks can be reconstructed, allowing for a better understanding of the molecular mechanisms underpinning specific cellular phenotypes.

1.4.1 Computational methods for reconstruction of TRNs

Being that generation of molecular data such as high throughput sequencing data is no longer a limitation in the field of genomics, it is the modelling and representation of these biological networks which becomes the next challenge to overcome. TRNs are classically represented by a graph in which the genes are depicted by nodes and their regulatory interactions are the edges connecting these nodes. One of the most commonly used data sources for reconstruction, likely due its earlier availability, is bulk transcriptomic data. There are several methods by which this can be done. First, correlation methods and mutual information can be used to predict regulatory relationships. With these, it becomes possible to correlate the variation in a TF's expression with expression of other genes. Mutual information allows for the consideration of non-linear dependencies between TFs and their targets¹³⁰. If one wants to predict directed interactions, however, correlation falls flat. Regression methods can do this, with the caveat that it is assumed that

a TF and their direct target genes will vary linearly. This fails to take into consideration that TFs are transcriptionally and post-transcriptionally highly regulated, and their binding to DNA motifs is highly dependent on availability of the chromatin as well as interaction with co-factors¹³¹. Typically, a method of feature selection will be used to select the TFs to be used in the regularized regression models¹³². Bayesian networks can also be used to represent interactions between TFs and target genes. These probabilistic graphical models represent a set of random variables and their dependencies. In the context of TRNs, edges will stand for the conditional dependencies between genes. There are two main steps. First, the structure of the network is learnt, allowing for potential model improvements. Second, TF activities are predicted from the network model. A likelihood score assigns each target gene to a predictor TF. A major advantage of this model stems from its ability to learn from previous observations¹³³. Methods using ordinary differential equations (ODEs) allow the consideration of quantitative and dynamic interactions among TFs and their targets, meaning this approach is well suited to data over multiple time points¹³⁴. Given that biological regulatory interactions are rarely linear, it has been of interest to consider qualitative representation of regulatory networks. This type of logic modelling is based on the idea that a variable can only take a discrete number of states or values. The state of a variable is posited to be the logical combination of the states of other variables¹³⁵.

Although using bulk transcriptomic data allows for reconstruction of regulatory networks, this type of approach requires many samples to achieve sufficient robustness, and may lack TF regulatory information. Therefore, in several approaches, information about TF binding and chromatin availability is included. This can be in the form of Chromatin Immuno-precipitation (ChIP) – Sequencing or even Assay for transposase-

accessible chromatin (ATAC) – Sequencing. Having the ability to integrate information about TF-promoter physical interactions or chromatin availability is invaluable, as TF RNA levels are not particularly good predictors of their regulatory activities^{119,136}.

1.5 Hypothesis and Aims

In the context of disease or of physiological dysregulation, it is interesting to study the ways by which cells will adapt. This can be done by looking at their transcriptional regulatory networks (TRNs) and examining how these can change depending on the cellular environment. This can help us to learn more about the mechanisms behind disease processes, as well as allow us to therapeutically target specific regulatory steps. The E1A protein acts as a molecular hub protein, interacting with many structurally diverse targets, thus defining a specific pathway or phenotype⁹⁷. As E1A is unable to bind to DNA, it must interact with many different regulatory proteins and transcription factors, to instil these transcriptional changes benefitting viral replication. It has over 50 reported primary protein targets, which will bind to different parts of its sequence⁹⁷. Although studies have been done regarding the effects of these mutations on viral replication and infection, little is known about how these affect host cell transcriptional regulatory networks. Our objective is to reconstruct host cell TRNs, following infection with various adenovirus E1A deletion mutants. This will allow us to further characterize the effects of each of these deleted regions on the cellular transcriptional landscape. To execute this goal, we will focus on three specific aims, outlined below:

- i. We will explore the transcriptional landscapes of host cells infected with adenovirus E1A mutants
- ii. We will reconstruct the TRNs induced upon infection with different E1A mutants

- iii. We will examine the enriched promoter binding regions in our mutants

2 - Methods

2.1 Cell lines + virus used

Cultured cells used in all experiments are of the IMR-90 cell line (ATCC, Catalogue#CCL-186). They are fibroblasts isolated from normal lung tissue. They were shipped by the manufacturer on dry ice and were stored in liquid nitrogen until needed. All virus constructs were obtained from the Mymryk lab. The WT virus used as a control was the *dl309* construct, which has an intact E1A region, but which has deletions in the E3 gene, namely the 10.4K, 14.5K and 14.7K genes are deleted, while the 6.7K gene has a two-AA deletion¹³⁷. As compared to the WT virus, *dl309* has enhanced cytopathogenicity in several cell lines, while also leading to reduced late gene expression¹³⁸. Additionally, this also leads to increased apoptosis throughout the cell cycle¹³⁸. This virus is the backbone for all other E1A deletion mutants used in this study⁹⁶. Among these are *dl312* (Δ E1A), in which the entirety of the E1A gene is deleted, ensuring there is no production of a legible protein¹³⁷. The *dl520* (Δ E1A289R) mutant expresses the E1A 243R 12S isoform, but does not express the large 13S (289 AA) transcript of the E1A protein, and is mainly replication deficient, for the most part¹³⁹. The *dl975* (E1A12S Δ 186-222) construct expresses only a mutated 12S (143 AA) transcript of the E1A protein, which lacks the first 36 AAs of the second exon, but is still capable of immortalization¹⁴⁰. The *dl1119* (Δ 1-139) is a deletion of the entire first exon of the 13S transcript of the E1A protein⁹⁶. The remaining mutants, from *dl1101* (Δ 4-25), *dl1102* (Δ 26-35), *dl1103* (Δ 30-49), *dl1104* (Δ 48-60), *dl1105* (Δ 70-81), *dl1106* (Δ 90-105), *dl1107* (Δ 111-123), *dl1108* (Δ 124-127), *dl1109* (Δ 128-138), *dl1110* (Δ 140-160), *dl1112* (Δ 161-168), *dl1113* (Δ 169-177), *dl1114* (Δ 178-184), *dl1115* (Δ 188-204), *dl1116* (Δ 205-221), *dl1132* (Δ 224-238), *dl1133* (Δ 239-254), *dl1134* (Δ 255-

270), *dl1135* (Δ 271-284), *dl1136* (Δ 285-289), *dl1141* (Δ 61-69) and *dl1142* (Δ 82-92) contain small deletions within the 13S transcript of the E1A protein^{56,116}. They will, collectively, delete almost the entire protein⁹⁶.

2.2 Infection

IMR90 cells were maintained in DMEM growth media (Wisent, Catalogue# 319-016-CL) supplemented with 10% FBS (VWR, Catalogue# 10221-86) and 1% P/S antibiotics (Wisent, Catalogue# 450-201-EL). Cells were seeded in 15cm dishes (two dishes for each the uninfected control, the WT virus, and all mutants). Plates were incubated at 37°C with 5% CO₂ until they reached 100% confluency. They were incubated for 48 hours in the same incubator to ensure cells were growth arrested. The media was removed and DMEM + 1% P/S was added prior to infection. Two plates were infected with each HAdV mutant species, and two plates were treated with virus buffer (Tris-HCl pH 7.4 at 150mM, NaCl at 680mM, KCl at 25mM, Na₂HPO₄ at 3.5mM and D-Glucose at 28mM) as a control, at an MOI of 5. Cells were collected at the 72-hour time point p.i. Two replicates were included for each of the experimental conditions.

2.3 RNA-Sequencing Library Preparation

RNA was extracted from cells following instructions from the GENEzol TriRNA Pure Kit from Geneaid (SKU#GZ050). Briefly, 1450 μ l of GENEzol reagent was added directly to culture dishes to lyse cells. The samples were incubated for 5 minutes at room temperature before being transferred to 1.5mL microcentrifuge tubes. Samples were centrifuged to remove cell debris. Absolute ethanol was added to the mixture at a ratio of 1:1. This mixture was transferred to a column for RNA binding. To reduce DNA contamination, a DNase I solution was pipetted onto the column and incubated for 15 minutes. Repeated centrifugation with wash buffer were performed to ensure purity of the

sample. Lastly, columns were incubated for 3 minutes with RNase-free water (Ambion, Catalogue# AM9930) for RNA elution. Centrifugation allowed the capture of a small volume of pure RNA. The concentration of RNA in solution was measured using the NanoDrop machine. For the remaining protocol, 0.1-1 μg of RNA should be used as starting material.

Library preparation was performing using a dUTP magnetic beads preparation protocol. The first step involves the binding of RNA to magnetic beads. First, 20 μL oligo(dT) magnetic beads (NEB, Catalogue# S1419S) per sample were added to a 1.5 mL microcentrifuge tube. Beads were washed twice with an equal volume of 2xDTBB (20 mM Tris-HCl (pH 7.5), 1 M LiCl, 2 mM EDTA, 1% lithium dodecyl sulfate, 0.1% Triton X-100) using magnet and vacuum aspirator. Beads were resuspended in 100 μL of 2xDTBB per sample. To prepare the RNA, 0.1-1 μg of RNA is prepared in 50 μL of ddH₂O in 200 μL PCR tubes. For each sample, 50 μL of the prepared magnetic beads were aliquoted into the PCR tubes. The remaining beads were left on ice. Samples were vortexed briefly to allow RNA binding to the beads. Tubes were centrifuged briefly to ensure all liquid remain at the tube bottom. Tubes were incubated for 2 minutes at 65°C in a thermocycler with heated lid. Tubes were incubated for 10 minutes at RT on a rotator. Beads were collected on a magnet for 5 minutes, washed twice with 200 μL of RNA wash buffer (10mM Tris-HCl (pH 7.5), 0.15 M NaCl, 1 mM EDTA). RNA was eluted through resuspension in 50 μL of RNA elution buffer (10mM Tris-HCl pH 7.5, 1mM EDTA). Samples were vortexed and centrifuged briefly before incubation for 2 minutes in a thermocycler at 80°C. After incubation, samples were placed on ice immediately. Beads were isolated using a magnet, and the RNA eluate (liquid) was transferred to a new PCR tube. Binding to beads was

repeated as before. The remaining prepared beads were aliquoted to the tubes (50 μ L per tube). Samples were vortexed and centrifuged before incubation for 2 minutes at 65°C in the thermocycler. Samples were incubated for 10 minutes at RT on a rotator. After collection of beads using a magnet, samples were washed twice with 200 μ L of RNA wash buffer. The second step of library preparation involves fragmentation. First, fragmentation buffers A (6 μ L 5X FS buffer (Life Technology, Catalogue# 18080-044) + 24 μ L ddH₂O for each sample) and B (4 μ L 5X FS buffer + 5 μ L ddH₂O + 1 μ L DTT 100mM (Life Technology, Catalogue# 18080-044) for each sample) should be prepared. After the removal of the last 200 μ L wash of 2x DTBB, beads were washed with 30 μ L of buffer A, and then resuspended in 10 μ L of buffer B. Samples were incubated in a thermocycler for 9 minutes at 94°C, and then placed on ice. Using a magnet, the supernatant (~10 μ L) was transferred to a new PCR tube.

The third step is the first strand cDNA synthesis. Before starting, both the RT mix A (0.5 μ L of Random Primers (Life Technology, Catalogue# 48190-011) + 0.5 μ L of oligo dT primers at 50 μ M (Ambion, Catalogue# AM5730G) + 0.5 μ L of SUPERase-In (Ambion, Catalogue# AM2694) + 1 μ L of dNTPs at 10 mM (Life Technology, Catalogue# 18427088) for each sample) and RT mix B (5.8 μ L of ddH₂O + 0.1 μ L Actinomycin D at 2 μ g/ μ L (Thermo-Fisher Scientific, Catalogue# A7592) + 1 μ L DTT at 100 mM + 0.2 μ L 1% Tween 20 (BioShop, Catalogue# TWN510.500) + 0.5 μ L M-Mulv Reverse Transcriptase (Invitrogen, Catalogue# 28025013) for each sample) should be prepared. For each sample, 2.5 μ L of the RT mix A should be added. Samples can be incubated for 1 minute at 50°C in a thermocycler, before being placed on ice. Next, 7.5 μ L of RT mix B is added to each sample. They can be incubated at 25°C for 10 minutes, and then 50°C for 50

minutes, before being placed back on ice. For each sample, 36 μL of RNAClean XP beads (Beckman Coulter, Catalogue# A63987) were added. Samples were vortexed and centrifuged to encourage binding to beads. Samples were incubated for 15 minutes on ice before collection on magnet for 5 minutes. Samples were washed twice with 75% ethanol. With PCR tube caps removed, samples were left on magnet to air dry for 10 minutes. Samples were eluted using 11 μL of ddH₂O. The magnet was applied to allow the transfer of the 10 μL eluate to a new PCR tube.

The fourth step led to the second strand cDNA synthesis with dUTP. Before starting, the second strand reaction mix (1.5 μL 10X Blue Buffer (Enzymatics, Catalogue# P7050L) + 1 μL dNTP mix w/ dUTP at 10 mM (Affymetrix (Core), Catalogue# 77330) + 0.1 μL dUTP at 100 mM (Affymetrix (Core), Catalogue# 77206) + 0.15 μL 1% Tween 20 + 1.05 μL ddH₂O + 0.2 μL RNase H (Enzymatics, Catalogue# Y9220L + 1 μL DNA polymerase I (Enzymatics, Catalogue# P7050L) can be prepared. To each tube, we added 5 μL of the second-strand reaction mix. Tubes were incubated at 16°C for 2.5 hours in the thermocycler with the heated lid, and then left at 4°C in the same thermocycler overnight. The next day, Speedbeads should be prepared in the following way. In a microcentrifuge tube, a volume of 1.5 μL of Speedbeads solution (Thermo, Catalogue# 6515-2105-050250) per sample was added. Beads were isolated using a magnet, and a vacuum was used to remove the supernatant. A volume of 28 μL of 20% PEG8000/2.5M NaCl per sample was added to the beads, which were re-suspended. An aliquot of 28 μL of the speedbeads solution was added to each sample. Samples were vortexed and centrifuged prior to an incubation at RT for 10 minutes. Beads were collected on a magnet for 5 minutes. Beads were washed twice with 200 μL of 75% ethanol and air dried for 10 minutes. Samples were

eluted using 40 μL of TE buffer (Bio Basic, Catalogue# USD8211). The next step requires the end repair mixture (5 μL 10x T4 buffer (Enzymatics, Catalogue# B6030) + 1 μL dNTP mix at 10 mM + 0.5 μL 1% Tween 20 + 2.9 μL ddH₂O + 0.3 μL T4 DNA polymerase (Enzymatics, Catalogue# P7050L) + 0.3 μL T4 PNK (Enzymatics, Catalogue# Y9049L) + 0.06 μL Klenow (Enzymatics, Catalogue# P7060L) for each sample). To each PCR tube, 10 μL of the end repair mixture was added. This was vortexed and centrifuged, prior to 10 minutes of incubation at 20°C and then placed on ice. Another 93 μL of the PEG8000/2.5M NaCl solution was added to each sample. Samples were vortexed and centrifuged. They were incubated for 15 minutes at RT. Beads were collected along magnets for 5 minutes. They were washed twice with 200 μL of 80% ethanol and air dried for 10 minutes. Elution was done with 15 μL of TE buffer.

In the next step, barcoding the samples takes place. The tailing reaction mix (3 μL of 10X Blue buffer + 0.6 μL dATP mix at 10 mM (Life Technology, Catalogue# 10216-018) + 0.3 μL 1% Tween 20 + 10.8 μL ddH₂O + 0.3 μL Klenow 3'-5' Exo (Enzymatics, Catalogue# P7010-LC-L) must first be prepared. A volume of 15 μL of the tailing mix was added to each sample. Samples were incubated at 37°C for 30 minutes in a thermocycler and then placed on ice. Then, 55.8 μL of PEG8000/2.5M NaCl was added to each sample. Samples were vortexed, centrifuged and then incubated at RT for 15 minutes. Beads were collected on a magnet for 5 minutes and then washed twice with 200 μL of 80% ethanol. Samples were left out to air dry for 10 minutes before eluting with 14 μL of TE buffer. Different IDT barcode adapters, in a volume of 5 μL were added to each sample (IDT adapters should be diluted 1:50 for a final concentration of 5 μM). A volume of 15.83 μL of the adapter ligation reaction mix (15 μL 2x rapid ligation buffer (Enzymatics,

Catalogue# L603-LC-L) + 0.33 μ L 1% Tween 20 + 0.5 μ L T4 DNA ligase HC (Enzymatics, Catalogue# L6030-HC-L) for each sample) was added to each sample. Samples were incubated at RT for 15 minutes and then put on ice. A volume of 7 μ L of 20%PEG8000/2.5M NaCl was added to each sample. They were centrifuged, vortexed and then incubated at RT for 15 minutes. Beads were collected on magnets for 5 minutes and washed using 200 μ L of 80% ethanol. Beads were left out to air dry for 10 minutes. Then, 20 μ L of elution mix (20 μ L TE buffer + 1 μ L UDG (Enzymatics, Catalogue# G5010L) per sample) was added to each sample. Beads were collected on a magnet and the eluate was transferred to new PCR tubes. Samples were incubated for 30 minutes at 37°C in a thermocycler to digest the second strand.

Lastly, samples were amplified and library sizes were selected using gel electrophoresis. To do this, 10 μ L of each sample was transferred to a new PCR tube, along with 10 μ L of the final PCR reaction mix (4 μ L 5x Phusion HF buffer (Thermo, Catalogue# F530L) + 1 μ L dNTP mix at 10 mM + 4.35 μ L ddH₂O + 0.2 μ L 1GA primer at 100 μ M (IDT, Sequence: AATGATACGGCGACCACCGA) + 0.2 μ L 1GB primer at 100 μ M (IDT, Sequence: AATGATACGGCGACCACCGA) + 0.25 μ L Phusion polymerase (Thermo, Catalogue#F530L) per sample). The thermocycler was set to run the initial denaturation at 95°C for 3 minutes, amplify for 8-14 cycles (98°C for 20 seconds, 60°C for 30 seconds and 72°C for 20 seconds) and elongate at 72°C for 1 minute. A 10% TBE gel (4 mL 30% Acrylamide/Bis solution (29:1) (BioBasic, Catalogue#A0010) + 3.6 mL H₂O + 1.2 mL 10x TBE + 150 μ L 10% APS + 20 μ L TEMED (BioBasic, Catalogue#TB0508) was cast and loaded with 20 μ L of each sample, along with a 100 bp ladder. After running the gel for 30-45 minutes, 1 μ L of SYBRGold stain (Life Technology, Catalogue# S11494)

was used to stain the gel directly for 30 seconds. Visualization was done using a BluPAD (Bio-Helix, Catalogue# BP001CU). Products between 200-350 bp were excised using a razor blade and placed inside 0.8 mL perforated microcentrifuge tubes, which are themselves inside of 1.5 mL microcentrifuge tubes. Gel slices were shredded through centrifugation at 14,000xg through the perforations. The perforated tubes were discarded and 100 μ L of diffusion buffer (500 mM ammonium acetate at pH 8.0, 0.1% SAS, 1 mM EDTA, 10 mM magnesium acetate) was added to each tube. Samples were incubated for 45 minutes on the rotator. The cDNA libraries were purified using the Zymo DNA Clean and Concentrator kit (Zymo, Catalogue# D4004). Briefly, each tube received 500 μ L of the DNA binding buffer (Zymo, Catalogue# D4004) and was centrifuged at 16,000xg for 1 minute. The supernatant was transferred to the provided spin columns in a collection tube. Tubes were centrifuged for 30 seconds, and the flow-through discarded. Then, 200 μ L of the DNA wash buffer (Zymo, Catalogue# D4004) was added to the column, followed by 30 seconds of centrifugation. This step was repeated. Lastly, 25 μ L of H₂O was added directly to the column matrix, which was incubated for 1 minute at RT. Columns were transferred into clean microcentrifuge tubes for the elution. The DNA concentrations in each of the tubes was measured using the Qubit dsDNA HS Assay Kit (Invitrogen, Catalogue# Q32854).

Samples were sent to the IGM facility at UCSD for Illumina Sequencing on a Nextseq550 single end x75 to a depth of ~20 million reads per sample. Due to inadequate sequencing depth in one of two replicates, the same samples were sent once again. This led to adequate data, according to QC measures.

2.4 Pipeline for processing Sequencing data

Fastq files were obtained directly from the sequencing facility. Quality control metrics were evaluated through the FASTQC software, by Babraham Bioinformatics, a program designed to identify potential problems in high throughput sequencing datasets. This ensured the sequencing quality was sufficient and the read depth uniform before moving on to the next step. The STAR alignment tool was employed to align reads to the human genome, outputting BAM (Binary Alignment Map) files for each replicate. All lanes from each replicate were merged, using the merge function of the samtools package.

The next steps were carried out using the HOMER suite of computational tools. Developed at UCSD, these command line executables allow for the easy analysis of high throughput sequencing data. First, tag directories were created from the BAM files, allowing for the organization of mapped and indexed reads by chromosome and speeding up downstream processes. From there, reads were quantified using the HOMER analyzeRepeats package. Both raw and RPKM values (reads per kilobase mapped reads) can be generated from this program. Differential gene expression analysis (DGEA) was carried out using EdgeR. ATAC-Sequencing peaks were called using HOMER findPeaks, and both sample replicates were merged using Irreproducible Discovery Rate (IDR) analysis.

2.5 Building TRNs

To reconstruct a bipartite TRN, in which the TFs and target genes represent the nodes we carried out the following steps: 1 – select the candidate TFs using ATAC-Seq motif enrichment – 2 – select the target genes using results from DGEA – 3 – infer TF-target gene regulatory relationships using Elastic Net Regression.

First, candidate TFs were selected. TFs will be selected based on potential binding to enriched regions in the infected sample. All human TFs were ranked according to highest RPKM expression. Of that ranked list, only the top 20 TFs were retained for TRN reconstruction. Second, target genes for the network were chosen through a differential gene expression analysis. After running DGEA for each of the virus treatments, comparing them to the WT virus, all identified differentially expressed genes were pooled across all conditions. A gene was considered differentially expressed if the $|\log_2 \text{fold change}| > 1.5$ and the false discovery rate (FDR) < 0.0001 . All repeated genes were eliminated to give a smaller pool of differentially expressed genes across all conditions. Lastly, TFs selected in the first step were eliminated from this pool to remove the possibility of self-edges in the final product. Lastly, relationships between TFs and target genes were inferred to reconstruct biologically relevant TRNs. It is important to consider that biological networks are sparse, and that central regulatory hubs will control most observed regulatory effects. For these reasons, Elastic Net is the best choice of regression model for this reconstruction. It is a modified form of linear regression in which two penalties are imposed on the mean squared error loss function. First, the Lasso penalty imposes sparsity, mimicking biological standard. Second, the Ridge penalty shrinks coefficients towards zero, approaching a topology which is closer to biology. The loss function can be expressed as shown below:

$$\frac{1}{2n} \|y - Xw\|_2^2 + \alpha \rho \|w\|_1 + \frac{1}{2} \alpha (1 - \rho) \|w\|_2^2$$

In this expression, X represents the matrix of TF expression, y the vector of target gene expression and w the vector of weights (coefficients) to optimize. The first term in this function is the mean squared loss, while the second and third terms represent the Lasso

and Ridge penalties, respectively. The α parameter controls the penalization while the ρ parameter is the mixing parameter between the two penalty terms. We use the ElasticNetCV function from the scikit-learn python package to optimize α for each target gene by cross-validation. The parameter ρ was set to 0.5. Prior to regression, all gene expression data was log-transformed and scaled to a standard Normal distribution.

A global TRN will be constructed using expression data from all experimental conditions, due to the small number of replicates. Then, TRNs will be reconstructed in which one experimental condition is removed, one at a time. The subtraction of the “leave-one out” TRN from the global TRN allows for the determination of one condition’s contribution to the overall phenotype. This allowed for comparison. We then used graph edit distance to quantify the difference between experimental conditions.

3 – Results

3.1 E1A deletion mutants have different RNA expression profiles

Following pre-processing of RNA-Sequencing data, gene expression of mutants was quantified using the EdgeR package in R. Differentially expressed genes were identified based on a log2 fold change of $|\geq 1|$ and an FDR-corrected p-value less than or equal to 0.01 when compared with the E1A WT virus (*dl309*). Log2 fold change values are illustrated in Figure 3A, for all genes for which expression was detected. This illustrates the apparent heterogeneity present in this data, in that gene expression changes (as compared to the WT virus) vary greatly depending on the mutant. It is also possible to represent the spread of differentially regulated genes for the totality of mutants as a boxplot, in which each gene is represented by a single data point. In Figure 3B, the spread of gene regulation for most mutants appears to be centered, with approximately equal numbers of up or downregulated genes, as compared to the WT. Surprisingly, for both *dl1101* ($\Delta 4-25$), and *dl1132* ($\Delta 224-238$), the distribution is skewed to the left, with a higher number of genes seemingly downregulated rather than upregulated. It should be noted that in these mutants there appears to be several outliers which are upregulated, indicating the left skew may not be representative of the data but driven by outliers.

To examine the differences between mutant expression profiles in a more functional manner, a KEGG pathway enrichment analysis was performed. Based on the differential gene expression, KEGG functional pathways were deemed either up or downregulated, as demonstrated in Figure 3C. In this heatmap, the 25 most highly enriched pathways related to downregulated genes were identified in red, while pathways related to upregulated genes are identified in blue. Once again, this enrichment varies depending on the location of the

deletion. Of note are the deletion mutants *dl1101* (Δ 4-25), *dl1136* (Δ 285-289), *dl1119* (Δ 1-139) and *dl1132* (Δ 224-238), in which the highly-enriched pathways in most mutants (red), do not appear differentially regulated. However, several of the downregulated pathways do appear to be highly enriched in these mutants. Additionally, mutations in *dl1113* (Δ 169-177), *dl1114* (Δ 178-184), *dl1115* (Δ 188-204), *dl1116* (Δ 205-221), appear to lead to a higher number of upregulated pathways and very few downregulated pathways. In the other deletion mutants, namely *dl1102* (Δ 26-35), *dl1103* (Δ 30-49), *dl1104* (Δ 48-60), *dl1105* (Δ 70-81), *dl1106* (Δ 90-105), *dl1107* (Δ 111-123), *dl1108* (Δ 124-127), *dl1109* (Δ 128-138), *dl1110* (Δ 140-160), *dl1112* (Δ 161-168), *dl1133* (Δ 239-254), *dl1134* (Δ 255-270), *dl1135* (Δ 271-284), *dl1141* (Δ 61-69) and *dl1142* (Δ 82-92), there appear to be approximately equal numbers of both up and downregulated functional KEGG pathways. A similar analysis can be done using Gene Ontology (GO) terms, in which over and underrepresented GO terms can be identified for each of the mutants, as compared to wildtype virus (*dl309*). Interestingly, when the top 50 most highly represented and top 50 most highly underrepresented GO terms are shown in a heatmap certain mutants appear to dominate the analysis (Figure 3D). In the overrepresented terms, mutants *dl1113* (Δ 169-177), *dl1114* (Δ 178-184), *dl1115* (Δ 188-204), *dl1116* (Δ 205-221) show a much lower FDR adjusted p-value than many of the other mutants, most of which appear not to show a significant difference with the WT virus. Overrepresented GO terms related to upregulated genes include “cell cycle”, “checkpoint signaling”, “nuclear fission” and are mostly centered on cell cycle regulation and cell division. Mutants which showed highly enriched GO terms related to downregulated genes included *dl1101* (Δ 4-25) and *dl1132* (Δ 224-238). GO terms which were highly represented here, with very low FDR adjusted p-values, included “cell

cycle”, “mitotic cell cycle”, “cell division” and several other terms related to cell proliferation and cycling. Although the terms included in this analysis were chosen based on a stringent p-value cutoff for both the underrepresented and overrepresented terms separately, both appear to focus on similar functional pathways. It appears that mutations in different regions of the E1A protein may have opposing effects on cell cycle (Figure 3D).

As the cell cycle-related genes and pathways appear to be highly involved in the transcriptional changes brought about by E1A, it was of interest to take a closer look at the top 50 most highly up/downregulated genes which are involved in the cell cycle regulation KEGG pathway. Their differential expression, as compared to the WT virus is expressed in log fold change in Figure 3E. It is immediately noticeable that, in mutants *d1113-1116*, most of the cell cycle related genes are overexpressed, as compared to the WT. On the other hand, in any of the other mutants, this overarching cell cycle regulation is not evident. In fact, in mutants *d1101-1112*, most of the genes appear to be similar to WT, or highly downregulated, suggesting different roles for different portions of E1A within the confines of cell cycle regulation. In mutations found within the latter portion of the second exon, including CR4, mutants appear not be confined to either a majorly up or down regulated phenotype, with a mixed phenotype appearing (Figure 3E). We can compare the heatmap derived from cell cycle related genes to that of another pathway, not expected to be highly involved in E1A mediated changes. In Figure 3F, genes related to the ribosome KEGG pathway are represented. While the ribosome biogenesis pathway is modulated through E1A’s interaction with TRRAP, the ribosome pathway itself, which modulates ribosomal function and protein production is not reported to be affect by E1A⁷³. Log fold change

values immediately appear to be much less striking than seen in the cell cycle heatmap. There is no observable trend among any of the mutants, as is expected of a pathway which will not be affected by mutations in E1A. This comparison brings to light the potential importance of the cell cycle in regulation of E1A-mediated transcriptional regulation.

Having explored host gene expression, it was important to look to adenovirus gene expression, as this can inform infection dynamics. Interestingly, the only genes which could be detected at the 72-hour time point were virus associated genes I and II (VAI and VAII). The expression of these viral genes varies lightly depending on the mutant, with average count values represented in Figure 3G. In the uninfected sample, the E1A K/O, *dl520* (Δ E1A289R) and *dl1119* (Δ 1-139), there is negligible to no viral gene production that can be detected. This is in line with expectations in that the first exon of the E1A protein is especially important in regulation of viral gene expression. Additionally, it can be noted that expression levels of VAI and VAII in *dl1105*, *dl1109* and *dl1114* are almost as high, if not higher than the WT virus, which could indicate a gain of function in these mutations or alternatively, innate repressive functions on viral gene expression as a means of expression control (Figure 3G).

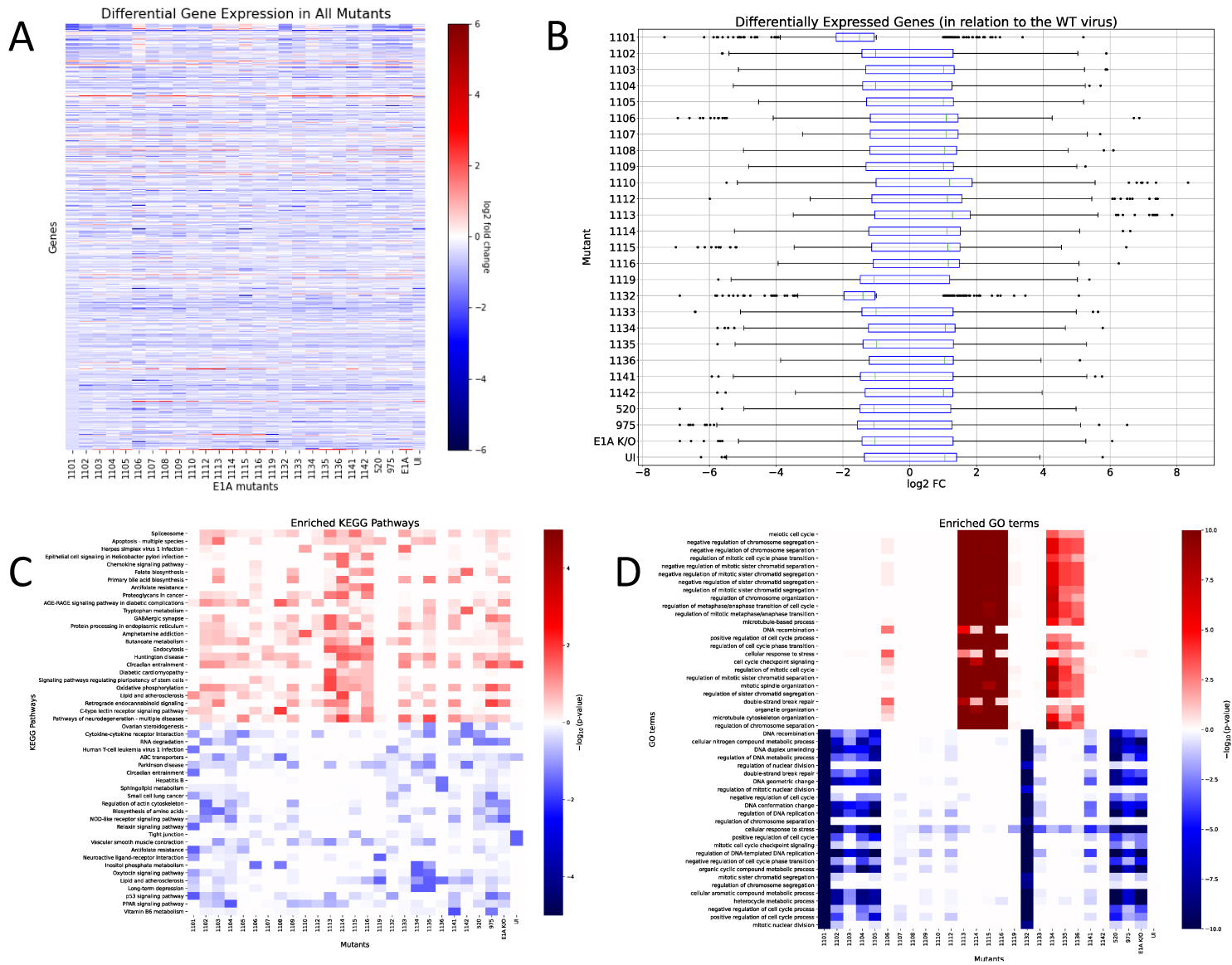


Figure 3: Expression Profiles of E1A Mutants. (A) Heatmap of differential Gene Expression in all mutants (B) Dispersion of differentially expressed genes expressed as a boxplot, with all mutants on the y axis (C) Enriched KEGG pathways for both upregulated genes (red) and downregulated genes (blue) (D) GO enrichment analysis for both upregulated (red) genes and downregulated (blue) genes

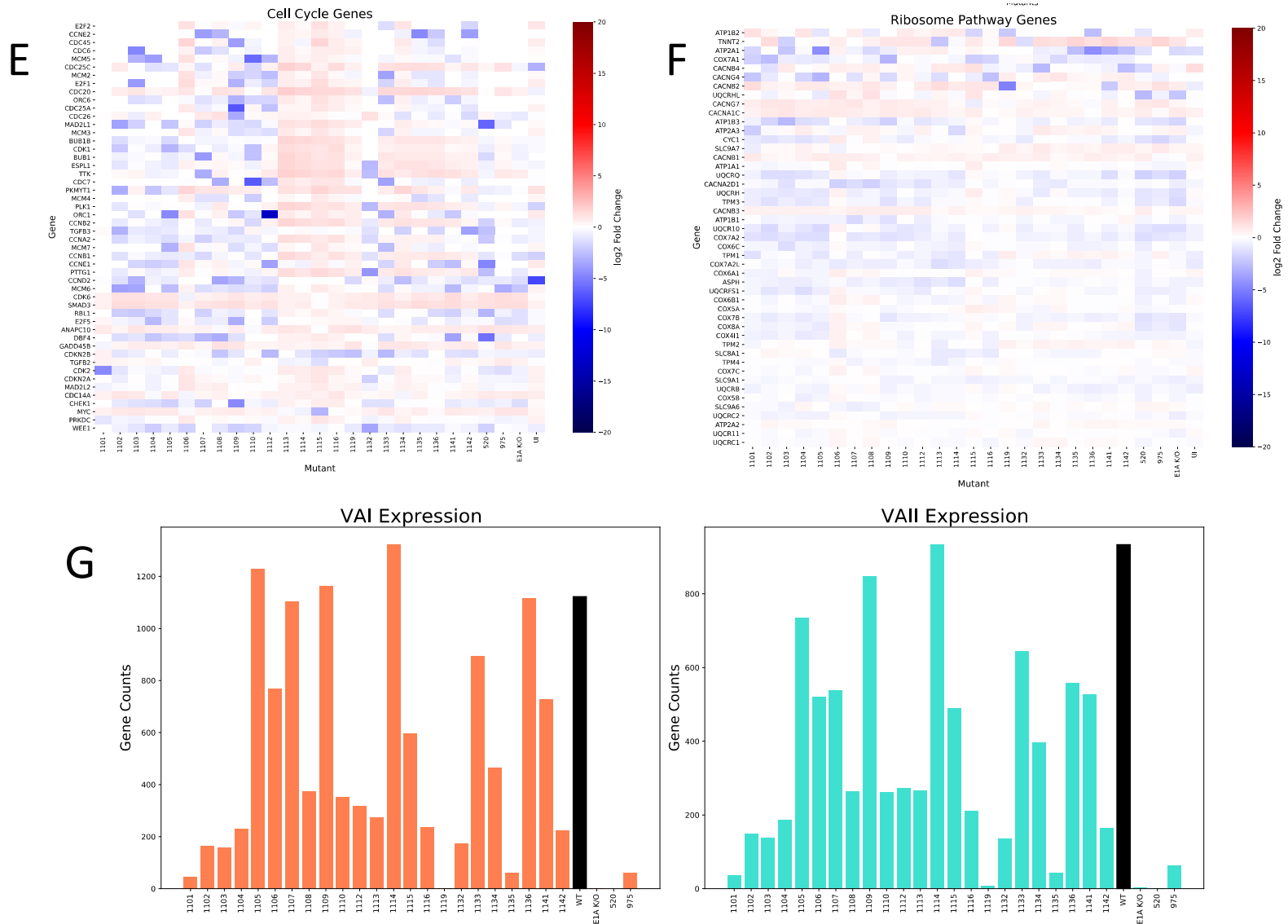


Figure 3: Expression Profiles of E1A Mutants (cont) (E) Heatmap of the top 50 most highly expressed genes related to the cell cycle KEGG pathway. (F) Heatmap of the top 50 most highly expressed genes related to the ribosome KEGG pathway. (G) Expression of viral genes VAI and VAII in all mutants. The WT viral expression is highlighted in black.

3.2 Hierarchical clustering reveals associations between mutants in similar regions of the E1A gene

A principal component analysis (PCA) was used as a visualization tool to identify any clustering trends among all mutants. Samples were color-coded in 4 different ways. First, coloring was based on whether mutants are found within a conserved region (CR) or non-conserved region (NCR) of the E1A genome (Figure 4A). Second, colors were based on whether mutants are found in exon 1 (X1), exon 2 (X2) or the unique region separating them (U) (Figure 4B). Third, the colors represented the specific conserved or non-conserved region in which mutants are found (Figure 4C). Lastly, colors were based on whether mutants are non-conserved, or within one of 4 identified conserved regions (Figure 4D). In all 4 cases, there is no apparent trend among the clustered data points. It doesn't appear as though the location of the mutation affects the way these mutants cluster. Considering this, we used a second method of clustering which relies on a more global view of gene expression differences, hierarchical clustering. When samples are clustered using a hierarchical clustering method, which assumes that all samples come from a common ancestor, in this case the common WT background virus, samples do appear to cluster in a region dependent manner. In Figure 4E, there are two small clusters of mutants found within the first exon of the protein (AA 1-139). First, *dl1119* (Δ 1-139), a deletion of the entire first exon clusters along with *dl1101* (Δ 4-25), *dl1102* (Δ 26-35), *dl1103* (Δ 30-49), *dl1104* (Δ 48-60), and *dl1108* (Δ 124-127). Similarly, mutants *dl1106* (Δ 90-105), *dl1107* (Δ 111-123), *dl1141* (Δ 61-69) and *dl1142* (Δ 82-92) cluster more closely together, with *dl1105* (Δ 70-81) seemingly not associated with any strong clusters. Interestingly, the mutants found in the unique region (AA 139-186) all appear to cluster closely together. This includes *dl1113* (Δ 169-177), *dl1110* (Δ 140-160), *dl1112* (Δ 161-168), *dl1114* (Δ 178-

184), a cluster which also seems associated with *dll109* (Δ 128-138), *dll115* (Δ 188-204) and the previously mentioned *dll105* (Δ 70-81). This hierarchical clustering analysis appears to indicate the existence of region dependent functions for the E1A protein.

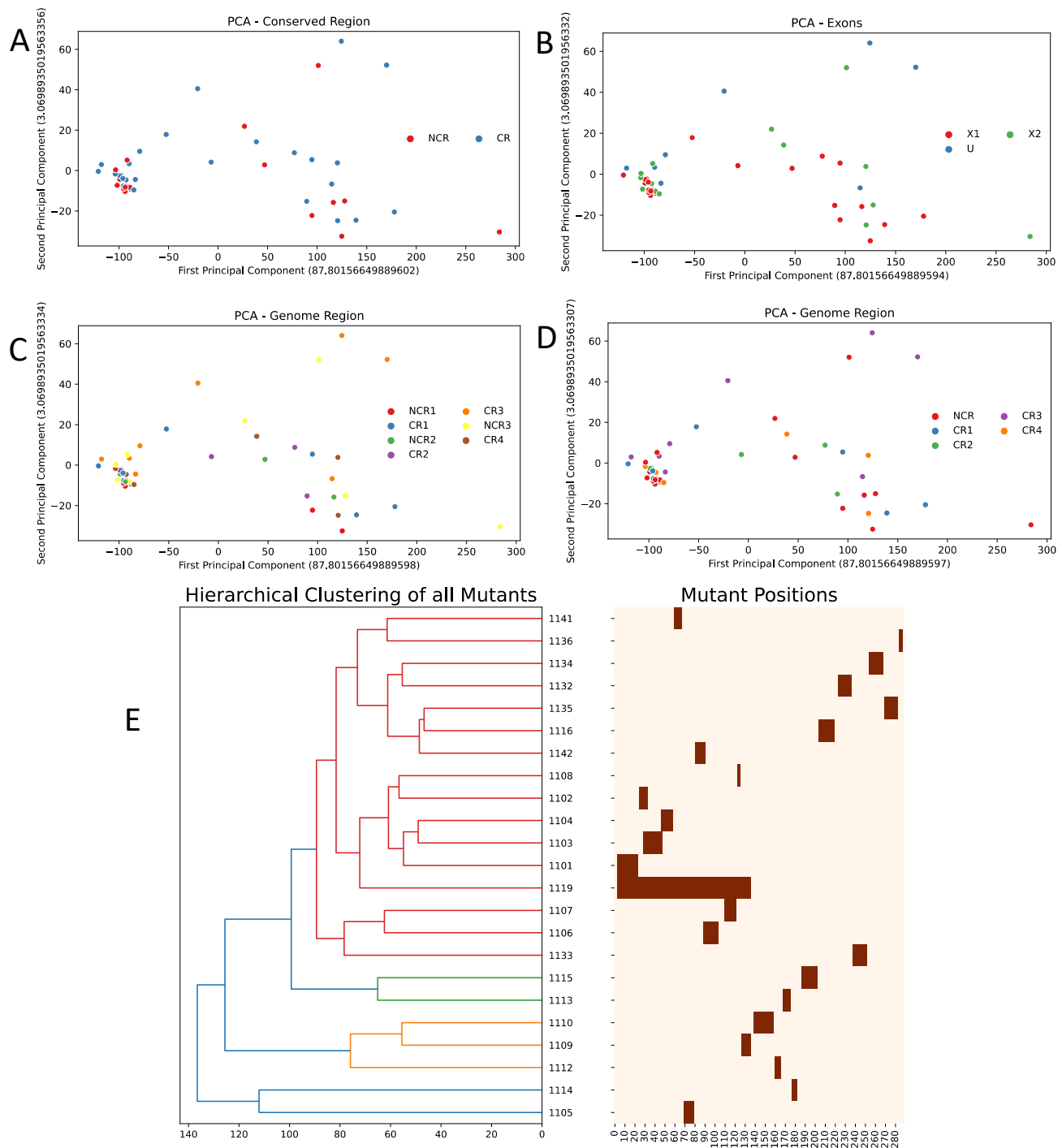


Figure 4: Clustering E1A deletion mutants. (A-D) Mutants were clustered using a Principle Component Analysis (A) Mutants were identified based on their belonging to either a conserved or non-conserved region of the E1A genome. (B) Mutants were color-coded based on their location within either exon 1, exon 2 or the unique region between both. (C) Mutants were organized based on their presence within any of the four conserved regions, or any of the 3 non-conserved regions. (D) Mutants were classified based on their location within any of the previously mentioned four conserved regions, or a non-conserved region. (E) A hierarchical clustering algorithm was used to cluster all mutants, as shown on the left of the diagram. The right side of the diagram shows the mutant positions within the entirety of the E1A gene (width).

3.3 Constructing TRNs

The process used to approximate TRNs for each mutant is demonstrated in Figure 5A. The TRN is composed of two aspects, TFs and differentially expressed genes (DEGs). These represent the nodes of the TRN. Edges are then built between each TF and DEGs to represent the effect the TF has on each genes expression. Once RNA-Sequencing data has gone through pre-processing, all human TFs expressed were ranked by mean RPKM, across all mutants. The top 20 TFs (according to rank) were chosen. They will become the TF nodes responsible for half the bipartite TRNs. Then, DGEA was performed for all mutants, as compared to the WT virus. The union of all differentially expressed genes, across all mutants was taken. All chosen TFs were removed from this list, to remove the possibility of self-regulation. Through increasingly stringent filtering of DEGs using FDR p-value ($< 10^{-8}$) and log fold change (> 1.5), 282 target genes were selected. These, along with the TFs selected, appear to be the most important in regulating the phenotypic changes seen in these mutants. An elastic net regression model was used to estimate TF-target gene relationships. This TRN will contain all the information from all mutants. In the rest of this work, this will be referred to as the “Original TRN”. This reconstruction process can be repeated sequentially, each time removing the expression information regarding one single mutant. This “leave one out” TRN allows for the estimation of TF-target relationships without the influence of that one removed mutant. Once the “leave one out” TRN is subtracted from the original TRN, the resultant interaction network can be said to quantify the impact of each individual mutant on TF-target relationships.

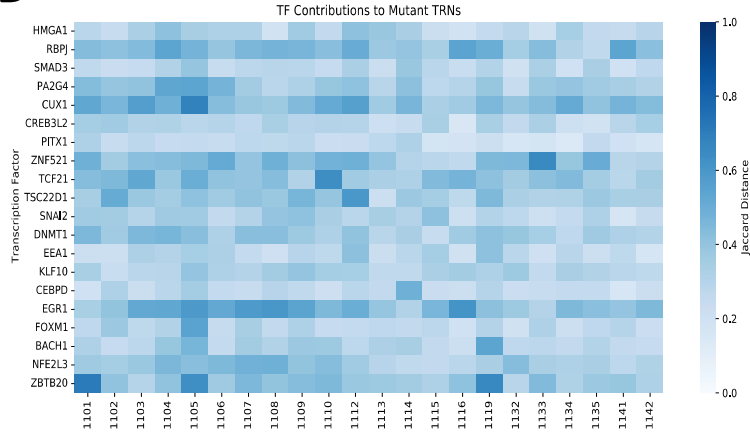
There are two main ways by which TRNs can be represented visually. First, as shown in Figure 5B, there is the classical bipartite representation, with all TFs on one side and all target genes on the other. However, this can become difficult to interpret. Figure

5C represents the classical network representation in which all genes and TFs are spread across the graphical space. This visual representation is preferred to the bipartite as it allows for a more intuitive visualization of the represented relationships. Additionally, it allows us to highlight specific highly connected TFs. In the rest of this work, all TRNs will be represented in this way. Blue edges (lines) represent negative regulatory effects, while red edges represent positive regulatory interactions. The width of each edge is proportional to the strength of the regulatory interaction it represents. The red points (nodes) represent TFs, with their size being proportional to the number of regulatory interactions they have with target genes. Grey nodes represent target genes.

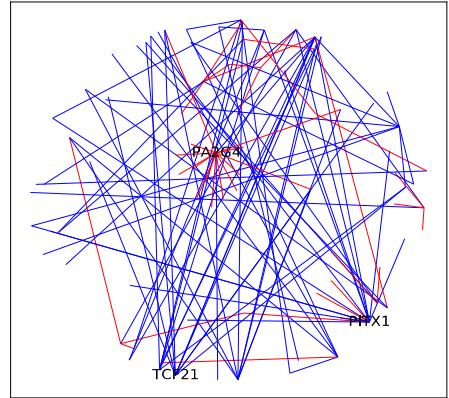
The original TRN mentioned previously is pictured in Figure 5D, which shows a visual representation of the regulatory effects of TFs on their target genes. In this TRN the 3 most highly connected TFs are labelled on the graph (PA2G4, TCF21, PITX1). Additionally, it is evident that in this TRN, most regulatory interactions are negative, with very few positive interactions. This is interesting given that this TRN represents expression information from all mutants. This evidence is indicative of the importance of the entirety of the E1A protein in regulation of gene function. In the interest of directly comparing each of the mutants, the Jaccard distance was calculated for each individual TF-mutant intersection (Figure 5E). To do so, the number of number of features found in either the original TRN or that of a specific mutant was divided by the union of common features found in both. This allows for the determination of the individual TF contributions to each specific TRN. A higher distance (darker blue) indicates a TF is contributing very differently to a specific TRN, as opposed to the original TRN. On the other hand, a lower Jaccard distance indicates similarity between mutant specific and overall TRNs. For

instance, in the case of the PITX1 TF, the Jaccard distance is small over the entirety of the heatmap, indicating its contribution is incredibly similar across the E1A gene. On the other hand, the EGR1 TF shows a generally high, but quite variable, distance across mutants, indicating it contributes differently depending on the mutation. In Figures 5F and 5G, the genes which PITX1 and EGR1, respectively, are interacting with are represented. The number of interactions which are unique to a specific mutant TRN are represented on the outer edge of each petal, while the number of genes common to two or more TRNs are represented on the inner edge of the petals. It is evident that, for the most part, there are more common interactions as compared to “unique” ones, as would be expected. Additionally, it is of interest that, for the same threshold, PITX1 has many more gene interactions in almost all mutant TRNs than does EGR1. While PITX1 has interactions in every single mutant TRN, EGR1 appears only in some, and seems to hold few interactions, both unique and common. This may be due to their differing roles, as shown previously in Figure 5D.

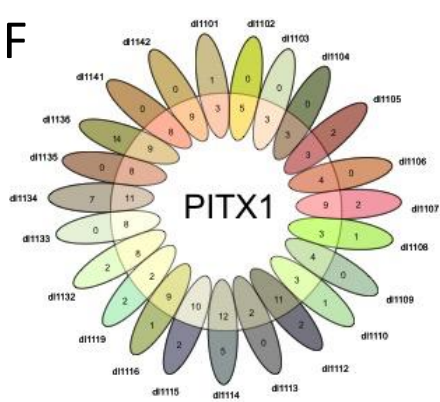
D



E



F



G

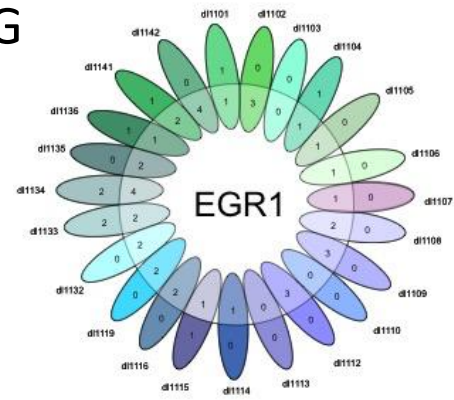


Figure 5: TRN Reconstruction and comparison (cont.) (D) Heatmap representation Jaccard distance values for each TF in each mutant TRN. The Jaccard distance was calculated by dividing the number of features found in either the original or mutant TRN by the number of features found in both TRNs. (E) Flower plot representing the unique and common interaction the TF PITX1 has in each mutant TRN. Interactions were deemed common when present in two or more TRNs. Otherwise, they were unique. (F) Flower plot representing the unique and common interaction the TF EGR1 has in each mutant TRN. Interactions were deemed common when present in two or more TRNs. Otherwise, they were unique.

3.4 Exploring the downregulated phenotype of *dll101* and *dll132*

Both the *dll101* and *dll132* mutants showed similar downregulated phenotypes in exploratory analysis (Figure 1). It is interesting to compare how their TRNs may differ (or be more similar). Figure 6A illustrates the TRN built for the *dll101* mutant, with an interaction threshold of 0.115. The most highly connected TFs are CUX1, PITX1 and KLF10, in descending order. Both the KLF10 and PITX1 TFs appear to contribute similarly in several TRNs, as is seen in Figure 5C. However, CUX1 appears to contribute differently in *dll101-1105* specifically. It is regulating 12 different targets in this network and appears to have more downregulatory rather than upregulatory interactions. The expression of the CUX1 gene (RPKM) across all mutants is illustrated in Figure 6B. Expression in the *dll101* mutant is especially high, with levels in all other mutant being close to background. They are comparable to the CUX1 expression levels seen in the uninfected sample. This is also true of the CYTH2 gene, which is a closely related interaction partner of CUX1 (see Figure 6C). The expression pattern of CYTH2 in all mutants shows an upregulation in *dll101*. This extreme trend is not seen in any other TFs (of the 20 chosen for TRN reconstruction), for any of the mutants. These Figures are shown in the supplementary section. The TRN representation of *dll132* is shown in Figure 6D, with a threshold of 0.115, once again. Visually, it is evident that this TRN contains more positive interactions than seen for *dll101*, which is surprising, given the overall downregulatory phenotype seen previously. The two most highly connected TFs in this case are PITX1 and CREB3L2, both of which appear to have mostly uniform contributions to TRNs across the board (Figure 5D). Additionally, gene expression of these two TFs in mutants doesn't show a strongly biased response towards *dll132* which may suggest that these TFs are targeted by E1A to modulate gene expression rather than indirectly

upregulated which would allow for increased promoter binding. As an effort to better compare both TRNs, the graphs (given the same filtering thresholds) were used to calculate the graph edit distance (GED). This quantitative measure allows for an estimate of the similarity between two graphs to be estimated. The GED calculates the minimal number of “steps” or changes which need to be made to the first TRN network in order to get to the second network. Figure 6E demonstrates the comparison between both *dll101* and *dll132* being 165 steps, which is markedly lower than the distance which is seen when either of these TRNs are compared to the original TRN (589 and 568 for both *dll101* and *dll132*, respectively). This is indicative of the potential similarity of these graphs to one another, and perhaps the functional roles of these regions within E1A.

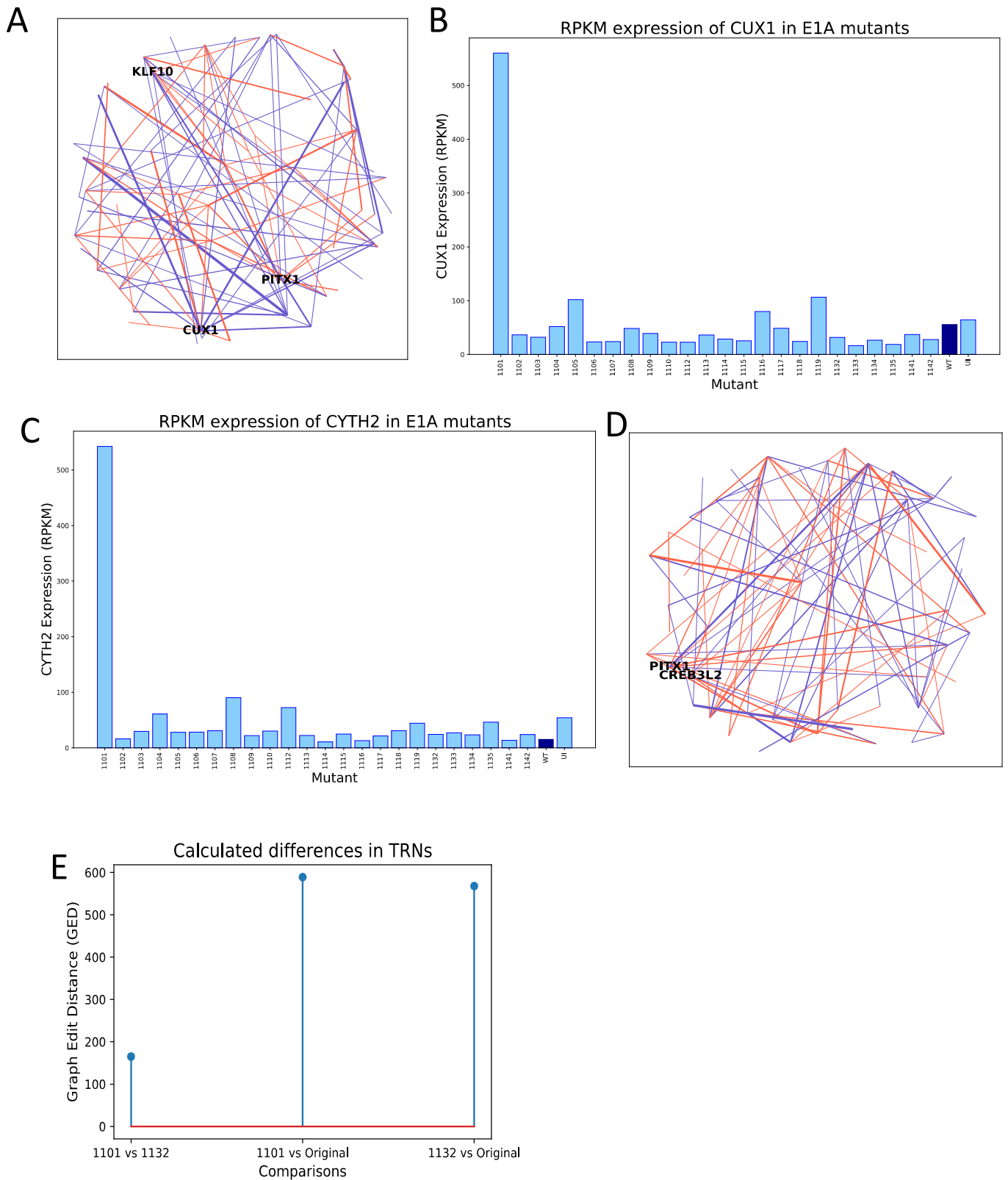


Figure 6: Exploring *dl1101* and *dl1132* TRNs. (A) TRN network representation for *dl1101*. (B) RPKM expression of the CUX1 gene across all mutants. The WT virus is highlighted in darker blue to act as a comparison. (C) RPKM expression of the CYTH2 gene across all mutants. The WT virus is highlighted in darker blue to act as a comparison. (D) TRN network representation for *dl1132*. (E) Stem plot shows the calculated graph edit differences as shown on the x axis. This algorithm calculates the required minimal number of steps to get from one graph to another. When running this algorithm, a maximum number of 200 secs was given to arrive to an answer.

3.5 Exploring the upregulated phenotype of *dl1113-dl1116*

Having previously looked at the seemingly downregulated *dl1101* (Δ 4-25) and *dl1132* (Δ 224-238) mutants, we next take our attention to the E1A subsection covered by *dl1113* (Δ 169-177) - *dl1116* (Δ 205-221), which covers AAs 169-221. The TRN representation for *dl1113* (Δ 169-177) is illustrated in Figure 7A. Both the PITX1 and CRE3L2 TFs are highlighted as being highly connected once again. Interestingly, TCF21 also appears important with both a very high number of connections, as well as one extremely strong positive regulatory interaction with the gene RFC2. The regulatory interactions shown do not appear distinctly upregulated, as was expected from this section, although it is evident there is more upregulation than is found in the original TRN (Figure 5B). Interestingly, when looking at the visual representation of the TRN for *dl1114* (Δ 178-184), shown in Figure 7B, two of the most highly connected TFs are identical to *dl1113* (Δ 169-177), that is PITX1 and TCF21. Here, the most highly connected TF, by far, is not either of the above, but instead PA2G4. It appears to have a wide variety of both stronger/weaker positive and negative regulatory effects. PITX1 is again present in the TRN constructed for *dl1115* (Δ 188-204). Among these TFs highly involved is SMAD3, which is unique to this TRN (Figure 7C). Despite these TRNs having identical interaction thresholds to determine the number of interactions displayed, this TRN shows few positive interactions, and many more negative ones. The same can be said of *dl1116* (Δ 205-221), pictured in Figure 7D, which shows a sparse structure, with few interactions strong enough to pass the filtering threshold imposed, and even fewer of these being positive regulatory effects. PITX1, ZBTB20 and NFE2L2 are the most highly connected TFs here, none of which appear to contribute differently across the genome, similarly to PITX1, as discussed earlier. The

graph edit distance was calculated for all combinations of these 4 TRNs with each other as well as with the original TRN, as shown in Figure 7E. TRNs constructed for *dl1113*, *dl1114*, *dl1115* or *dl1116* appear to be much more similar to one another, when directly compared. However, when the GED was calculated for each mutant TRN, as compared to the original TRN, the distance found was uniformly larger. Figure 7F is a visual representation of the most highly connected TFs in all four mutants. TFs which are shared amongst the mutants are more likely to highly contribute to the general section phenotype. Here, the importance of PITX1 is evident in that all four TRNs (*dl1113*, *dl1114*, *dl1115*, *dl1116*) feature it prominently (Figure 7F). Additionally, the TF TCF21 is shared between the TRNs constructed for *dl1113* and *dl1114*, while CREB3L2 was shared by TRNs built for *dl1113* and *dl1115*, which is indicative of potentially shared roles or regulators among these regions of the E1A gene (Figure 7F). This Venn diagram highlights the close relationship shared by *dl1113*, *dl1114* and *dl1115*. On the other hand, *dl1116* does not seem nearly as closely related with only PITX1 as a related TF, despite its GED seemingly close to the other TRNs (Figures 7E-F).

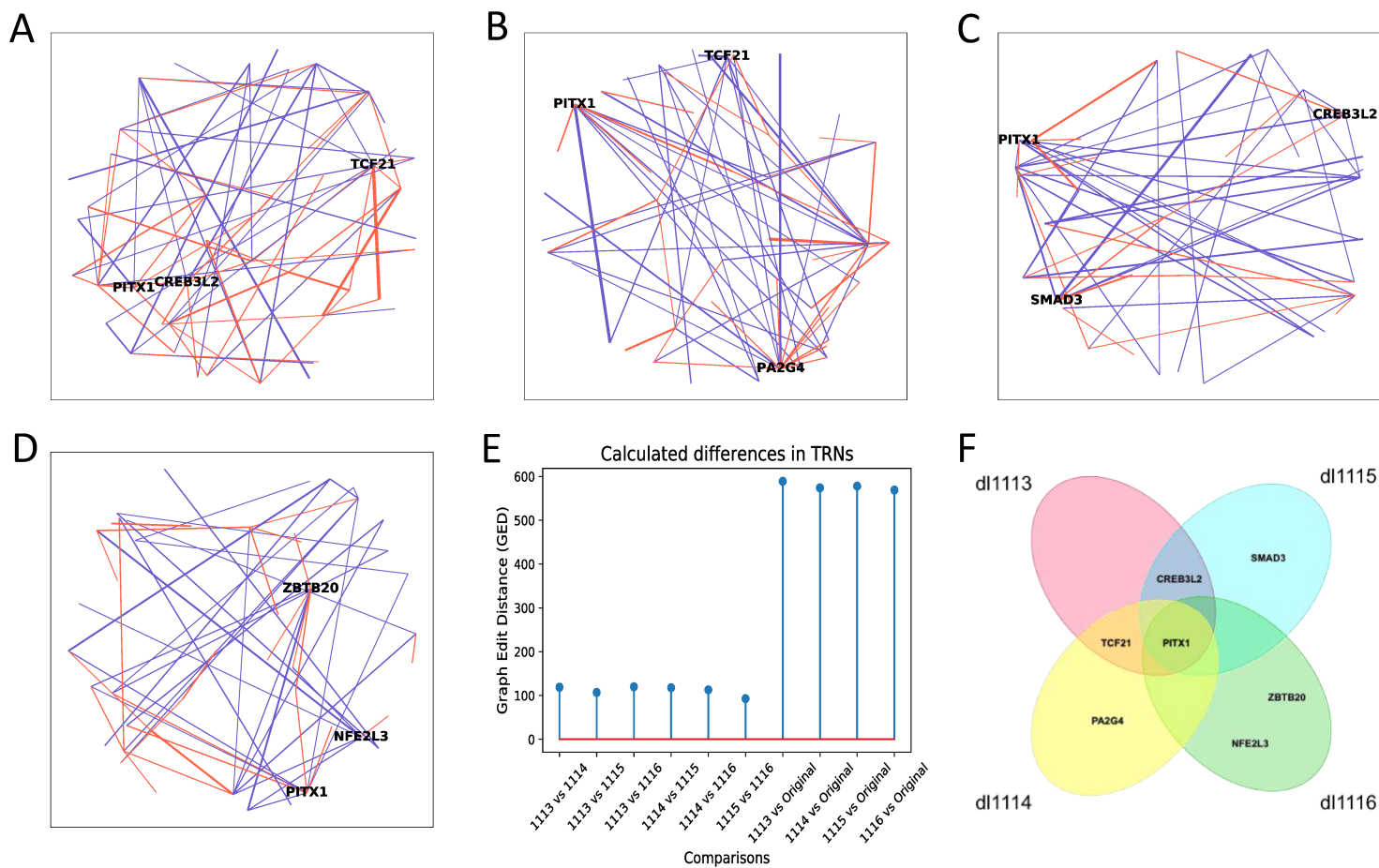


Figure 7: Exploring *dl1113-dl1116* TRNs.. (A) TRN network representation for *dl1113*. (B) TRN network representation for *dl1114*. (C) TRN network representation for *dl1115*. (D) TRN network representation for *dl1116*. (E) Stem plot shows the calculated graph edit differences as shown on the x axis. This algorithm calculates the required minimal number of steps to get from one graph to another. When running this algorithm, a maximum number of 200 secs was given to arrive to an answer. (F) Venn diagram for all four discussed mutants demonstrating the intersections of highly connected TFs.

3.6 Motif Analysis reveals potentially important motifs in a global manner

To control the expression of genes, TFs must bind to regulatory sequences in DNA. These TF binding sites often contain conserved AA sequences which allow the binding to occur. These are known as DNA binding motifs. To better understand the regulatory interactions at hand, motif enrichment analysis was performed based on the union of previously determined differentially expressed genes for all mutants, as compared to the WT virus. The promoter regions of these genes were searched for enriched (or overrepresented) motifs. Figure 8A represents the output resulting from the use of HOMER's findMotifs program, and depicts the most enriched DNA binding motif as the CHR site (cell cycle genes homology region), a DNA element with a role in late cell cycle signaling. The top 25 most enriched motifs are shown in Figure 8B. Visually, it is evident that 4 motifs appear to be more present than the rest, that being CHR, E2F, NFY and E2F. Each of these identified factors appear to play important roles in cell growth, differentiation as well as cell cycle regulation. Next, motif enrichment analysis was performed for each mutant separately, by analyzing the promoter regions of each of the separate lists of differentially expressed genes. The top 50 most highly enriched motifs in all mutants are shown in Figure 8C. Interestingly, the 3 most highly expressed motifs are in conjunction with the previous analysis done with the union of all genes, that is CHR, NFY and E2F, with an apparent importance in mutants *dll101*, *dll102*, *dll113*, *dll114*, *dll115* and *dll116*, which all appear to be driving the apparent importance of these binding sites (Figure 8C). Other motifs of interest include the TATA box binding protein interaction motif, which is known to interact with E1A, as well as SMAD3's binding motif, as this was one of the TFs previously chosen for TRN reconstruction (Figure 8C).

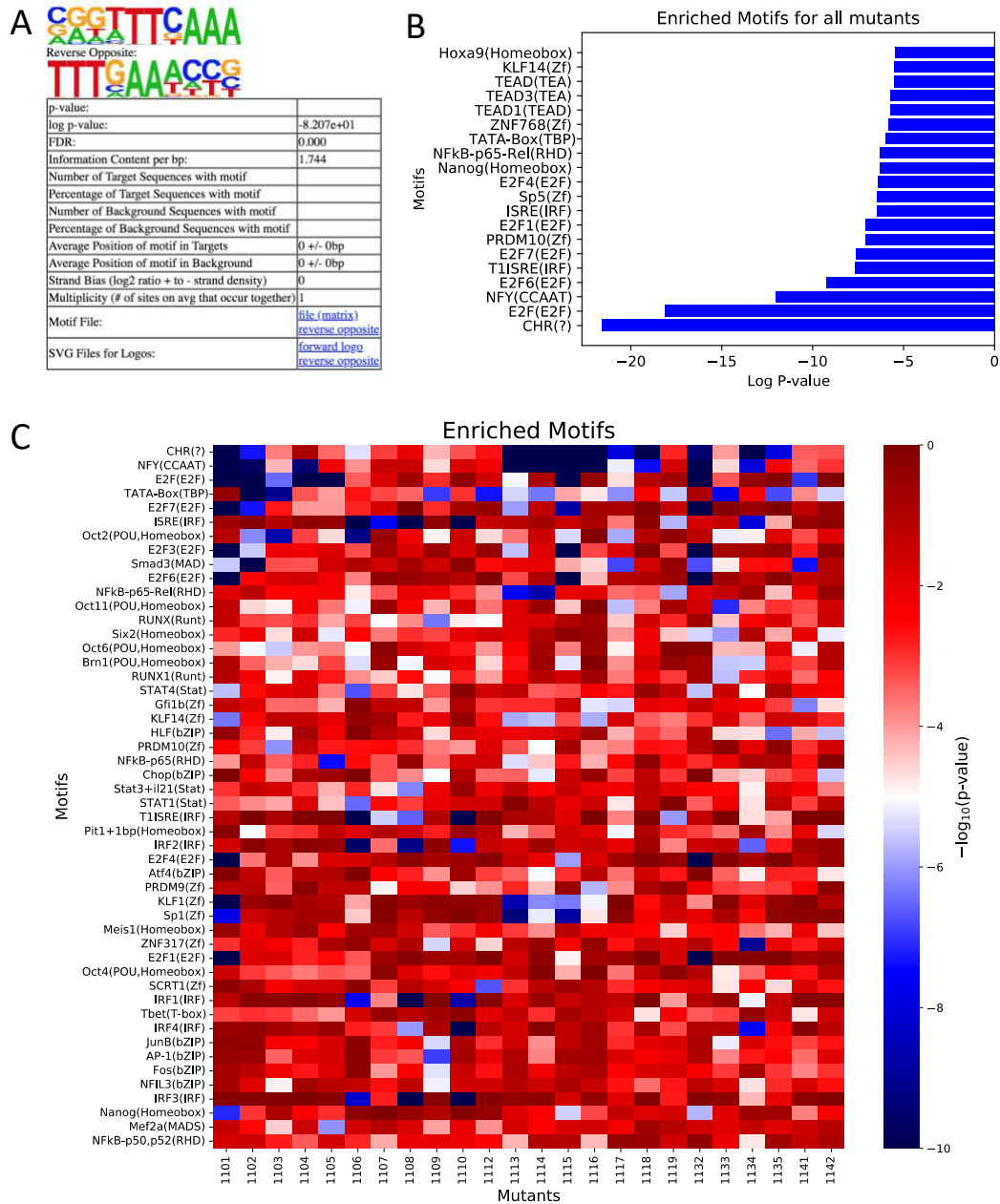


Figure 8: Motif Analysis for all mutants. (A) Output of HOMER's findMotifs, for the most highly enriched motif, CHR. (B) Bar graph demonstrates log p-value of top 25 most enriched motifs, for the union of all mutants. (C) Heatmap represents the log transformed p-values associated with the top 50 most enriched motifs, for each E1A deletion mutant. Blue indicates lower p-value, while red indicates a higher value.

4 - Discussion

4.1 E1A as a molecular hub protein

Adenovirus, and its E1A proteins, have been widely studied in the past, with many works focusing on infection kinetics, or molecular interactions with E1A. Additionally, mutational analyses have further supported structural data suggesting a largely disordered structure for the E1A protein, as small insertions or deletions fail to affect the global E1A phenotype. Rather, these mutations have effects on specific subsets of the function only⁹⁵. Additionally, it was shown that recombinant E1A can be fully denatured by boiling and will retain its ability to induce transcription and translocate to the nucleus¹⁴¹. Upon comparison of the CR1 and CR2 regions of E1A to the E7 protein of papillomavirus 45 (its own molecular hub protein), they were found to share some sequence similarity¹⁴². Disordered regions have been found in E7, further pointing to this same disorder being present in E1A¹⁴³. Disentangling this lack of structure is important, as it allows for a deeper understanding of the molecular mechanisms which could be used to potentially target this protein. The way by which E1A can bind to and control the function of many of these cellular factors concurrently is something which still requires research. In this work, we attempt to dissect and disentangle the roles of different portions of the E1A protein through the analysis of TRNs reconstructed for several small E1A deletion mutants.

A preliminary examination of RNA expression in all deletion mutants reinforces the disorder which is hypothesized to be a feature of E1A. Although deletions made ranged from approximately 3 to 20 AAs, all mutants were found to have differing RNA expression profiles compared to the WT virus (Figure 3A).¹⁴⁴ One aspect of this exploration which stood out was the overwhelmingly negative/down-regulation which was seen in many

genes and pathways. This may point to the importance of E1A in its role as a molecular hub protein and transcriptional activator.

4.2 E1A control of cell cycle as a tumor suppressor and oncogene

E1A holds dual roles as both a tumor suppressor and an oncogene⁵⁵. In the case of oncogene properties, to ensure the successful transformation of rodent cells, both the E1A and E1B viral genes must be present⁶. However, E1A can also transform normal cells in cooperation with host oncogenes, such as activated RAS⁶. Further, studies using oncolytic adenovirus vectors for therapeutic means have extracted and explored the roles of E1A and E1B, separately from one another¹⁴⁵. It was found that E1A alone was sufficient in maintaining both the oncolysis and replication of replication-competent adenoviruses (RCAd). It was also able to mediate the replication and transduction of replication-deficient adenoviruses (RDAd). On the other hand, both the E1B proteins studied were found to have opposite effects, with E1B-19 kDa enhancing the effects of E1A, and E1B-55 kDa antagonizing its effects¹⁴⁵. In the case of tumour suppressor properties, E1A has also been shown to reverse the transformed phenotype of human tumor cells and repress the growth of primary human tumors¹⁴⁶. This finding seemed to suggest that, contrary to popular belief at the time, the E1A protein may act as a tumor suppressor. Further studies found that overexpression of E1A induces apoptosis in host cells. This directed researchers back towards the oncogene theory, in that most growth-promoting oncogenes also tend to induce apoptosis, a mechanism which must be countered by tumor cells in the context of mutations. Another feature of E1A which hints to its importance as a tumor suppressor is its ability to convert human tumor cell lines into cells displaying an epithelial phenotype. This is the opposite of what we would expect from an oncogene, which typically mediates the epithelial to mesenchymal transition, or EMT¹⁴⁷. Given what is now known about E1A,

its interactions with host cellular proteins, and the overall plasticity of its role, perhaps it cannot be referred to as either oncogene or tumor suppressor, in the classical sense. This evidence points to the overarching importance of E1A in regulating and activating several different molecular hubs, which would begin to explain why small deletions of its functional regions may lead to downregulation.

As mentioned, due to its important dual role in oncogenesis, E1A must have a tight handle on cell cycle related genes to induce these cell wide phenotypic changes which will inevitably increase the virus production ability of the host. The duality of its role may begin to explain the differences seen in our study, in which cell cycle related genes were differentially regulated depending on the E1A region which was deleted (Figures 3D-E). While mutations made in the region between AAs 169-221 appeared to greatly increase expression of cell cycle genes and lead to enrichment of cell cycle related GO terms, mutations made from AAs 1-168 appeared to lead to the opposing phenotype. However, in support of tumor suppressive activities, E1A regions also had control over apoptosis and cell death. Not unexpectedly, mutations found in similar regions of the E1A protein appeared to cluster closely with one another when using a hierarchical clustering method based on gene expression. While spatially related regions share similar activities, they still maintain distinct effects on gene expression with specific differential gene sets (Figure 4E). This points to the importance of efficient viral packaging, as viruses have very little genomic space with which to encode all their functions. Thus, there are few redundancies, and every small section serves a purpose¹⁴⁸. However, while redundancies are few, they do exist. Several of E1A's binding partners, such as pRb, p300/400 and even E2F reportedly require interactions with multiple regions of the E1A protein. This may explain why certain

mutations occurring in distanced regions of the protein still appear to cluster closely to one another, as they affect similar functions. For instance, *dll102*, which is found in the N-terminus region of E1A clusters closely with *dll104* (Figure 4E), found in CR1. Both regions appear important in binding to p300/400⁷¹. Importantly, these distinctions are only observable when using appropriate clustering methods. This was not something which was seen when PCA was used, with the caveat that use of a PCA requires a strong correlation between samples to present defined clusters. If this is not the case, its output is nonsensical, as is the case here. On the other hand, hierarchical clustering methods will always find and calculate clusters. This should be considered when determining the validity of the clusters identified here¹⁴⁹.

4.3 Viral Gene Expression

In our study, only two viral genes were found to be expressed, that is VAI and VAII (Figure 3G). As only the late time point of the early phase of adenoviral infection, that is the 72-hour mark, was assayed, it is expected that the bulk of RNA expression has not yet occurred to a degree by which it can be measured. While many viral genes were not present in the analysis of RNA-Sequencing data, this does not indicate they were completely absent, but simply that they may not be highly expressed enough for their presence to be detected by the protocol used. Further studies are warranted, in which later time points should be assayed. VAI and VAII are small virally encoded RNA molecules, which have dsRNA structure like miRNAs. They form tightly structured stems which confer resistance to cellular defense systems¹⁵⁰. Throughout infection, they accumulate in the cytoplasm where they block expression of specific reporter genes through binding of complementary sequences¹⁵¹. Prior studies have demonstrated they are necessary to ensure efficient protein synthesis during late adenoviral infection. For instance, VAI itself has been shown to

promote an accumulation of its own transcript, which may explain its heightened presence in this case¹⁵². This evidence may indicate an abundance of VAI and VAII as compared to other viral genes, which can represent the reasoning behind their solitary presence in this analysis. When comparing the mutants to one another, it was evident that many demonstrated similar VAI/VAII expression when compared to the WT virus. However, several mutants, notably, *dl1101*, *dl1132* and *dl1135* had incredibly low viral gene expression, which may indicate they hold important functions in viral gene expression and regulation.

4.4 TRN construction

In first constructing our original TRN (Figure 5E), which is inclusive of all regulatory information from all mutants, two TFs came to light.

4.4.1 PA2G4

First, while many of the strong regulatory interactions mapped appeared to be negative, PA2G4 itself appeared to be a strongly connected regulator, with mostly positive interactions. Also known as proliferation-associated G4, PA2G4 encodes an RNA binding protein which is involved in growth regulation. It is heavily associated with several cancers, with both oncogene and tumour suppressor roles. Additionally, this protein has a predicted amphipathic helical domain allowing it to interact directly with DNA and proteins. It holds an important role in binding to the E2F1 promoter element, with the help of the Rb protein and Sin3A¹⁵³. It appears that PA2G4 is positively regulating several genes related to cell cycle pathways, which may indicate its role could be dampened in a fully functional E1A protein, to ensure proper cell cycling according to viral needs. Although there is currently no strong association between E1A and PA2G4, the gene has been found to be one of many

differentially expressed genes involved in proliferation during the early phase of adenovirus infection ¹⁵⁴.

4.4.2 PITX1

Another TF of interest is PITX1, also known as paired-like homeodomain transcription factor 1. This TF appears to be similarly involved in TFs constructed for most, if not all, mutants, which could indicate either an essential role for this TF throughout E1A, or an unimportance. PITX1, as such, has not been associated with adenovirus, or E1A, but is hypothesized to play an important role in cancer formation, as a tumour suppressor. It is a known repressor of the *hTERT* gene, which leads to the inhibition of elongation of the telomere repeat sequences necessary to cellular immortalization¹⁵⁵. In certain cells, PITX1 has even been reported to increase expression of the p53 gene at both the mRNA and protein levels, a gene which negatively regulates cell division, an process which would negatively affect the virus¹⁵⁶. The seemingly important role of PITX1 in regulating cell proliferation may be one of the aspects which makes it so uniformly involved in all TRNs constructed here.

4.4.3 EGR1

The EGR1 gene (early growth response factor 1) was also highlighted in our study as being differentially functional in many of the mutant TRNs, with a strong contribution when deletions were made from AAs 20-168 and 205-221. Essentially, its behavior opposes that of PITX1. EGR1 plays a dual role in different signaling pathways, and has also been shown to have an importance in cancer formation. It has been found to regulate the attachment and survival of normal cells, but to induce apoptosis in abnormal cells with

decreased adhesion¹⁵⁷. In several studies, EGR1 has also been shown to increase proliferation of tumor cells through modulation of the cell cycle^{158,159}. These roles may indicate the importance of EGR1 in the context of E1A, in that E1A fulfills an important role in cell cycle function, namely the induction of S phase to allow for virus multiplication. The DNA binding region of EGR1 is composed of 3 classical zinc finger motifs and is commonly found co-localized with CpG islands throughout the genome. ChIP-chip studies have shown that this TF may bind to a variety of promoters of actively transcribed genes throughout the genome, and, through its binding, may change the current occupancy of these sites by other binding proteins¹⁶⁰. This is indicative of the various number of functions this TF can enact. EGR1 may interact with several regions of E1A, possible both directly and indirectly, to enact many different types of functions, explaining why its functions appear to change depending on the mutant region.

4.4.4 *dl1101* vs. *dl1132* – The downregulated phenotype

In previous studies, a deletion in *dl1101* was shown to decrease the hepatotoxicity of adenoviral therapy vectors greatly, indicating an important role in viral replication¹⁶¹. This is backed up by viral kinetics studies which have shown *dl1101* to have a significantly reduced ability to synthesize viral genomes as well as to induce S phase in host cells⁵⁶. Additionally, induction of cell cycle genes in *dl1132* was significantly lower than in the WT virus¹¹⁶. This evidence aligns with our findings in which both the *dl1101* mutant, which is a deletion of AAs 4-25, and the *dl1132* mutant, which is deletion of AAs 224-238, appear to have a visibly downregulated phenotype both in terms of host and viral genes. It should not be ignored that both of these mutants also demonstrate a reduced ability to induce viral gene expression, as compared to the WT virus⁵⁶. Thus, it may be possible that

the “downregulated” phenotype observed here is simply due to the lack of an ability to induce viral gene expression, as opposed to a specific phenotype relating to the cell cycle related genes. This caveat should be considered in the rest of this section. When TRNs reconstructed for both mutants are compared using the quantitative GED approach, both mutant TRNs are markedly closer to one another than they are to the wildtype E1A TRN (Figure 6E). However, this is expected given that, as the wildtype E1A TRN includes regulatory information from all mutants at once, it is more likely to contain a larger number of regulatory interactions which surpass the filtering threshold. As such, it is likely that the graph will be less sparse, and thus always largely different from any of the mutants. Although both *d11101* and *d11132* demonstrate downregulated phenotypes, the evidence present cannot allow us to determine whether they perform similarly within the context of E1A regulation. In fact, it is much more likely that, while both sections are important in the positive regulation of genes in a normally functioning E1A, they do so in completely different ways. In fact, the most prominent TF in the TRN constructed for *d11101* is CUX1, which is a TF with four different DNA-binding domains (Figure 6A). Prior studies have shown an importance in cancer formation, especially relating to CUX1’s role in cell proliferation, differentiation and migration¹⁶². Many of CUX1’s downstream target genes have also been reported to play a role in cell cycle progression, including CYTH2, which is involved in cytoskeletal rearrangement¹⁶³. As shown previously, the mRNA levels of both CYTH2 and CUX1 across all mutants showed an incredibly highly skewed response towards *d11101*, which may indicate the importance of this N-terminal region of E1A in suppressing the activity of this TF.

4.4.5 *dl1113-dl1116* – The upregulated phenotype

While both *dl1101* and *dl1132* demonstrated interestingly downregulated phenotypes, *dl1113* ($\Delta 169-177$), *dl1114* ($\Delta 178-184$), *dl1115* ($\Delta 188-204$) and *dl1116* ($\Delta 205-221$) all showed an upregulated phenotype, more specifically regarding the cell cycle related genes. Mutations within this region almost seem to garner a gain of function phenotype. It is important to consider that while *dl1115-1116* are both found within the second exon of the E1A protein, *dl1113-1114* are mutations made within the unique region of E1A, also known as conserved region 3. These regions may contribute more to this overall phenotype, something which is confirmed by abundant interactions present within their respective TRNs. This region is only present in the largest protein isoform (13S) and has important roles in transactivation, although these roles have been stipulated to extend up to AA 204, which would also include *dl1115*⁹⁶. This may start to explain the closer relationship observed between these three graphs in terms of common TFs (Figure 7F). In prior kinetics studies, *dl1116* mutants have shown viral titers and genome counts comparable to that of WT. However, expression of related viral genes was found to be downregulated¹¹⁶. It is interesting that a deletion of either of these 4 deletion spots would lead to such increased expression of cell cycle related genes, especially given this section plays a role in transactivation. This could be related to the virus' need to slow or dampen the cell cycle, to ensure that necessary resources for viral replication are efficiently allocated instead of being used for cellular division. In fact, according to previous studies, these mutations have a negative effect on CR3 transactivation, with the exception of *dl1116*⁹⁶. In fact, there is little known about *dl1113-1115*, as shown in Table 1. On the other hand, *dl1116* shows a poor ability to induce viral gene expression⁵⁶. This evidence points towards all 4 mutants having downregulatory phenotypes. The exact opposite of this

effect is observed within the cell cycle related genes, thus indicating this could be a true effect and not simply related to increased viral replicaton (as is seen for *dll101* and *dll132*).

The construction of TRNs here was performed using only RNA-Sequencing data, which does incur some caveats. Measuring RNA expression can be incredibly useful as it provides information regarding the flow of genomic information through to the mRNA state. However, it fails to account for the state of chromatin and the availability of TF binding sites on the genome¹⁶⁴. This type of information is critical to the proper selection of TFs in the reconstruction process. As such, it becomes important to evaluate the results obtained from such TRNs critically, and to validate any assumptions with further confirmation experiments.

4.5 TF binding to DNA motifs

When analyzing regulation at the transcriptional level, it becomes important to consider not only gene expression, which can give information as to which genes are or are not being stimulated, but also to the elements which ensure their regulation. TFs themselves bind to discrete sets of DNA sequences, which are referred to as a binding motif¹⁶⁵. Extracting information about the available DNA binding motifs within a specific condition allows for a better estimation of the regulatory interactions occurring. One such enriched motif, as per the totality of the genes, is the cell cycle genes homology region (CHR), which has been identified as a DNA element with an important role in transcriptional regulation of late cell cycle genes (Figure 8A). As the prominent role of E1A in regulation of the cell cycle has already been established, this is hardly surprising. There is a large spectrum of CHR elements within the genome, with four different variants of the motif known. Of the four, one motif, TAGGAA has been found to be functional with an adjacent E2F binding

site¹⁶⁶. Moreover, several E2F binding domains also appear in the list of enriched domains. Of interest is the fact that PA2G4, one of the most highly influential TFs found in our original TRN is known to bind to E2F binding motifs as well. This highlights the potential importance of this TF in the scope of E1A's functionality. Of the other enriched motifs found in this analysis, it should be noted that there were no other found matches between TFs and their motifs, from the TFs selected for TRN reconstruction (Figure 8C). This is not altogether surprising given the importance of transcriptional regulation not only through *cis*, or local mechanisms, but also through *trans* pathways. TFs may bind at the promoter regions, as were queried in this motif enrichment analysis, but there also several interactions which occur at distal enhancer sites, which are unaccounted for here¹⁶⁴.

4.6 Future Directions

In future works, we look forward to improving upon our established pipeline in order to achieve more meaningful results, and to confirm the observations made thus far. First, it would be of interest to repeat the reconstruction of TRNs, but with the use of either ATAC or ChIP sequencing information. Chromatin accessibility is crucial to ensure gene expression and its regulation, and both these techniques represents ways by which this accessibility could be queried¹⁶⁷. Motif enrichment analysis should then be repeated, to detect overrepresented DNA binding motifs found within the accessible regions of the human genome. These motifs can then aid in identifying TFs which are more likely to be phenotypically relevant. In this case, TFs were chosen based on mRNA expression alone, which may not represent their ongoing functionality. In fact, some groups have even started using epigenetic data (ChIP-Seq) to better inform the modeling of network relationships. These computational methods allows for the detection of events of co-regulation in which

two TFs are acting concurrently to enact gene regulation¹⁶⁸. This may represent a potential avenue for future studies, in which the reconstruction of TRNs can be based on the exchange of information between epigenetically accessible motifs. Additionally, our current analysis focuses on gene expression at a single time point of adenovirus infection, that is 72 hours after infection. In IMR90 cells, as were used here, this represents late infection. It would be of interest to collect data at several other time points, such as 24 and 48 hours. This would allow for a more nuanced understanding of the transcriptional regulatory networks which surround E1A's function, and the way by which they change throughout infection. It has been established that E1A has several functions, all of which are important at various time points throughout infection, such as activation of viral early gene expression, induction of cellular S-phase and immune evasion, to name a few⁵⁷. As such, it is logical that E1A would modify its regulatory interactions to ensure only the correct regulatory modules are activated at the precisely correct time to ensure viral reproduction. Lastly, this study has identified several putative regulators of E1A induced TRNs (CUX1, PA2G4, PITX1, EGR1). However, with the predictive nature of the work done here, it gets important to validate such approximations through experimental work. The intent is to repeat the current experimental setup, with cells expressing specific TF K/O. This would allow for an analysis of E1A's function without the presence of a specific TF, resulting in a confirmation of our findings.

5 – Conclusion

In this work, we have demonstrated the diversity of interactions and regulatory effects which E1A may have within host cells, further cementing its importance as a molecular hub protein. We have identified several TFs previously unrelated to the HAdV-5 E1A protein, that is PITX1, CUX1, EGR1 and PA2G4. Confirmation experiments should be performed to solidify the importance of these proteins in adenoviral infection. Lastly, important TF binding motifs were identified in the promoter regions of differentially regulated genes. These provide a preliminary insight into the DNA binding which occurs in these experimental conditions. A deeper analysis using genome accessibility information will be required. In the future, this pipeline can be used for a longitudinal time course analysis of the infection. This work has opened avenues for studying adenovirus' E1A protein more closely, as well as other viral infections and related proteins, using TRNs. The importance of this advancement can be clearly found in the potential of having the ability to determine the exact regulatory mechanisms in play during infection.

6 - Bibliography

- 1 Jeong, H., Mason, S. P., Barabási, A. L. & Oltvai, Z. N. Lethality and centrality in protein networks. *Nature* **411**, 41-42, doi:10.1038/35075138 (2001).
- 2 Purushothaman, P., Uppal, T. & Verma, S. C. Human DNA tumor viruses and oncogenesis. *Animal Biotechnology*, 131-151, doi:10.1016/B978-0-12-811710-1.00007-0 (2020).
- 3 Chaitanya, K. V. Structure and Organization of Virus Genomes. *Genome and Genomics*, 1-30, doi:10.1007/978-981-15-0702-1_1 (2019).
- 4 Khanal, S., Ghimire, P. & Dhamoon, A. S. The Repertoire of Adenovirus in Human Disease: The Innocuous to the Deadly. *Biomedicines* **6**, doi:10.3390/biomedicines6010030 (2018).
- 5 Lynch, J. P., 3rd & Kajon, A. E. Adenovirus: Epidemiology, Global Spread of Novel Serotypes, and Advances in Treatment and Prevention. *Semin Respir Crit Care Med* **37**, 586-602, doi:10.1055/s-0036-1584923 (2016).
- 6 Miller, D. L., Myers, C. L., Rickards, B., Collier, H. A. & Flint, S. J. Adenovirus type 5 exerts genome-wide control over cellular programs governing proliferation, quiescence, and survival. *Genome Biol* **8**, R58, doi:10.1186/gb-2007-8-4-r58 (2007).
- 7 Davison, A. J., Benkő, M. & Harrach, B. Genetic content and evolution of adenoviruses. *Journal of General Virology* **84**, 2895-2908, doi:<https://doi.org/10.1099/vir.0.19497-0> (2003).
- 8 Trentin, J. J., Yabe, Y. & Taylor, G. The quest for human cancer viruses. *Science* **137**, 835-841, doi:10.1126/science.137.3533.835 (1962).
- 9 Gallimore, P. H. & Turnell, A. S. Adenovirus E1A: remodelling the host cell, a life or death experience. *Oncogene* **20**, 7824-7835, doi:10.1038/sj.onc.1204913 (2001).
- 10 Gallimore, P. H., McDougall, J. K. & Chen, L. B. In vitro traits of adenovirus-transformed cell lines and their relevance to tumorigenicity in nude mice. *Cell* **10**, 669-678, doi:10.1016/0092-8674(77)90100-3 (1977).
- 11 . (!!! INVALID CITATION !!! 4).
- 12 Wu, P.-Q. *et al.* Clinical manifestations and risk factors of adenovirus respiratory infection in hospitalized children in Guangzhou, China during the 2011-2014 period. *Medicine (Baltimore)* **99**, e18584-e18584, doi:10.1097/MD.00000000000018584 (2020).
- 13 Walsh, M. P. *et al.* Computational analysis of two species C human adenoviruses provides evidence of a novel virus. *J Clin Microbiol* **49**, 3482-3490, doi:10.1128/JCM.00156-11 (2011).
- 14 Chow, L. T., Gelinas, R. E., Broker, T. R. & Roberts, R. J. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* **12**, 1-8, doi:10.1016/0092-8674(77)90180-5 (1977).
- 15 Condezo, G. N. & San Martín, C. Localization of adenovirus morphogenesis players, together with visualization of assembly intermediates and failed products, favor a model where assembly and packaging occur concurrently at

- the periphery of the replication center. *PLOS Pathogens* **13**, e1006320, doi:10.1371/journal.ppat.1006320 (2017).
- 16 van Raaij, M. J., Mitraki, A., Lavigne, G. & Cusack, S. A triple beta-spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein. *Nature* **401**, 935-938, doi:10.1038/44880 (1999).
 - 17 Sohn, S.-Y. & Hearing, P. Adenoviral strategies to overcome innate cellular responses to infection. *FEBS Letters* **593**, 3484-3495, doi:<https://doi.org/10.1002/1873-3468.13680> (2019).
 - 18 Greber, U. F. & Flatt, J. W. Adenovirus Entry: From Infection to Immunity. *Annu Rev Virol* **6**, 177-197, doi:10.1146/annurev-virology-092818-015550 (2019).
 - 19 Zhang, Y. & Bergelson, J. M. Adenovirus receptors. *J Virol* **79**, 12125-12131, doi:10.1128/jvi.79.19.12125-12131.2005 (2005).
 - 20 Wodrich, H. *et al.* Adenovirus core protein pVII is translocated into the nucleus by multiple import receptor pathways. *J Virol* **80**, 9608-9618, doi:10.1128/jvi.00850-06 (2006).
 - 21 Georgi, F. & Greber, U. F. The Adenovirus Death Protein – a small membrane protein controls cell lysis and disease. *FEBS Letters* **594**, 1861-1878, doi:<https://doi.org/10.1002/1873-3468.13848> (2020).
 - 22 Kennedy, M. A. & Parks, R. J. Adenovirus Virion Stability and the Viral Genome: Size Matters. *Molecular Therapy* **17**, 1664-1666, doi:10.1038/mt.2009.202 (2009).
 - 23 Ahi, Y. S. & Mittal, S. K. Components of Adenovirus Genome Packaging. *Frontiers in Microbiology* **7**, doi:10.3389/fmicb.2016.01503 (2016).
 - 24 Guimet, D. & Hearing, P. in *Adenoviral Vectors for Gene Therapy (Second Edition)* (ed David T. Curiel) 59-84 (Academic Press, 2016).
 - 25 Charman, M., Herrmann, C. & Weitzman, M. D. Viral and cellular interactions during adenovirus DNA replication. *FEBS Letters* **593**, 3531-3550, doi:<https://doi.org/10.1002/1873-3468.13695> (2019).
 - 26 Saha, B., Wong, C. M. & Parks, R. J. The adenovirus genome contributes to the structural stability of the virion. *Viruses* **6**, 3563-3583, doi:10.3390/v6093563 (2014).
 - 27 Kulanayake, S. & Tikoo, S. K. Adenovirus Core Proteins: Structure and Function. *Viruses* **13**, 388 (2021).
 - 28 Hay, R. T. The origin of adenovirus DNA replication: minimal DNA sequence requirement in vivo. *Embo j* **4**, 421-426, doi:10.1002/j.1460-2075.1985.tb03645.x (1985).
 - 29 Kazlauskas, D., Krupovic, M. & Venclovas, Č. The logic of DNA replication in double-stranded DNA viruses: insights from global analysis of viral genomes. *Nucleic Acids Res* **44**, 4551-4564, doi:10.1093/nar/gkw322 (2016).
 - 30 Lam, E., Stein, S. & Falck-Pedersen, E. Adenovirus detection by the cGAS/STING/TBK1 DNA sensing cascade. *J Virol* **88**, 974-981, doi:10.1128/jvi.02702-13 (2014).

- 31 Barlan, A. U., Griffin, T. M., McGuire, K. A. & Wiethoff, C. M. Adenovirus membrane penetration activates the NLRP3 inflammasome. *J Virol* **85**, 146-155, doi:10.1128/jvi.01265-10 (2011).
- 32 Nociari, M., Ocheretina, O., Murphy, M. & Falck-Pedersen, E. Adenovirus induction of IRF3 occurs through a binary trigger targeting Jun N-terminal kinase and TBK1 kinase cascades and type I interferon autocrine signaling. *J Virol* **83**, 4081-4091, doi:10.1128/jvi.02591-08 (2009).
- 33 Weitzman, M. D. & Ornelles, D. A. Inactivating intracellular antiviral responses during adenovirus infection. *Oncogene* **24**, 7686-7696, doi:10.1038/sj.onc.1209063 (2005).
- 34 Stingele, J., Bellelli, R. & Boulton, S. J. Mechanisms of DNA-protein crosslink repair. *Nat Rev Mol Cell Biol* **18**, 563-573, doi:10.1038/nrm.2017.56 (2017).
- 35 Lau, L., Gray, E. E., Brunette, R. L. & Stetson, D. B. DNA tumor virus oncogenes antagonize the cGAS-STING DNA-sensing pathway. *Science* **350**, 568-571, doi:10.1126/science.aab3291 (2015).
- 36 Hendrickx, R. *et al.* Innate immunity to adenovirus. *Hum Gene Ther* **25**, 265-284, doi:10.1089/hum.2014.001 (2014).
- 37 Olanubi, O., Frost, J. R., Radko, S. & Pelka, P. Suppression of Type I Interferon Signaling by E1A via RuvBL1/Pontin. *J Virol* **91**, doi:10.1128/jvi.02484-16 (2017).
- 38 Sohn, S. Y. & Hearing, P. Adenovirus sequesters phosphorylated STAT1 at viral replication centers and inhibits STAT dephosphorylation. *J Virol* **85**, 7555-7562, doi:10.1128/jvi.00513-11 (2011).
- 39 Look, D. C. *et al.* Direct suppression of Stat1 function during adenoviral infection. *Immunity* **9**, 871-880, doi:10.1016/s1074-7613(00)80652-4 (1998).
- 40 Fonseca, G. J. *et al.* Adenovirus evasion of interferon-mediated innate immunity by direct antagonism of a cellular histone posttranslational modification. *Cell Host Microbe* **11**, 597-606, doi:10.1016/j.chom.2012.05.005 (2012).
- 41 Hwang, W. W. *et al.* A conserved RING finger protein required for histone H2B monoubiquitination and cell size control. *Mol Cell* **11**, 261-266, doi:10.1016/s1097-2765(02)00826-2 (2003).
- 42 Fonseca, G. J., Cohen, M. J., Nichols, A. C., Barrett, J. W. & Mymryk, J. S. Viral Retasking of hBre1/RNF20 to Recruit hPaf1 for Transcriptional Activation. *PLOS Pathogens* **9**, e1003411, doi:10.1371/journal.ppat.1003411 (2013).
- 43 Zemke, N. R. & Berk, A. J. The Adenovirus E1A C Terminus Suppresses a Delayed Antiviral Response and Modulates RAS Signaling. *Cell Host Microbe* **22**, 789-800.e785, doi:10.1016/j.chom.2017.11.008 (2017).
- 44 Castel, P., Rauen, K. A. & McCormick, F. The duality of human oncoproteins: drivers of cancer and congenital disorders. *Nat Rev Cancer* **20**, 383-397, doi:10.1038/s41568-020-0256-z (2020).
- 45 Todd, R. & Wong, D. T. Oncogenes. *Anticancer Res* **19**, 4729-4746 (1999).
- 46 Santos, E., Tronick, S. R., Aaronson, S. A., Pulciani, S. & Barbacid, M. T24 human bladder carcinoma oncogene is an activated form of the normal human homologue of BALB- and Harvey-MSV transforming genes. *Nature* **298**, 343-347, doi:10.1038/298343a0 (1982).

- 47 Rous, P. A SARCOMA OF THE FOWL TRANSMISSIBLE BY AN AGENT SEPARABLE FROM THE TUMOR CELLS. *J Exp Med* **13**, 397-411, doi:10.1084/jem.13.4.397 (1911).
- 48 Bishop, J. M. Viral oncogenes. *Cell* **42**, 23-38, doi:10.1016/s0092-8674(85)80098-2 (1985).
- 49 Haigis, K. M., Cichowski, K. & Elledge, S. J. Tissue-specificity in cancer: The rule, not the exception. *Science* **363**, 1150-1151, doi:doi:10.1126/science.aaw3472 (2019).
- 50 Flavahan, W. A., Gaskell, E. & Bernstein, B. E. Epigenetic plasticity and the hallmarks of cancer. *Science* **357**, doi:10.1126/science.aal2380 (2017).
- 51 Greten, F. R. & Grivnickov, S. I. Inflammation and Cancer: Triggers, Mechanisms, and Consequences. *Immunity* **51**, 27-41, doi:10.1016/j.immuni.2019.06.025 (2019).
- 52 Nelson, C. M. & Bissell, M. J. Of extracellular matrix, scaffolds, and signaling: tissue architecture regulates development, homeostasis, and cancer. *Annu Rev Cell Dev Biol* **22**, 287-309, doi:10.1146/annurev.cellbio.22.010305.104315 (2006).
- 53 Merino, M. M., Levayer, R. & Moreno, E. Survival of the Fittest: Essential Roles of Cell Competition in Development, Aging, and Cancer. *Trends Cell Biol* **26**, 776-788, doi:10.1016/j.tcb.2016.05.009 (2016).
- 54 Schneider, G., Schmidt-Suprian, M., Rad, R. & Saur, D. Tissue-specific tumorigenesis: context matters. *Nat Rev Cancer* **17**, 239-253, doi:10.1038/nrc.2017.5 (2017).
- 55 Frisch, S. M. & Mymryk, J. S. Adenovirus-5 E1A: paradox and paradigm. *Nature Reviews Molecular Cell Biology* **3**, 441-452, doi:10.1038/nrm827 (2002).
- 56 Costa, R. *et al.* Characterization of Adenovirus 5 E1A Exon 1 Deletion Mutants in the Viral Replicative Cycle. *Viruses* **12**, 213 (2020).
- 57 Radko, S., Jung, R., Olanubi, O. & Pelka, P. Effects of Adenovirus Type 5 E1A Isoforms on Viral Replication in Arrested Human Cells. *PLOS ONE* **10**, e0140124, doi:10.1371/journal.pone.0140124 (2015).
- 58 Graves, D., Akkerman, N., Bachus, S. & Pelka, P. Differential Splicing of Human Adenovirus 5 E1A RNA Expressed in cis versus in trans. *J Virol* **95**, doi:10.1128/jvi.02081-20 (2021).
- 59 Lillie, J. W., Loewenstein, P. M., Green, M. R. & Green, M. Functional domains of adenovirus type 5 E1a proteins. *Cell* **50**, 1091-1100, doi:[https://doi.org/10.1016/0092-8674\(87\)90175-9](https://doi.org/10.1016/0092-8674(87)90175-9) (1987).
- 60 Berk, A. J. Recent lessons in gene expression, cell cycle control, and cell biology from adenovirus. *Oncogene* **24**, 7673-7685, doi:10.1038/sj.onc.1209040 (2005).
- 61 Fattaey, A. R., Harlow, E. & Helin, K. Independent regions of adenovirus E1A are required for binding to and dissociation of E2F-protein complexes. *Mol Cell Biol* **13**, 7267-7277, doi:10.1128/mcb.13.12.7267-7277.1993 (1993).
- 62 Haynes, C. *et al.* Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. *PLoS Comput Biol* **2**, e100, doi:10.1371/journal.pcbi.0020100 (2006).

- 63 Singh, G. P., Ganapathi, M. & Dash, D. Role of intrinsic disorder in transient interactions of hub proteins. *Proteins* **66**, 761-765, doi:10.1002/prot.21281 (2007).
- 64 Dunker, A. K., Cortese, M. S., Romero, P., Iakoucheva, L. M. & Uversky, V. N. Flexible nets. *The FEBS Journal* **272**, 5129-5148, doi:<https://doi.org/10.1111/j.1742-4658.2005.04948.x> (2005).
- 65 Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradović, Z. & Dunker, A. K. Intrinsic Disorder in Cell-signaling and Cancer-associated Proteins. *Journal of Molecular Biology* **323**, 573-584, doi:[https://doi.org/10.1016/S0022-2836\(02\)00969-5](https://doi.org/10.1016/S0022-2836(02)00969-5) (2002).
- 66 Mohan, A. *et al.* Analysis of molecular recognition features (MoRFs). *J Mol Biol* **362**, 1043-1059, doi:10.1016/j.jmb.2006.07.087 (2006).
- 67 Berk, A. J. Functions of adenovirus E1A. *Cancer Surv* **5**, 367-387 (1986).
- 68 Schaack, J. *et al.* E1A and E1B proteins inhibit inflammation induced by adenovirus. *Proc Natl Acad Sci U S A* **101**, 3124-3129, doi:10.1073/pnas.0303709101 (2004).
- 69 Yousef, A. F., Fonseca, G. J., Cohen, M. J. & Mymryk, J. S. The C-terminal region of E1A: a molecular tool for cellular cartography. *Biochemistry and Cell Biology* **90**, 153-163, doi:10.1139/o11-080 %M 22292450 (2012).
- 70 O'Connor, M. J., Zimmermann, H., Nielsen, S., Bernard, H. U. & Kouzarides, T. Characterization of an E1A-CBP interaction defines a novel transcriptional adapter motif (TRAM) in CBP/p300. *J Virol* **73**, 3574-3581, doi:10.1128/jvi.73.5.3574-3581.1999 (1999).
- 71 Ferrari, R. *et al.* 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**, 663-676, doi:10.1016/j.chom.2014.10.004 (2014).
- 72 Liu, Z. *et al.* Prognostic role of upregulated P300 expression in human cancers: A clinical study of synovial sarcoma and a meta-analysis. *Exp Ther Med* **18**, 3161-3171, doi:10.3892/etm.2019.7906 (2019).
- 73 Zhao, L. J., Loewenstein, P. M. & Green, M. Adenovirus E1A TRRAP-targeting domain-mediated enhancement of MYC association with the NuA4 complex activates a panel of MYC target genes enriched for gene expression and ribosome biogenesis. *Virology* **512**, 172-179, doi:10.1016/j.virol.2017.08.010 (2017).
- 74 Tworowski, K. A. *et al.* Adenovirus E1A targets p400 to induce the cellular oncoprotein Myc. *Proc Natl Acad Sci U S A* **105**, 6103-6108, doi:10.1073/pnas.0802095105 (2008).
- 75 Ikura, T. *et al.* Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis. *Cell* **102**, 463-473, doi:10.1016/S0092-8674(00)00051-9 (2000).
- 76 Reid, J. L., Bannister, A. J., Zegerman, P., Martínez-Balbás, M. A. & Kouzarides, T. E1A directly binds and regulates the P/CAF acetyltransferase. *Embo j* **17**, 4469-4477, doi:10.1093/emboj/17.15.4469 (1998).

- 77 Flak, M. B. *et al.* p21 promotes oncolytic adenoviral activity in ovarian cancer and is a potential biomarker. *Molecular Cancer* **9**, 175, doi:10.1186/1476-4598-9-175 (2010).
- 78 Alevizopoulos, K., Catarin, B., Vlach, J. & Amati, B. A novel function of adenovirus E1A is required to overcome growth arrest by the CDK2 inhibitor p27(Kip1). *Embo j* **17**, 5987-5997, doi:10.1093/emboj/17.20.5987 (1998).
- 79 Turnell, A. S. *et al.* Regulation of the 26S proteasome by adenovirus E1A. *Embo j* **19**, 4759-4773, doi:10.1093/emboj/19.17.4759 (2000).
- 80 Somasundaram, K. *et al.* Repression of a matrix metalloprotease gene by E1A correlates with its ability to bind to cell type-specific transcription factor AP-2. *Proc Natl Acad Sci U S A* **93**, 3088-3093, doi:10.1073/pnas.93.7.3088 (1996).
- 81 Taylor, D. A., Kraus, V. B., Schwarz, J. J., Olson, E. N. & Kraus, W. E. E1A-mediated inhibition of myogenesis correlates with a direct physical interaction of E1A12S and basic helix-loop-helix proteins. *Mol Cell Biol* **13**, 4714-4727, doi:10.1128/mcb.13.8.4714-4727.1993 (1993).
- 82 Meng, X. *et al.* Cellular context of coregulator and adaptor proteins regulates human adenovirus 5 early region 1A-dependent gene activation by the thyroid hormone receptor. *Mol Endocrinol* **17**, 1095-1105, doi:10.1210/me.2002-0294 (2003).
- 83 Song, C. Z., Loewenstein, P. M., Toth, K. & Green, M. Transcription factor TFIID is a direct functional target of the adenovirus E1A transcription-repression domain. *Proc Natl Acad Sci U S A* **92**, 10330-10333, doi:10.1073/pnas.92.22.10330 (1995).
- 84 Howe, J. A., Mymryk, J. S., Egan, C., Branton, P. E. & Bayley, S. T. Retinoblastoma growth suppressor and a 300-kDa protein appear to regulate cellular DNA synthesis. *Proc Natl Acad Sci U S A* **87**, 5883-5887, doi:10.1073/pnas.87.15.5883 (1990).
- 85 Cobrinik, D. Pocket proteins and cell cycle control. *Oncogene* **24**, 2796-2809, doi:10.1038/sj.onc.1208619 (2005).
- 86 Ghosh, M. K. & Harter, M. L. A viral mechanism for remodeling chromatin structure in G0 cells. *Mol Cell* **12**, 255-260, doi:10.1016/s1097-2765(03)00225-9 (2003).
- 87 Liu, X. & Marmorstein, R. Structure of the retinoblastoma protein bound to adenovirus E1A reveals the molecular basis for viral oncoprotein inactivation of a tumor suppressor. *Genes Dev* **21**, 2711-2716, doi:10.1101/gad.1590607 (2007).
- 88 Zhang, H. S. & Dean, D. C. Rb-mediated chromatin structure regulation and transcriptional repression. *Oncogene* **20**, 3134-3138, doi:10.1038/sj.onc.1204338 (2001).
- 89 Lowe, S. W. & Ruley, H. E. Stabilization of the p53 tumor suppressor is induced by adenovirus 5 E1A and accompanies apoptosis. *Genes Dev* **7**, 535-545, doi:10.1101/gad.7.4.535 (1993).
- 90 Avvakumov, N., Kajon, A. E., Hoebe, R. C. & Mymryk, J. S. Comprehensive sequence analysis of the E1A proteins of human and simian adenoviruses. *Virology* **329**, 477-492, doi:10.1016/j.virol.2004.08.007 (2004).

- 91 Zhang, A. *et al.* The Transcriptional Repressor BS69 is a Conserved Target of the E1A Proteins from Several Human Adenovirus Species. *Viruses* **10**, 662, doi:<https://doi.org/10.3390/v10120662> (2018).
- 92 Yousef, A. F. *et al.* Identification of a molecular recognition feature in the E1A oncoprotein that binds the SUMO conjugase UBC9 and likely interferes with polySUMOylation. *Oncogene* **29**, 4693-4704, doi:10.1038/onc.2010.226 (2010).
- 93 Zhang, X. *et al.* The targeting of the proteasomal regulatory subunit S2 by adenovirus E1A causes inhibition of proteasomal activity and increased p53 expression. *J Biol Chem* **279**, 25122-25133, doi:10.1074/jbc.M403287200 (2004).
- 94 Zhao, L. Y. & Liao, D. Sequestration of p53 in the cytoplasm by adenovirus type 12 E1B 55-kilodalton oncoprotein is required for inhibition of p53-mediated apoptosis. *J Virol* **77**, 13171-13181, doi:10.1128/jvi.77.24.13171-13181.2003 (2003).
- 95 Bayley, S. T. & Mymryk, J. S. Adenovirus e1a proteins and transformation (review). *Int J Oncol* **5**, 425-444, doi:10.3892/ijo.5.3.425 (1994).
- 96 Jelsma, T. N. *et al.* Use of deletion and point mutants spanning the coding region of the adenovirus 5 E1A gene to define a domain that is essential for transcriptional activation. *Virology* **163**, 494-502, doi:10.1016/0042-6822(88)90290-5 (1988).
- 97 Pelka, P., Ablack, J. N., Fonseca, G. J., Yousef, A. F. & Mymryk, J. S. Intrinsic structural disorder in adenovirus E1A: a viral molecular hub linking multiple diverse processes. *J Virol* **82**, 7252-7263, doi:10.1128/jvi.00104-08 (2008).
- 98 Vijayalingam, S. & Chinnadurai, G. Adenovirus L-E1A activates transcription through mediator complex-dependent recruitment of the super elongation complex. *J Virol* **87**, 3425-3434, doi:10.1128/jvi.03046-12 (2013).
- 99 Liu, F. & Green, M. R. Promoter targeting by adenovirus E1a through interaction with different cellular DNA-binding domains. *Nature* **368**, 520-525, doi:10.1038/368520a0 (1994).
- 100 Rasti, M. *et al.* Roles for APIS and the 20S proteasome in adenovirus E1A-dependent transcription. *Embo j* **25**, 2710-2722, doi:10.1038/sj.emboj.7601169 (2006).
- 101 Cohen, M. J., King, C. R., Dikeakos, J. D. & Mymryk, J. S. Functional analysis of the C-terminal region of human adenovirus E1A reveals a misidentified nuclear localization signal. *Virology* **468-470**, 238-243, doi:<https://doi.org/10.1016/j.virol.2014.08.014> (2014).
- 102 Douglas, J. L. & Quinlan, M. P. Efficient nuclear localization and immortalizing ability, two functions dependent on the adenovirus type 5 (Ad5) E1A second exon, are necessary for cotransformation with Ad5 E1B but not with T24ras. *J Virol* **69**, 8061-8065, doi:10.1128/jvi.69.12.8061-8065.1995 (1995).
- 103 Subramanian, T., La Regina, M. & Chinnadurai, G. Enhanced ras oncogene mediated cell transformation and tumorigenesis by adenovirus 2 mutants lacking the C-terminal region of E1a protein. *Oncogene* **4**, 415-420 (1989).
- 104 Zhao, L. J., Subramanian, T., Vijayalingam, S. & Chinnadurai, G. PLDLS-dependent interaction of E1A with CtBP: regulation of CtBP nuclear localization and

- transcriptional functions. *Oncogene* **26**, 7544-7551, doi:10.1038/sj.onc.1210569 (2007).
- 105 Frisch, S. M. & Francis, H. Disruption of epithelial cell-matrix interactions induces apoptosis. *J Cell Biol* **124**, 619-626, doi:10.1083/jcb.124.4.619 (1994).
- 106 Chinnadurai, G. Transcriptional regulation by C-terminal binding proteins. *Int J Biochem Cell Biol* **39**, 1593-1607, doi:10.1016/j.biocel.2007.01.025 (2007).
- 107 Subramanian, T., Zhao, L. J. & Chinnadurai, G. Interaction of CtBP with adenovirus E1A suppresses immortalization of primary epithelial cells and enhances virus replication during productive infection. *Virology* **443**, 313-320, doi:10.1016/j.virol.2013.05.018 (2013).
- 108 Glenewinkel, F. *et al.* The adaptor protein DCAF7 mediates the interaction of the adenovirus E1A oncoprotein with the protein kinases DYRK1A and HIPK2. *Sci Rep* **6**, 28241, doi:10.1038/srep28241 (2016).
- 109 Becker, W. Emerging role of DYRK family protein kinases as regulators of protein stability in cell cycle control. *Cell Cycle* **11**, 3389-3394, doi:10.4161/cc.21404 (2012).
- 110 Doronin, K. *et al.* Tumor-Specific, Replication-Competent Adenovirus Vectors Overexpressing the Adenovirus Death Protein. *Journal of Virology* **74**, 6147-6155, doi:doi:10.1128/JVI.74.13.6147-6155.2000 (2000).
- 111 Zoller, M. J. & Smith, M. Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template. *DNA* **3**, 479-488, doi:10.1089/dna.1.1984.3.479 (1984).
- 112 Kalderon, D., Oostra, B. A., Ely, B. K. & Smith, A. E. Deletion loop mutagenesis: a novel method for the construction of point mutations using deletion mutants. *Nucleic Acids Res* **10**, 5161-5171, doi:10.1093/nar/10.17.5161 (1982).
- 113 Mymryk, J. S. & Bayley, S. T. Multiple pathways for gene activation in rodent cells by the smaller adenovirus 5 E1A protein and their relevance to growth and transformation. *Journal of General Virology* **74**, 2131-2141, doi:<https://doi.org/10.1099/0022-1317-74-10-2131> (1993).
- 114 Krantz, C. K., Routes, B. A., Quinlan, M. P. & Cook, J. L. E1A second exon requirements for induction of target cell susceptibility to lysis by natural killer cells: implications for the mechanism of action. *Virology* **217**, 23-32, doi:10.1006/viro.1996.0089 (1996).
- 115 Boyd, J. M. *et al.* A region in the C-terminus of adenovirus 2/5 E1a protein is required for association with a cellular phosphoprotein and important for the negative modulation of T24-ras mediated transformation, tumorigenesis and metastasis. *Embo j* **12**, 469-478, doi:10.1002/j.1460-2075.1993.tb05679.x (1993).
- 116 Crisostomo, L. *et al.* The Influence of E1A C-Terminus on Adenovirus Replicative Cycle. *Viruses* **9**, 387 (2017).
- 117 Mymryk, J. S. *et al.* Disruption of the coordinate expression of muscle genes in a transfected BC3H1 myoblast cell line producing a low level of the adenovirus E1A transforming protein. *Biochem Cell Biol* **70**, 1268-1276, doi:10.1139/o92-173 (1992).

- 118 Levine, M. & Davidson, E. H. Gene regulatory networks for development. *Proc Natl Acad Sci U S A* **102**, 4936-4942, doi:10.1073/pnas.0408031102 (2005).
- 119 Saint-André, V. Computational biology approaches for mapping transcriptional regulatory networks. *Comput Struct Biotechnol J* **19**, 4884-4895, doi:10.1016/j.csbj.2021.08.028 (2021).
- 120 Blais, A. & Dynlacht, B. D. Constructing transcriptional regulatory networks. *Genes Dev* **19**, 1499-1511, doi:10.1101/gad.1325605 (2005).
- 121 Busiello, D. M., Suweis, S., Hidalgo, J. & Maritan, A. Explorability and the origin of network sparsity in living systems. *Scientific Reports* **7**, 12323, doi:10.1038/s41598-017-12521-1 (2017).
- 122 Babu, M. M., Luscombe, N. M., Aravind, L., Gerstein, M. & Teichmann, S. A. Structure and evolution of transcriptional regulatory networks. *Current Opinion in Structural Biology* **14**, 283-291, doi:<https://doi.org/10.1016/j.sbi.2004.05.004> (2004).
- 123 Yu, H., Luscombe, N. M., Qian, J. & Gerstein, M. Genomic analysis of gene expression relationships in transcriptional regulatory networks. *Trends in Genetics* **19**, 422-427, doi:[https://doi.org/10.1016/S0168-9525\(03\)00175-6](https://doi.org/10.1016/S0168-9525(03)00175-6) (2003).
- 124 Guelzim, N., Bottani, S., Bourguin, P. & Képès, F. Topological and causal structure of the yeast transcriptional regulatory network. *Nature Genetics* **31**, 60-63, doi:10.1038/ng873 (2002).
- 125 Tavazoie, S., Hughes, J. D., Campbell, M. J., Cho, R. J. & Church, G. M. Systematic determination of genetic network architecture. *Nature Genetics* **22**, 281-285, doi:10.1038/10343 (1999).
- 126 Gutiérrez-Ríos, R. M. *et al.* Regulatory network of Escherichia coli: Consistency between literature knowledge and microarray profiles. *Genome Research* **13**, 2435-2443, doi:10.1101/gr.1387003 (2003).
- 127 He, B. & Tan, K. Understanding transcriptional regulatory networks using computational models. *Curr Opin Genet Dev* **37**, 101-108, doi:10.1016/j.gde.2016.02.002 (2016).
- 128 Ament, S. A. *et al.* Transcriptional regulatory networks underlying gene expression changes in Huntington's disease. *Molecular Systems Biology* **14**, e7435, doi:<https://doi.org/10.15252/msb.20167435> (2018).
- 129 Emmert-Streib, F., Dehmer, M. & Haibe-Kains, B. Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks. *Frontiers in Cell and Developmental Biology* **2**, doi:10.3389/fcell.2014.00038 (2014).
- 130 Margolin, A. A. *et al.* Reverse engineering cellular networks. *Nature Protocols* **1**, 662-671, doi:10.1038/nprot.2006.106 (2006).
- 131 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).
- 132 Segal, E. *et al.* Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nature Genetics* **34**, 166-176, doi:10.1038/ng1165 (2003).

- 133 Sun, N., Carroll, R. J. & Zhao, H. Bayesian error analysis model for reconstructing transcriptional regulatory networks. *Proceedings of the National Academy of Sciences* **103**, 7988-7993, doi:doi:10.1073/pnas.0600164103 (2006).
- 134 Gardner, T. S., di Bernardo, D., Lorenz, D. & Collins, J. J. Inferring Genetic Networks and Identifying Compound Mode of Action via Expression Profiling. *Science* **301**, 102-105, doi:doi:10.1126/science.1081900 (2003).
- 135 Thieffry, D. & Thomas, R. Qualitative analysis of gene networks. *Pac Symp Biocomput*, 77-88 (1998).
- 136 Ren, B. *et al.* Genome-Wide Location and Function of DNA Binding Proteins. *Science* **290**, 2306-2309, doi:10.1126/science.290.5500.2306 (2000).
- 137 Jones, N. & Shenk, T. Isolation of adenovirus type 5 host range deletion mutants defective for transformation of rat embryo cells. *Cell* **17**, 683-689, doi:[https://doi.org/10.1016/0092-8674\(79\)90275-7](https://doi.org/10.1016/0092-8674(79)90275-7) (1979).
- 138 Hibma, M. H. *et al.* Increased apoptosis and reduced replication efficiency of the E3 region-modified dl309 adenovirus in cancer cells. *Virus Research* **145**, 112-120, doi:<https://doi.org/10.1016/j.virusres.2009.06.016> (2009).
- 139 Alexa Bieler, A. B., Sybille Staerk, Anka Grosso, Bernd Gaensbacher, Per Sonne Holm. 612. Effect of Adenovirus dl520 as Replicating Agent in Combination with Fractionated Radiotherapy in a Glioma Xenograft Model. *Molecular Therapy* **9**, S232, doi:10.1016/j.ymthe.2004.06.537 (2004).
- 140 Malladi, A. & Quinlan, M. P. Mutations in CR1 of E1A 12S yield dominant negative suppressors of immortalization and the lytic cycle. *Virology* **233**, 51-62, doi:10.1006/viro.1997.8605 (1997).
- 141 Krippl, B., Ferguson, B., Rosenberg, M. & Westphal, H. Functions of purified E1A protein microinjected into mammalian cells. *Proc Natl Acad Sci U S A* **81**, 6988-6992, doi:10.1073/pnas.81.22.6988 (1984).
- 142 Phelps, W. C., Yee, C. L., Munger, K. & Howley, P. M. The human papillomavirus type 16 E7 gene encodes transactivation and transformation functions similar to those of adenovirus E1A. *Cell* **53**, 539-547, doi:10.1016/0092-8674(88)90570-3 (1988).
- 143 Ohlenschlager, O. *et al.* Solution structure of the partially folded high-risk human papilloma virus 45 oncoprotein E7. *Oncogene* **25**, 5953-5959, doi:10.1038/sj.onc.1209584 (2006).
- 144 Snitkin, E. S. *et al.* Model-driven analysis of experimentally determined growth phenotypes for 465 yeast gene deletion mutants under 16 different conditions. *Genome Biology* **9**, R140, doi:10.1186/gb-2008-9-9-r140 (2008).
- 145 Singhal, G., Leo, E., Setty, S. K., Pommier, Y. & Thimmapaya, B. Adenovirus E1A oncogene induces rereplication of cellular DNA and alters DNA replication dynamics. *J Virol* **87**, 8767-8778, doi:10.1128/jvi.00879-13 (2013).
- 146 Frisch, S. M. Antioncogenic effect of adenovirus E1A in human tumor cells. *Proc Natl Acad Sci U S A* **88**, 9077-9081, doi:10.1073/pnas.88.20.9077 (1991).
- 147 Frisch, S. M. E1a induces the expression of epithelial characteristics. *J Cell Biol* **127**, 1085-1096, doi:10.1083/jcb.127.4.1085 (1994).

- 148 Sun, S., Rao, V. B. & Rossmann, M. G. Genome packaging in viruses. *Curr Opin Struct Biol* **20**, 114-120, doi:10.1016/j.sbi.2009.12.006 (2010).
- 149 Maugeri, A., Barchitta, M., Basile, G. & Agodi, A. Applying a hierarchical clustering on principal components approach to identify different patterns of the SARS-CoV-2 epidemic across Italian regions. *Scientific Reports* **11**, 7082, doi:10.1038/s41598-021-86703-3 (2021).
- 150 Sano, M., Kato, Y. & Taira, K. Sequence-specific interference by small RNAs derived from adenovirus VAI RNA. *FEBS Letters* **580**, 1553-1564, doi:<https://doi.org/10.1016/j.febslet.2006.01.085> (2006).
- 151 Aparicio, O., Razquin, N., Zaratiegui, M., Narvaiza, I. & Fortes, P. Adenovirus Virus-Associated RNA Is Processed to Functional Interfering RNAs Involved in Virus Production. *Journal of Virology* **80**, 1376-1384, doi:doi:10.1128/JVI.80.3.1376-1384.2006 (2006).
- 152 Strijker, R., Fritz, D. T. & Levinson, A. D. Adenovirus VAI-RNA regulates gene expression by controlling stability of ribosome-bound RNAs. *Embo j* **8**, 2669-2675, doi:10.1002/j.1460-2075.1989.tb08407.x (1989).
- 153 Stevenson, B. W. *et al.* A structural view of PA2G4 isoforms with opposing functions in cancer. *Journal of Biological Chemistry* **295**, 16100-16112, doi:<https://doi.org/10.1074/jbc.REV120.014293> (2020).
- 154 Zhao, H., Granberg, F. & Pettersson, U. How adenovirus strives to control cellular gene expression. *Virology* **363**, 357-375, doi:<https://doi.org/10.1016/j.virol.2007.02.013> (2007).
- 155 Ohira, T. *et al.* PITX1 protein interacts with ZCCHC10 to regulate hTERT mRNA transcription. *PLoS One* **14**, e0217605, doi:10.1371/journal.pone.0217605 (2019).
- 156 Liu, D. X. & Lobie, P. E. Transcriptional activation of p53 by Pitx1. *Cell Death & Differentiation* **14**, 1893-1907, doi:10.1038/sj.cdd.4402209 (2007).
- 157 Wang, B. *et al.* The Role of the Transcription Factor EGR1 in Cancer. *Frontiers in Oncology* **11**, doi:10.3389/fonc.2021.642547 (2021).
- 158 Gregg, J. & Fraizer, G. Transcriptional Regulation of EGR1 by EGF and the ERK Signaling Pathway in Prostate Cancer Cells. *Genes Cancer* **2**, 900-909, doi:10.1177/1947601911431885 (2011).
- 159 Johnson, A. E., Le, I. P., Buchwalter, A. & Burnatowska-Hledin, M. A. Estrogen-dependent growth and estrogen receptor (ER)- α concentration in T47D breast cancer cells are inhibited by VACM-1, a cul 5 gene. *Molecular and Cellular Biochemistry* **301**, 13-20, doi:10.1007/s11010-006-9392-3 (2007).
- 160 Kubosaki, A. *et al.* Genome-wide investigation of in vivo EGR-1 binding sites in monocytic differentiation. *Genome Biol* **10**, R41, doi:10.1186/gb-2009-10-4-r41 (2009).
- 161 Shashkova, E., Spencer, J., Wold, W. & Doronin, K. Shashkova, EV, Spencer, JF, Wold, WS and Doronin, K. Targeting Interferon-alpha Increases Antitumor Efficacy and Reduces Hepatotoxicity of E1A-mutated Spread-enhanced Oncolytic Adenovirus. *Mol Ther* **15**: 598-607. *Molecular therapy : the journal of the*

- American Society of Gene Therapy* **15**, 598-607, doi:10.1038/sj.mt.6300064 (2007).
- 162 Liu, N. *et al.* CUX1, A Controversial Player in Tumor Development. *Front Oncol* **10**, 738, doi:10.3389/fonc.2020.00738 (2020).
- 163 Wang, Y., Çil, Ç., Harnett, M. M. & Pineda, M. A. Cytohesin-2/ARNO: A Novel Bridge Between Cell Migration and Immunoregulation in Synovial Fibroblasts. *Frontiers in Immunology* **12**, doi:10.3389/fimmu.2021.809896 (2022).
- 164 Patel, N. & Bush, W. S. Modeling transcriptional regulation using gene regulatory networks based on multi-omics data sources. *BMC Bioinformatics* **22**, 200, doi:10.1186/s12859-021-04126-3 (2021).
- 165 Boeva, V. Analysis of Genomic Sequence Motifs for Deciphering Transcription Factor Binding and Transcriptional Regulation in Eukaryotic Cells. *Frontiers in Genetics* **7**, doi:10.3389/fgene.2016.00024 (2016).
- 166 Müller, G. A. *et al.* The CHR site: definition and genome-wide identification of a cell cycle transcriptional element. *Nucleic Acids Res* **42**, 10331-10350, doi:10.1093/nar/gku696 (2014).
- 167 Sun, Y., Miao, N. & Sun, T. Detect accessible chromatin using ATAC-sequencing, from principle to applications. *Hereditas* **156**, 29, doi:10.1186/s41065-019-0105-9 (2019).
- 168 Sonawane, A. R., DeMeo, D. L., Quackenbush, J. & Glass, K. Constructing gene regulatory networks using epigenetic data. *npj Systems Biology and Applications* **7**, 45, doi:10.1038/s41540-021-00208-3 (2021).