

Transcriptomics of E1A and Arglu1 in the Cell Cycle, Polymerase Pausing, and DNA Damage Response

by
Rafe Helwer

A Thesis Submitted to the Faculty of Graduate Studies of The University of Manitoba in
partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Department of Microbiology
University of Manitoba
Winnipeg, Manitoba, Canada

Copyright © 2024 Rafe Helwer

Abstract	3
Acknowledgments	5
Permissions	7
List of Figures	8
List of Tables	9
 1 Introduction	 10
1.1 Adenovirus	
1.2 E1A	
1.3 Arglu1	
1.4 Arglu1 Structure and Splicing	
1.5 Arglu1 in DNA Damage, Cell Cycle, Growth, and Cancer	
1.6 Objectives	
 2 Methods	 28
2.1 RNA-Seq	
2.2 Cell Cycle Analysis	
2.3 Actinomycin D Drug Focus Forming Assay	
2.4 Proteomics	
2.5 Deletion Mutant Plasmids	
2.6 Stable Cell Lines	
2.7 ChIP-Seq	
 3 Results	 38
3.1 Adenovirus E1A RNA-Seq	
3.2 Arglu1 Phenotype	
3.3 Arglu1 Proteomics	
3.4 Arglu1 Deletion Mutant Plasmids and Stable Cell Lines	
3.5 Arglu1 ChIP-Seq	
 4 Discussion	 93
4.1 Adenovirus E1A Isoforms and YCD Mutant	
4.2 Arglu1 Cell Cycle	
4.3 Arglu1 Transcription and Pausing	
4.4 Arglu1 DNA Damage	
4.5 Limitations and Looking Ahead	
 References	 107
 Appendix	 120

Abstract

The Adenovirus E1A protein transforms human cells in by disrupting the binding of the retinoblastoma protein (pRb or Rb) to the transcription factor E2F. When not repressed by pRb, E2F-dependent transcription activates S phase or replication initiation in the cell cycle. Although there have been previous studies on the differences between E1A 13S and 12S, the two largest and earliest expressed isoforms, their global effect on human gene expression is less understood. Previously my lab made mutants of E1A 13S (YCD) and 12S (12S YCD) that deleted residues 2-11 and mutates Y47 and C124, resulting in a loss of binding to pRb and E2F, deficient S phase activation, and reduced viral growth. However, unlike E1A 12S YCD, the 13S YCD mutant was able to drive S phase in a BrdU assay, while 12S had a reduced amount of S phase initiation compared to WT and 13S. BrdU is a thymidine analogue that is incorporated into replicating DNA and can be detected with anti-BrdU antibodies. To generate hypothesis for how 13S YCD could do this, I analyzed RNA-sequencing of primary human lung fibroblast cells (IMR-90s) infected with wild type adenovirus and three adenovirus mutants, the first expressing only the E1A 13S isoform, the second expressing predominantly E1A 12S, and the E1A 13S YCD mutant. I subsequently performed differential expression analysis, principal component analysis, clustering, and gene set enrichment analysis on the data. I found that in both the individual isoforms and the YCD mutant, transcripts associated with DNA replication and the cell cycle are downregulated. Additionally, the E1A 13S isoform has increased expression of heat shock protein transcripts, cytokines, and transcripts associated with vesicles. While the E1A 12S isoform has reduced signaling and heat shock protein transcripts but an

increase in transcripts associated with Golgi bodies. Even more than E1A 13S, the YCD mutant has a substantial increase in cytokine, inflammatory, and immune activating transcripts. In particular, I identified 28 cell cycle-associated upregulated transcripts in YCD/12S that could explain how YCD is able to drive S phase, including MYC, Cyclins D and J, DDX3X, and PPME1.

Arglu1 is a human protein that increases the cellular growth rate, the amount of paused RNA polymerase II, and is localized to sites of DNA damage. I further characterized Arglu1 by showing that Arglu1 overexpressing cells (HTA) increase resistance to the DNA damage inducing genotoxic drug Actinomycin D and have reduced duration of G1 and S-phase but slightly extended G2/M phase. To analyze how Arglu1 increased the cell cycle rate and resistance to DNA damage, I made deletion mutants of Arglu1, stable cell lines overexpressing these mutants, and performed proteomic and ChIP-seq analysis of HTA cells. Proteomics showed that HTA cells are enriched for GO terms including the cell cycle, the G1/S transition, DNA damage response, and transcription elongation. ChIP-seq data of RNAP2 and Arglu1 in HTA cells showed that 10690 transcripts were paused in HTA cells, 2470 genes were both paused and occupied by Arglu1, and 82 genes were paused, occupied by Arglu1, and differentially expressed as proteins in HTA cells. I found genes involved in DNA damage response, the cell cycle, and transcription that generate new hypotheses for how Arglu1 affects these pathways. This research should increase our understanding of cancer and tumor viruses and possibly lead to new oncogenes and drug targets.

Acknowledgements

I would like to thank my supervisor Dr. Peter Pelka, and the rest of his lab, in particular Scott Bachus, Nik Akkerman, Drayson Graves, and Lauren Fulham but also Jordan Rempel, Fazia Ait Zenati, Nathaniel Casiano, Amit Koul, Ian McNicol, Esha Kaul, Wenjun Zhu, and Anusha Memon. I would also like to thank the other members of my committee Dr. Karen Brassinga and Dr. Georg Hausner.

I have been financially supported by a Master's Scholarship from Research Manitoba and Travel Scholarships from UMGSA and the Faculties of Science and Graduate Studies. The lab has also been funded by the Natural Sciences and Engineering Research Council of Canada, the Canadian Institutes of Health Research, and the Canada Biomedical Research Fund.

I would like to thank Dr. Marike Palmer for invaluable assistance with ChIP-Seq programs and methodology, Dr. Aleeza Gerstein for instruction in R and biostatistics, Dr Christine Zhang for technical assistance with flow cytometry, and the rest of the Department of Microbiology. I would also like to thank Dr. Michael Charette and Dr. Vincent Chen for supervising my lab work during my undergraduate degree.

I would like to thank my parents, my friends, and my two cats.

Finally, I thank Thomas Pynchon, Mark Z. Danielewski, Roberto Bolaño, Emily Dickenson, Wallace Stevens, Autechre, Apichatpong Weerasethakul, Edward Yang, and Tsai Ming-liang for aesthetic support.

Permissions

Portions of this thesis were previously published as Bachus et al. (2024) by Oxford University Press on behalf of Nucleic Acids Research as an Open Access article distributed under the terms of the Creative Commons Attribution-Non Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which “permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited.”

Figure 1 is originally from (King et al. 2018) published by American Society of Microbiology (ASM), copyright © 2018 King et al. and is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license (<https://creativecommons.org/licenses/by/4.0/>), which allows sharing (“[copying and redistributing] the material in any medium or format for any purpose, even commercially”) under the condition of attribution or credit or the creators (“You must give appropriate credit , provide a link to the license, and indicate if changes were made . You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.”). No changes were made to any of the figures. Figure 2a is originally from (González-Foutel et al. 2022) (Figure 1a). Reproduced with permission from Springer Nature, License Number: 5933931261012 (via Copyright Clearance Center’s RightsLink® service). Figure 2b is originally from (Matthews et al. 2021) (Box 1 Figure). Reproduced with permission from Springer Nature, License Number: 5933940575811.

List of Figures

E1A RNA-Seq

- Figure 1:** E1A Protein Interactions
- Figure 2:** E1A Cell Cycle Mechanism
- Figure 3:** E1A Isoforms and BrdU Assay
- Figure 4:** RNA-Seq PCA, Scree Plot, Counts
- Figure 5:** RNA-Seq Venn diagrams
- Figure 6:** RNA-Seq Heat plot
- Figure 7:** RNA-Seq Volcano Plots
- Figure 8:** RNA-Seq GO Bar plots

Arglu1

- Figure 9:** Actinomycin D Assay
- Figure 10:** Flow Cytometry
- Figure 11:** Arglu1 Proteomics PCA
- Figure 12:** Arglu1 Proteomics Top Fold Change Heat Map
- Figure 13:** Arglu1 Proteomics Volcano Plot
- Figure 14:** Arglu1 Proteomics GSEA Bar Chart
- Figure 15:** Genome Ontology Comparison
- Figure 16:** ChIP-Seq of ILF3
- Figure 17:** ChIP-Seq Arglu1 GO Dot Plot
- Figure 18:** Pausing Index All Genes vs Arglu1 Occupied Genes
- Figure 19:** Pausing Index GO plot
- Figure 20:** Arglu1 Peaks-RNAP2 Pausing Index Venn and Volcano
- Figure 21:** Arglu1 Proteomics and Arglu1 ChIP-Seq Venn and Volcano

Appendix Figures

- Figure S1:** Sample to Sample Heat Map
- Figure S2:** PCA Top/Bottom Loading Genes
- Figure S3:** RNA-Seq GSEA bar charts
- Figure S4:** PCA 3 sample chart
- Figure S5:** PCA 3 samples Top/Bottom Loading Genes
- Figure S6:** Heat Map 3 Samples
- Figure S7:** RNA-Seq GO Bar plots (13S/UI, 12S/UI, YCD/UI)
- Figure S8:** Heat Map Clusters A
- Figure S9:** Heat Map Clusters B
- Figure S10:** Proteomics Normalized Intensity Distribution and Missing, Values, Sample to Sample Heat Map
- Figure S11:** Arglu1 Proteomics g:Profiler Figure
- Figure S12:** Arglu1 Proteomics GSEA Enrichment Map

List of Tables

Table 1: Deletion Mutant Components Table
Table 2: Software Used
Table 3: RNA-Seq Samples
Table 4: RNA-Seq Sample Comparisons
Table 5: Cell Cycle Up/Downregulated genes in YCD/12S
Table 6: Arglu1 Proteomics top proteins by p value
Table 7: Arglu1 Proteomics GSEA terms of note
Table 8: Arglu1 Proteomics g:Profiler terms of note
Table 9: Arglu1 Proteomics Proteins of Notes
Table 10: ChIP-Seq Sample Peaks/Motifs
Table 11: ChIP-Seq Arglu1 (/IgG) Peaks
Table 12: ChIP-Seq Arglu1 GO CC
Table 13: ChIP-Seq Arglu1 GO MF
Table 14: ChIP-Seq Arglu1 GO BP
Table 15: Pausing Index most paused genes (HTA/HT)
Table 16: Pausing Index GSEA (HTA/HT)
Table 17: Pausing Index GSEA important
Table 18: Arglu1 ChIP-Seq de novo Motif Enrichment
Table 19: Arglu1 Peaks-RNAP2 Pausing Index GO terms
Table 20: Arglu1 Proteomics and Arglu1 ChIP-Seq GO terms

Appendix Tables

Table S1: RNA-Seq DE, top 25 gens by padj (samples 1-5)
Table S2: RNA-Seq DE, top 25 gens by padj (samples 6-10)
Table S3: PCA PC1 Positive and Negative Loadings GO terms
Table S4: PCA PC2 Positive and Negative Loadings GO terms
Table S6: Cluster Table 1
Table S7: Cluster Table 2
Table S8: Full table of positively and negatively regulated cell cycle process genes
Table S9: Arglu1 Deletion Mutants and Cell Lines
Table S10: Arglu1 Proteomics top proteins by Fold Change
Table S11: Arglu1 Proteomics GSEA table
Table S12: Arglu1 Proteomics g:Profiler table
Table S13: ChIP-Seq RNAP2 HT/IgG Peaks
Table S14: ChIP-Seq RNAP2 HTA/IgG Peaks
Table S15: Arglu1 ChIP-Seq Known Motif Enrichment
Table S16: Genome Ontology Comparison
Table S17: Proteomics and Arglu1 ChIP-Seq Overlaps

1 Introduction

1.1 Adenovirus

Adenovirus (AdV) was originally discovered in 1953 as the cause of an acute respiratory illness in children. The name Adenoviridae is derived via the adenoid tissue from which the virus was isolated (Rowe et al. 1953; Hilleman and Werner 1954). Adenovirus has also been an essential tool for molecular biologists through the discovery of splicing (Berget, Moore, and Sharp 1977; Chow et al. 1977), critical cellular hub proteins (King et al. 2018), and as vaccines (Lion and Wold 2021).

Adenovirus Classification

There are six genera of the family Adenoviridae which have been isolated from all classes of vertebrates from fish to mammals (Walker et al. 2020). All known Human Adenovirus (HAdV) species are classified within the genus Mastadenovirus. Mastadenovirus is currently composed of 51 species infecting various mammals, of which 7 are HAdVs, denoted A to G, comprising 56-88 serotypes depending on classification method, which can include capacity for hemagglutination, transformation, and tumour formation as well as sequence similarity (Heim et al. 2003; Dhingra et al. 2019, Hearing 2021).

Human Adenoviruses

HAdV is usually spread by droplets, fomites, or fecal contamination (Dela Cruz et al. 2019). Symptoms of HAdV infections vary based on serotype but the most common is a mild respiratory infection producing flu-like symptoms such as fever and sore throat as well as

conjunctivitis, pneumonia, acute bronchitis, and, rarely, death. HAdV infections are higher risk for immunocompromised patients (Lion and Wold 2021).

Adenovirus Genome

Adenoviridae have a linear dsDNA genome of 26-45 kbp, HAdVs have about 36 kbp. The genome is divided into early (E), intermediate (I) and late (L) transcription units corresponding to when they are transcribed during infection (Davison, Benkő, and Harrach 2003). The early (E) region genes (E1A, E1B, E2, E3 and E4) activate the transcription of other viral genes necessary for infection. The I region (IX, IVa2, L4 intermediate, and E2 late) is transcribed shortly after E1A is produced (Crisostomo et al. 2019). The L region produces five late mRNAs (L1-L5) that code for structural and other proteins involved in capsid production, virion formation, and DNA packaging. The AdV genome also contains two non-protein coding genes that produce the Virus Associated RNA (VA RNA I and II). The functions of the VARNAs include preventing the cell from terminating protein synthesis and interacting with nuclear export proteins and Dicer (Vachon and Conn 2016). Adenoviruses code for six core proteins associated with viral replication (not counting isoforms) and seven involved in capsid formation and stability. The most important capsid proteins are the penton base, hexon, and fiber. The fiber proteins are used for viral attachment and entry.

Adenovirus Early Transcription Units

The early region codes for at least 25 proteins (counting isoforms). Of the early transcriptional units, the Early 1A (E1A) unit is particularly important as it drives cells into S phase (more below). E1B produces 2 proteins, E1B-19K which is a BCL-2 homologue that inhibits apoptosis by interacting with BAK and BAX (Cuconati and White 2002) and E1B-55K which degrades p53 as its primary function, amongst many other activities (Berk 2005). The E2 region codes for three proteins critical for viral genome replication. E2A, aka DNA binding protein (DPB), stabilizes the AdV genome intermediate during replication (Doerfler and Böhm 1995). E2B1 (pre-terminal protein) is a protein primer for viral DNA replication and E2B2 is the viral DNA polymerase that replicates the viral genome. The E3 region codes for 9 proteins involved in mediating and suppressing host infection response and virus release (Rosenfeld et al. 1992). The E4 unit codes for 6 proteins (E4orf1, orf2, orf3, orf4, orf6 and orf6/7) involved in DNA replication, apoptosis, transcription, RNA splicing, nuclear export, transformation, and cell growth (Hearing 2021, Täuber and Dobner 2001).

Adenovirus Life Cycle

The adenovirus life cycle is complex and not fully understood (Hearing 2021). Adenoviridae attach to many different cell surface receptors depending on species and serotype. Receptors HAdVs use include sialic acid, HSPG, MHCI, CD80, DSG2, GD1A, CD46, and CAR (Coxsackie Adenovirus Receptor). Of these the most common are CD46 and CAR. The CAR receptor, an immunoglobulin, is expressed in many cell types including primary epithelial and endothelial cells and in such organs as the heart, liver, lungs, prostate, testis, and intestines (Howitt, Anderson, and Freimuth 2001). After initial attachment, the

penton base interacts with integrins $\alpha\beta 3$ and $\alpha\beta 5$ for viral entry into the cell (Wickham et al. 1993). After viral endocytosis the virion is released into the cytosol, where it is moved to the nucleus through cellular microtubule and dynein complexes (Pied and Wodrich 2019). It is currently hypothesized that the virion binds to microtubules using hexon (Luftig and Weihing 1975). At the nucleus, the virion associates with a nuclear pore complex (NPC) and releases viral DNA into the nucleus (Kremer and Nemerow 2015). After viral DNA replication, virions are assembled within the nucleus and eventually lyse the cell (Charman et al. 2019).

1.2 E1A

E1A is the first viral protein transcribed during infection. E1A mRNA is detectible after only six hours post infection (Crisostomo et al. 2019). E1A has many functions including driving quiescent cells into S phase, activating viral gene expression, and immune suppression. (King et al. 2018). E1A and E1B together cause tumor formation in rat cells (Tessier et al. 2021). There are 5 isoforms of E1A (13S/298R, 12S/243R, 11S/217R, 10S/171R, and 9S/55R) with the first set of values referring to mRNAs (S is the Svedberg unit) and the second to proteins (R for amino acid residues). The 13S and 12S isoforms are detected earliest, peaking at around 24 hours post infection. The three smaller isoforms appear later and the 10S has the highest average abundance over the entire course of infection (Radko et al. 2015). 13S is generally regarded as a transcriptional activator and 12S as a transcriptional repressor (Pelka et al. 2009). There are four regions in E1A that are conserved through many different AdV E1A sequences, the 46 amino acid residue

sequence unique to the 289R isoform is the third conserved region 3 (CR3) (Ablack et al. 2012).

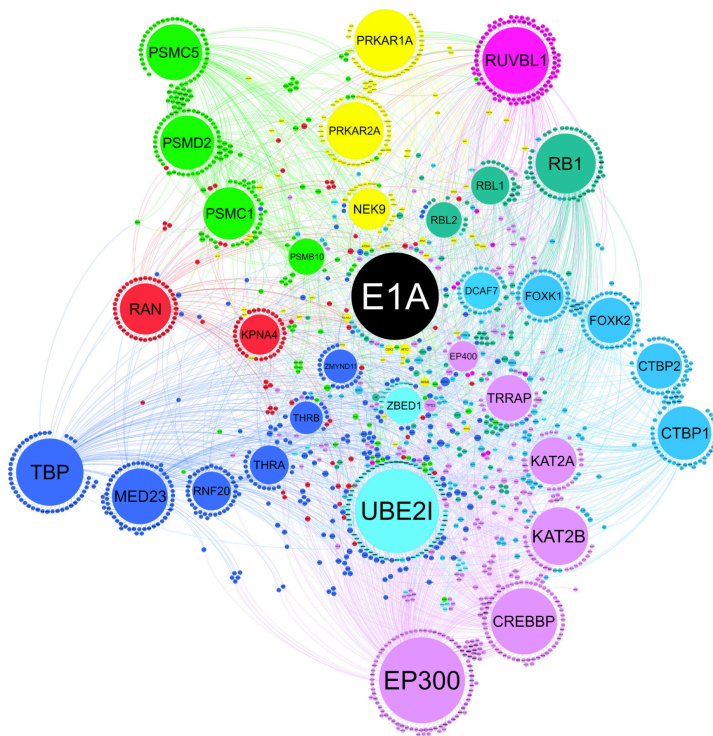
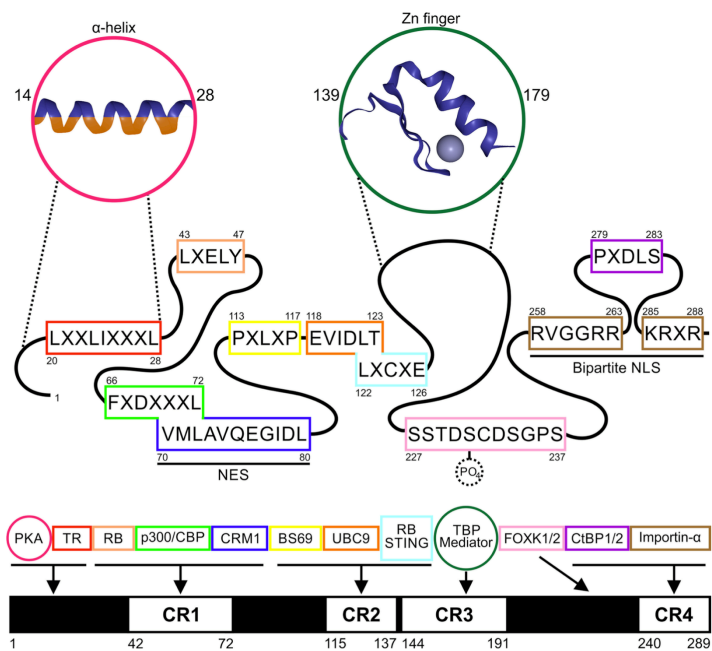


Figure 1: (top) Conserved Regions of E1A and their binding partners. (bottom) The HAdV-5 E1A interactome or protein interaction network. 32 primary (large circles) and 2,207

secondary binding partners are depicted along with the 4,087 unique edges between all depicted proteins [and] ... the size of each is proportional to the number of its own binding partners". Originally from (King et al. 2018)

E1A Conserved Regions

The four conserved regions in E1A (Figure 1) have the least amount of variation across species and are responsible for most of the important functions of E1A (Kimelman 1986).

CR1 spans residues 42-72 in HAdV-C5 E1A and is required for transcriptional repression and transformation (with CR2), DNA synthesis, and viral transduction (Svensson et al.

1991). CR1 interacts with human proteins including p300/CBP, pRb, and CRM1 (King et al. 2018). Of particular importance is E1A blocking the retinoblastoma protein (pRb) from

binding to E2F transcription factors (Figure 2). The E1A-pRb interaction prevents pRb from repressing transcription by binding and inactivating E2F, a vital transcription factor for cell cycle initiation (Ikeda and Nevins 1993). CR2 spans residues 115-137 of HAdV-C5 E1A and

is S-phase initiation (Avvakumov et al. 2004), transforming rodent cells in culture (with CR1), and transcriptional repression (Lillie et al. 1987, Pelka et al. 2008). Some of the

human proteins CR2 interacts with are pRb, STING, BS69, and UBC9 (King et al. 2018). CR3

spans residues 144-191 of HAdV-C5 E1A and contains a zinc finger and promotor targeting region (Culp et al. 1988; Liu and Green 1994). The zinc finger, possibly the only structured part of E1A (Ablack et al. 2012), is used to transactivate other early viral proteins (Berk et al.

1979; Jones and Shenk 1979) and interacts with the mediator complex subunit (MED) 23

(Singh et al. 2019) CR3 interacts with the human proteins hSug1, p300/CBP, and the TATA-box binding protein (Ablack et al. 2010; Arany et al. 1995). CR4 spans residues 240-288

(Avvakumov et al. 2004) at the C-terminus of E1A. CR4 contains a nuclear localization

signal and binds to the human proteins CtBP1 and 2, Ku70, DREF, Importin- α , and HAN11 (Frost et al. 2017; Radko et al. 2014; Zhao et al. 2008; Boyd et al. 1993; Schaeper et al. 1995a).

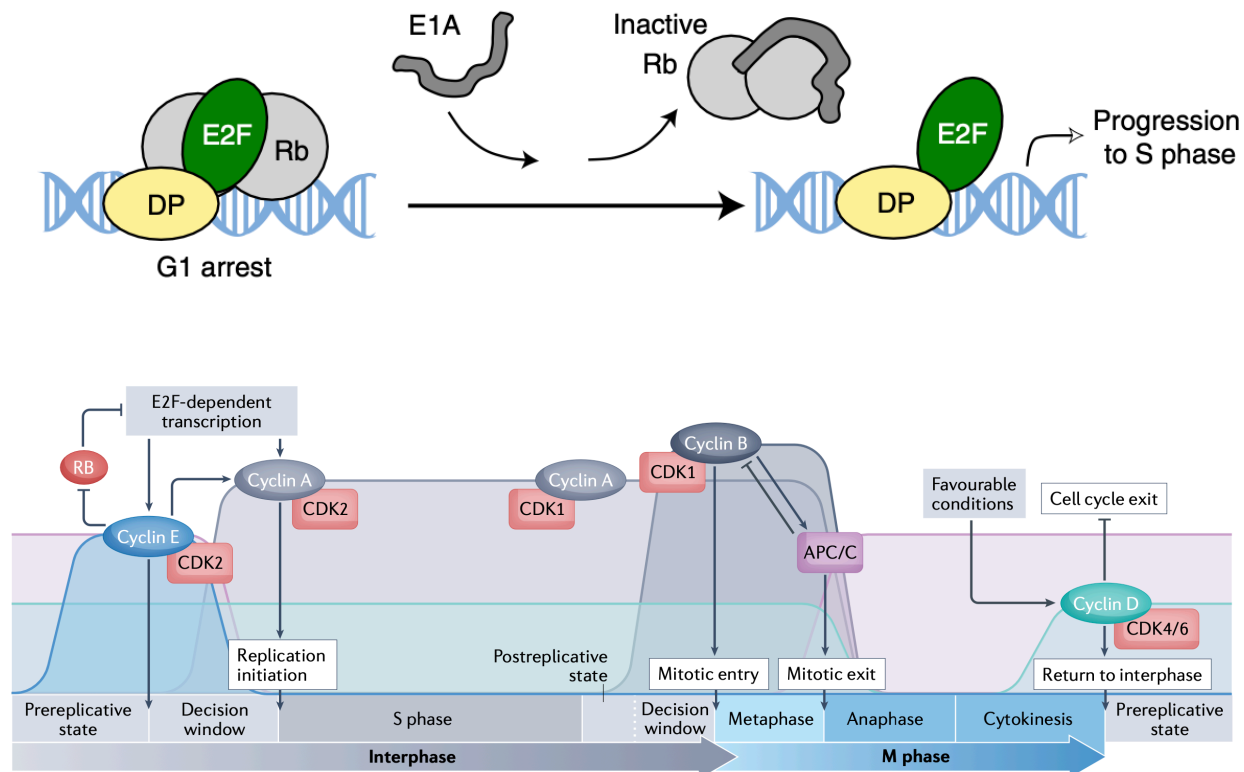


Figure 2: (top) schematic mechanism of E1A binding to the Retinoblastoma (Rb aka pRb) and E2F, causing Rb to be released from E2F and the transcription factor DP (dimerization factor). Rb is subsequently inactivated via phosphorylation (not by E1A) and unbound E2F becomes activates and (bottom left) E2F-dependent transcription activates Cyclins A and E which activates replication initiation and S phase. Figure 2 (top) is originally from (González-Foutel et al. 2022), figure 2 (bottom) is originally from (Matthews et al 2021). Reproduced with permission from Springer Nature.

E1A Structure

The structure of E1A is believed to be almost entirely intrinsically disordered. Interestingly, E1A affects transcription even though it never binds to DNA directly, rather it binds to transcription factor proteins and recruits them to promoters (Avvakumov et al. 2002). So far, E1A has been found to interact with 32 proteins directly and 2,207 proteins indirectly. Because of the small size of E1A and the rapid rate of viral evolution, cellular proteins that E1A interacts with tend to be important hub proteins (Liu and Green 1994). E1A interacts with proteins via molecular recognition features (MoRFs) aka short linear motifs (SLiMs), which are small 5-11 residue sequences that bind to host proteins (Davey et al. 2011; Davey et al. 2015; Yousef et al. 2010; Pelka et al. 2008). MoRFs have low binding affinity for their target sequences and their interactions are thus transient and reversible. For example, the E1A MoRF that interacts with CtBPs (C-terminal binding proteins) is PXDLS, where X varies among HAdV species (Quinlan et al. 2006; Schaeper et al. 1995b). Other viral and cellular proteins have also been found to use the same MoRF to interact with CtBPs (Pelka et al. 2008). E1A also uses 2 MoRFs in CR1 and CR2 to bind pRb and prevent it from binding to E2F (Liu and Marmorstein 2007).

E1A Functions

E1A has a variety of functions. HAdV needs RNA polymerases and transcription factors to transcribe its genes (Hearing 2021) and E1A is a crucial part of this by activating transcription of the other adenovirus early genes. The transactivation ability of E1A is traditionally thought to be in CR1 and CR3 (Yousef et al. 2009). The CR3 region binds several components of the TFIID (Transcription Factor II D) complex, which is involved in

the formation of the RNA Pol II Preinitiation Complex and includes TATA-box binding protein (TBP) and TBP-associated factors (TAFs) (King et al. 2019). The full effect of E1A on transcription requires MED23, a component of the Mediator complex. E1A indirectly recruits the Mediator complex by binding to TFs including members of the cAMP response element/activating transcription factor (ATF) family, activator protein-1 (AP-1), upstream stimulatory factor (USF), and Sp1 (Liu et al. 1990, Liu and Green 1994, Ström et al. 1998). CR3 also binds transcriptional coactivators (pCAF and p300/CBP) and transcriptional repressors (GCN5, BS69). E1A enhances transcriptional elongation through the Mediator complex (Vijayalingam et al. 2013) and alters host post translational modifications (PTMs) of histones and other proteins including acetylation (Ferrari et al. 2008, Horwitz et al. 2008) SUMOylation (Radko et al. 2014), ubiquitination (Isobe et al. 2009), and others. This eventually downregulates genes involved in differentiation or innate immunity genes in cell cycle and macromolecular synthesis (King et al. 2019). E1A interacts with pRB through SLiMs in CR1 and CR2, the second of which also interacts with the STING protein, which under normal functioning induces a type 1 interferon response (Liu and Marmorstein 2007, Lau et al. 2015). E1A can also bind to a component of the immunoproteasome, which is involved in the production of MHCI peptides (Berhane et al. 2011).

E1A and Host Immune Suppression

E1A has multiple methods of suppressing the host immune response. E1A blocks type 1 IFN-inducible gene expression through the CR1 region (Hendrickx et al. 2014). E1A also prevents transcription of MHC class II genes induced by type 2 interferons in response to

dsRNA (Kalvakolanu et al., 1991). Specifically, E1A blocks Stat2 transactivation by binding to RuvBL1 (Olanubi et al. 2017). Furthermore, in HAdV species 7 and 14, E1A sequesters STAT2, RuvBL1, 2, and RNAP2, further suppressing interferon- stimulated gene activation (Graves et al. 2024).

E1A and Apoptosis

E1A can induce apoptosis through both p53-dependent and p53-independent mechanisms. The p53 pathway is activated by E2F (Hale and Braithwaite 1999) and normally inhibited by MDM2, but E1A represses MDM2 (Ashcroft and Vousden 1999, Thomas and White 1998). So far, the mechanism of E1A p53-independent apoptosis is not well understood. In viral mutants lacking E1B, E1A 12S can induce apoptosis through a pathway involving caspase 8, FADD, BCL2, and TNF cascades (Nguyen et al. 1998). However, if E1B is present, 12S can only induce apoptosis with p53 (Teodoro et al. 1995). E1B 19K and 55K can both interrupt the apoptosis pathway (Burgert et al. 2002). The same regions of E1A involved in DNA synthesis and p53 activation also sensitize cells to TNF lysis through NF- κ B activation (Shisler et al. 1996, Shao et al. 1999). E1B and three E3 proteins protect infected cells from TNF cytolysis (Wold et al. 1995).

1.3 Arglu1

Arglu1 is a small 273 residue protein (33 kDa) that is predicted to be intrinsically disordered. Arglu1 potentiates breast cancer cell growth and estrogen receptor mediated gene transcription by binding to the Mediator subunit 1. Arglu1 is also involved in EBV-

associated lymphoma, stress hormone pathways, alternative splicing, head and neck squamous cell carcinoma (HNSC), ALS, hypertrophic scar formation, gastric cancer, apoptosis, DNA damage repair, cell growth, promotor proximal pausing, and possibly genetic disorders of development (Bachus et al. 2024).

1.4 Arglu1 Structure and Splicing

Arginine and glutamate rich 1 (ARGLU1) is a small 273 residue protein (33 kDa) that is likely mostly if not all intrinsically disordered, though AlphaFold does predict an alpha-helix at residues 93-255 with very high confidence (pLDDT > 90). AlphaMissense predicts that Arglu1 mutations around residues 80-90 and 110-270 (particularly 110-170) are likely to be highly functionally detrimental (Jumper et al. 2021, Varadi et al. 2024, Cheng et al. 2023). PrDOS also predicts that most of ARGLU1 is likely to be mostly unstructured. A sequence alignment found overlap between the BRD4 ET domain (residues 600–682) and ARGLU1 C-terminus (residues 101–273), suggesting that Arglu1 might compete for the BRD4 binding partner JMJD6. Structural comparison of Arglu1 and BRD4 ET domain shows both have alpha helices. (Baccus et al. 2024)

Arglu1 Literature Overview

As of November 2024, there are 18 articles in PubMed referencing Arglu1. Arglu1 was first discovered as preferentially expressed in human kidney glomerulus, mostly referred to as FLJ10154 (Cuellar et al. 2009). More important was the finding by Zhang et al. (2011) that Arglu1 bound the Mediator complex, which connects transcriptional activators and RNA

polymerase II. They also found that Arglu1 is required for estrogen-receptor mediated gene transcription and breast cancer cell growth. Arglu1 expression is elevated by the Epstein-Barr virus protein EBNA1 in EBV-associated lymphoma cells (Portal et al. 2011, 2). Pirnie et al. (2017) found that Arglu1 has three isoforms, the main one, an isoform with a retained intron that remains in the nucleus, and an isoform with an alternative exon that is targeted for nonsense mediated decay. Arglu1 has also been found to interact with components of the spliceosome. According to the GTEx database, the highest expression of Arglu1 is in the cerebellum and a loss of function mutation in Arglu1 was associated with neurological disorders (Gambin et al. 2017). It has also been shown that Arglu1 can mediate its own splicing via a non-coding ultra conserved element (Chan and Pek, 2022).

Arglu1, Splicing, and the Glucocorticoid Receptor

Magomedova et al. (2019) divide Arglu1 into 2 parts, an arginine rich beginning (3-74 aa) and a glutamate rich end (97-251 aa). The arginine rich region also has 3 SR (serine-arginine) regions, which are RNA binders, regulators, and connectors that regulate splicing (Jeong 2017) and a nuclear localization sequence (NLS). The glutamate rich region contains LLXXL (172–176 aa) and LXXIL (201–205 aa) motifs, which are canonical interaction sequences in nuclear receptor (NR) coactivators that allow binding to the AF2 domains. They used a *E. coli* luciferase Cortisol screen that identified Arglu1 as a transcriptional coactivator. Arglu1 is very highly conserved among mammals for example the sequence identity of mouse and human Arglu1 is 99%. They also found Arglu1 colocalizes with the nuclear speckle marker SRSF2. The glutamate rich c-terminus of

Arglu1 increased the transcriptional activity of several steroid receptors (ER α , MR, PR, ER β) in particular the glucocorticoid receptor (GR), although the LXXL/IL motifs were not required for activation. One unfortunate aspect of this study is that all experiments were done with HEK 293 cells which contain E1A that binds to Arglu1 and neural N2a mice cells not human neural cells. The arginine rich region of Arglu1 is associated with splicing. Analysis of RNA-seq data found Arglu1 knockdown results in alternate splicing through exon skipping and exon inclusion. Arglu1 knockdown also resulted in a more than 90% reduction in mRNA and protein. An RNAcomplete assay found RNA binding to Arglu1 and Arglu1 N-term but not C-term. They also found that loss of Arglu1 is embryonic lethal in mice and causes neurodevelopment defects in zebrafish.

Arglu1 and HNSC

Ye et al. (2022) used clinical data from head and neck squamous cell carcinoma (HNSC) patients from the TCGA database to construct a correlation network. ARGLU1 along with DDX39B and PRPF39 were identified as key splicing factors regulating alternate splicing. A mouse knockout model of Arglu1 found differential splicing events involving proteins such as Hnrnpa2b1, Zfp317, Tmem209, Mdm2, and Mdm4. (Yao et al. 2023)

1.5 Arglu1 in DNA Damage, Cell Cycle, Growth, and Cancer

Daneshafrooz et al. (2022) ran weighted gene go-expression network analysis (WGCNA) and LASSO regression analysis on transcriptomics of blood samples from ALS patients and

identified Arglu1 as a potentially diagnostic biomarker for ALS. They hypothesize that Arglu1 functions as an apoptotic factor, referencing how targeting of Arglu1 by miRNA causes apoptosis and proliferation inhibition in hypertrophic scar formation and breast cancer. Arglu1 mutation in human patients associated with intellectual disability, seizures, congenital heart defects, anorectal malformation and hearing loss, some of which could be caused by DNA damage repair, however Arglu1 is rarely the only gene mutated, which makes causal specification difficult (Yao et al. 2023).

Arglu1 and Neurogenesis

A mouse knockout model of Arglu1 caused microcephaly, delocalisation and detachment of radial glial cells, impaired neurogenesis due to delayed mitotic progression and in particular, an increase in apoptosis through upregulated CASP3, p53, PUMA, and NOXA. Removing p53 eliminated the apoptosis and microcephaly but not the delayed mitotic progression suggesting that Arglu1 mediated cell cycle effects are not via a p53 pathway (Yao et al. 2023).

Arglu1 and Hypertrophic Scar Formation

In samples taken from human patients, qPCR found that Arglu1 was upregulated in hypertrophic scar (HS) tissue and hypertrophic scar fibroblasts (HSFs). Hypertrophic scar (HS) formation is a dysfunctional wound healing process caused by defects in cell proliferation and excessive collagen biosynthesis and extracellular matrix (ECM) remodeling. Knocking down Arglu1 inhibited cell proliferation, collagen and ECM

production, as well as overall cell viability, probably caused by increased caspase-3 and -9 activity. It was also shown that the miRNA miR-6113-3p suppressed Arglu1 by binding to its 3'-UTR. It was previously found that overexpressing miR-6113-3p inhibited hepatocellular cancer cell proliferation and cell cycle progression and that miR-6113-3p was downregulated in colon cancer cell lines and upregulated in acute hepatitis B viral infection of liver cells and blood and neuron samples of patients with Parkinson's disease and Alzheimer's disease, all of which are associated with apoptosis. (Li et al. 2021)

Arglu1 and Colon Cancer

In the TCGA and GTEx databases, Arglu1 was significantly downregulated in colon cancer tissue samples. Knocking down Arglu1 in human colon cancer cell lines could significantly increase gastric cancer cell viability while upregulating Arglu1 decreased gastric cancer (GC) cell proliferation. Similar results occurred upon transplanting cells into nude mice. Gene Set Enrichment Analysis (GSEA) found that Arglu1 overexpression was positively correlated with DNA Mismatch Repair (MMR) terms (single base mismatches and small insertion/deletion loops). Overall, 23 MMR genes were associated with knockdown or overexpression of Arglu1 by qPCR including MLH3, MSH2, MSH3 and MSH6. Furthermore, ChIP-Seq found that Arglu1 bound the promotor regions of the above 4 genes and luciferase reporter assays found that transcription of these genes was highly affected by Arglu1. Combining transcription factor prediction with mass spectrometry identified SP1, YY1, MLLT1, POLR2A and RBPJ as potential Arglu1 binding partners. Coimmunoprecipitation found Arglu1 binding to SP1 and YY1 and ChIPs on Arglu1

knockdowns found significantly attenuated signals for promoter regions of MLH3, MSH2, MSH3 and MSH6 when pulling down with SP1 and YY1 antibodies. Arglu1 overexpression also inhibited GC growth in a mouse model. These results might seem to be inconsistent with other results (see below) but it has been shown that in gastric cancers, higher expression of mismatch repair (MMR) proteins is associated with delayed GC progression (Park et al. 2012, Schmitz-Winnenthal et al. 2007, Koedoot et al. 2019). This confirms the important effect of Arglu1 on genome stability and suggests that whether Arglu1 is oncogenic or tumor suppressive could be determined by the effect of specific types of DNA Damage Response on specific cancers (Li et al. 2021).

More on Arglu1 and Cancer

Arglu1 siRNA upregulated CSL, CCND3, HES1 and c-MYC in HNSCC (head and neck squamous cell carcinoma) cells. Arglu1 siRNA also increased cell proliferation and migration in HNSCC cells and tumor growth in nude mice. Also, the protein DTL degrades Arglu1 via ubiquitination-mediated proteasomal degradation to activate the CSV-dependent Notch signaling pathway, which can increase HNSCC proliferation and migration (Shi et al. 2024). The Wellcome Sanger Cancer Dependency Map designates ARGLU1 as a pan-cancer core fitness protein with high target priority score for many cancers, suggesting strong therapeutic potential. (Bachus et al. 2024)

Team Pelka Enters The Frey!

Arglu1 binds to E1A N-terminus (around 26-49 aa) and Arglu1 overexpression (HTA cells) decreases HAdV growth and gene expression in species C5 and B7. Searching the 26-39 aa sequence in HAdV-C5 E1A, that is bound by Arglu1, in protein BLAST identified Lysine Methyltransferase 2D (KMT2D), Arginine-Glutamic Acid Dipeptide Repeats (RERE), and JMJD6, among other proteins involved in transcription, chromatin remodelling, and histone methylation. Arglu1 siRNA knockdown increased viral growth and gene expression. HTA cells grow about 1.5 times faster than regular HT1080 (HT) cells (doubling time 20 h vs 30 h), HT siArglu1 cells grow about 1.5 times slower (doubling time 50h), possibly to the point of growth arrest. Arglu1 overexpression also increases cell resistance to the DNA damage inducing drug bleomycin, which binds to GC-rich regions and induces single strand breaks, eventually leading to double strand breaks (DSBs) (Murray et al. 2018). A comet assay found that HTA cells have less endogenous DNA damage than HT cells and that HTA cells are faster at repairing DNA damage from bleomycin than HT cells. Inhibiting Brd4 activity via the drug JQ1 decreased the amount of DNA damage response (DDR) in HTA cells but not HT cells. Knocking down Arglu1 also decreased DDR. Assays using U2OS cells found that HTAs increased both homologous recombination (HR) and non-homologous end joining (NHEJ) DDR and that Arglu1 reduces gene expression near a site of DNA damage. Immunofluorescent staining found that Arglu1 can colocalize to DSBs. RNA polymerase II (RNAP2) was also significantly more ubiquitinated in HTA cells, which indicates stalled RNAP2. Adding bleomycin did not change the ubiquitination. Inhibiting Brd4 with JQ1 significantly reduced the growth rate of HTA cells, while knocking down ERCC6, another protein involved in DDR, did not. HTA cells also have more paused RNAP2 on promoter

proximal regions of five genes: RHOB, ALDOA, PCNA, MCM4, and HAdV E4.

Overexpressing Arglu1 almost entirely eliminated BRD4-JMJD6 binding.

1.6 Objectives

1. Global expression analysis of how E1A isoforms and mutants (13S, 12S, and 13S YCD) effect human transcript expression with particular focus on how YCD effects the cell cycle.
 - Rationale: Exploratory data analysis for how YCD could drive S phase. Cell transformation is one of the most studied aspects of E1A and Adenovirus and further understanding of it could lead to a better understanding of cancer and possible new chemotherapies.
2. Make Relevant Deletion Mutant Plasmids and Stable Cell Lines
 - Rationale: Identify where important Arglu1 binding partners bind and correlate their deletion to loss of phenotype induced deletions (i.e. growth, etc.).
3. Use proteomics and ChIP-seq to generate new gene targets for how Arglu1 effects cell growth, DNA damage, and polymerase pausing.
 - Exploratory data analysis for how Arglu1 could affect the cell cycle, pausing, etc. Could identify other functions of Arglu1 and how Arglu1 effects known phenotypes and why it is an essential survival factor for cancer.

2 Methods

2.1 RNA-Seq

In IMR-90 cells, Four different Adenovirus strains (WT, E1A 13S, 12S, and YCD) were infected for 24h along with a mock/uninfected control and subsequently harvested and sequenced. IMR-90s (primary human lung fibroblasts) were used because they are commonly used in adenovirus transcriptomic studies such as in (Zhao et al 2019) and recommended in (Zhao et al 2020) because cancer cells such as HeLA and A549 have different gene expression compared to normal cells, more variable gene expression in individual cells, and a faster viral replication cycle. RNA sequencing generated between 38 to 44 million paired reads per library using the Illumina Hiseq 2000/2500 sequencer. Libraries were prepared using the first strand TrueSeq mRNA protocol. Base calls were made using the Illumina CASAVA pipeline. Trimmomatic was used to remove Illumina sequencing adapters and to trim and clip low quality reads (Bolger et al. 2014). The reads were aligned to the *Homo sapiens* reference genome GRCh37 using the STAR aligner (Dobin et al. 2013). Cufflinks were used to assemble aligned RNA-seq reads into transcripts and generate raw read counts, respectively (Trapnell et al. 2010, Anders et al. 2014). The above was performed by the McGill Innovation Center. Differential expression analysis (including normalization) was performed on raw reads using DESeq2 (Love et al. 2014). Genes with total counts less than 10 across all 10 samples were removed. The 'lfcshrinkage' function was used to reduce the variability of genes with low counts. Genes

were designated as differentially expressed if they had a fold change of greater than $|2|$ and an adjusted p value (Benjamin-Hochberg) of less than 0.05 as in (Zhao et al 2019). PCA charts were generated using pcaExplorer (Marini and Binder 2019), heat plots with Pheatmap (Kolde 2019), and volcano plots with ggplot2 (Wickham 2016). Hierarchical clustering was performed using hclust R function and gene enrichment analysis and GSEA was performed using the g:profiler and clusterProfiler r packages (Yu 2024).

RNA-Seq Normalization

RNA-seq data was first normalized using RPKM (single end reads) or FPKM (paired end reads). RPKM (Reads per Kilobase Million) normalizes for sequencing depth by dividing the read counts by 1 million, a number chosen for producing nice numbers i.e. without too many decimals. RPKM then normalizes for gene length by dividing by gene length (hence per kilobase). FPKM (Fragments Per Kilobase Million) is the same as RPKM except that it controls for cases where only one of two ends is mapped. The second commonly used method for normalization is TPM (Transcripts per Million), which is basically the same 2 steps as RPKM/FPKM except done in the reverse order. This allows for easier comparison of proportion of total reads compared RPKM because in RPKM each sample has a different total. However, these methods only normalize for library size (reads per sample i.e. sequencing depth) they do not account for differences in library composition like DESeq2 or edgeR (Love et. al. 2014). Why do we need to correct for library composition? Suppose you have read counts for 4 transcripts across 2 groups and genes A, B, and C have the same read counts in both groups. But one of these groups gene D has a very high read

count and in another it has 0 transcripts. If you just normalize for sequencing depth then it will look like all genes are differentially expressed, even though only D is.

In order to determine if a transcript is statistically differentially expressed, we must have some assumptions or a model (i.e. a probability distribution) to which transcripts do not have average or expected expression values. One common example of a probability distribution is a normal distribution. But RNA-seq counts data is not normally distributed. Instead there are more low values than high and the data does not fall within a specific range. For data like this, a better approach is instead to model it with a Poisson distribution. [Stats citation] However, the Poisson distribution assumes that the mean is equal to the variance, which is not the case in RNA-seq experiments, in particular genes with high counts tend to have higher variance than the mean and genes with low counts have high amounts of dispersion of variation i.e. different low genes will have a higher variance of variances compared to high count genes. DESeq2 accounts for this by using a negative binomial distribution with a fitted mean and a dispersion parameter (Love et. al. 2014).

The fitted mean in DESeq2 is equal to a size factor (specific to each transcript, used to account for library size and library composition) multiplied by a general linear model (aka linear regression) specific to the experiment design (ex. treated vs untreated). DESeq2 estimates the size factor using the median ratios method, which is robust to outliers which reduces library composition bias. A crucial assumption is that most genes are not differentially expressed. Samples are normalized for library composition and size by the

median of ratio method, briefly the untreated counts are normalized by dividing them by the median of the ratios of transcripts divided by their geometric mean (and similarly for the treated counts). This normalization method is used because it is robust to outliers. The dispersion parameter is calculated using a model that calculates a curve based on the relation between mean norm counts and dispersion. Finally, a Wald test is used to determine if there is sufficient evidence that the fold change of a transcript is not 0, producing a p-value reflecting the certainty (Love et. al. 2014). HOMER uses a similar approach to DESeq2 for analyzing which peaks are differentially occupied, at some points even using DESeq2 code (Heinz et. al. 2014).

2.2 Cell Cycle Analysis

HT1080 fibrosarcoma cells were used for this and subsequent experiments because we had a cell line overexpressing Arglu1 (HTA). "Double-thymidine block was performed essentially as previously described (Chen and Deng 2018). Briefly, HT and HT-A cells were plated at 1.5 million cells in 10 cm plates in 5% FBS DMEM and incubated at 37°C overnight. Thymidine was added to a final concentration of 2 mM and cells were incubated for 18h. Cells were washed with PBS, media [changed], and incubated for 9h. Thymidine was added again to a final concentration of 2 mM and cells incubated for 18 h. Cells were released from thymidine block by PBS rinse and media replacement. Time points were harvested every 2 h for 36 h with T0 harvested directly after thymidine block release.

Harvested samples were then stained using propidium iodide as described below.”

Methods previously described in Baccus et al. (2024).

2.3 Actinomycin D Drug Focus Forming Assay

“Two hundred thousand HT or HT-A cells were plated in 6-well plates and allowed to reach 90% confluency at which point 5 μ M or 50 nM bleomycin or actinomycin D was added, respectively, and kept on for the next 24 h. Media was then changed, and cells were allowed to recover and form foci for 7 days for bleomycin-treated cells and 21 days for actinomycin D-treated cells. For hygromycin B control, cells were treated with 350 μ g/ml of hygromycin B for 24 h, drug was removed and replaced with fresh media, and foci were allowed to form for 21 days. Media was then removed from the cells, the cells were rinsed with PBS twice and fixed using 4% formaldehyde in PBS for 10 min. Plates were subsequently rinsed with PBS two more times for 10 min, then stained with crystal violet for 10 min. After staining, cells were repeatedly rinsed with PBS until the removed PBS ran clear.” Methods previously described in Baccus et al. (2024).

2.4 Proteomics

Two batches of samples were sent for analysis in 2021 and 2023 using essentially the same methods. HT and HTA cells were harvested and lysed in PBS by sanitation. TCA (trichloroacetic acid) was used to precipitate 300 μ g per sample. Samples were sent to

Southern Alberta Mass Spectrometry Centre for analysis by 2D-LC-MS/MS DDA (data-dependent acquisition) with label free quantification was performed on a [mass spec instrument]. Mascot [v2.7.0] was used for peptide identification and Scaffold (v5.3.0) was used quantification. The fragment and parent tolerances were 0.60 Da and 10.0 PPM (Monoisotopic), respectfully. Fixed Modifications were +18 on and +57 on C (Carbamidomethyl) while +16 on M (Oxidation) was a variable modification. Trypsin was used as a digestion enzyme and 1 missed cleavage was allowed. The peptide threshold was 95% min with a FDR of 0.5% (Prophet). The protein threshold was 95% min with a minimum of 1 peptide and an FDR of 0.7% (0.5% for 2023) with the Peptide Prophet algorithm with Scaffold delta mass correction. Protein intensities were normalized and missing values were stochastically approximated using proDA (Ahlmann-Eltze and Anders 2019). Heat maps were generated using pheatmap (Kolde 2019) and GSEA analysis and bar graphs by clusterProfiler (Yu 2024).

2.5 Deletion Mutant Plasmids

Primers were designed by me and Peter Pelka. PCR was done using a pCAN-HA-Arglu1-Amp template and Thermo Scientific Phusion High-Fidelity DNA Polymerase Kit. 2.5 ul of pmol/ul of each primer was used (Table _). DNA Template (Arglu1) concentration varied from 10-100 ng/ul. The samples were mixed on ice. The temperature of annealing step on the thermocycler was varied from 55-72C depending on the optimal annealing temperature (from ThermoFisher Tm Calculator) and trial and error. The rough length of PCR products

was verified using a DNA gel. After DNA gel verification, PCR products were gel extracted using a Bio Basic Molecular Biology Kit using 500 ul of binding buffer. 17 ul of column purified

PCR products were then phosphorylated using NEB T4 Polynucleotide Kinase in double distilled water with T4 DNA Ligase Buffer, then ligated using NEB T4 DNA Ligase with added ATP. Sometimes the PNK solution was diluted by ½ before adding the DNA ligase. Solution was incubated for 2 hours at 14C then 2 hours at room temperature. After incubating, the plasmids were transformed into *E. Coli* resistant to Ampicillin with LB media and plated on LB Amp plates. 10-20 ul of ligase solution was mixed with 50 ul of *E. Coli* on ice for 30 min before heat shock at 42C for 30s at which point 50 ul of LB broth with ampicillin was added for recovery at which point cells were plated using sterile glass beads. Colonies were then grown in LB Amp media and were mini-prepped using Bio Basic Molecular Biology Kit. Plasmids purified by miniprep were then digested with NEB EcoRI and Xba1 and run on a 1.5% Agarose gel against a control of the template HA-Arglu1-Amp plasmid to verify the plasmids were smaller. At this point, plasmid concentration was determined using a Molecular Devices SpectraMax iD3 with SpectraDrop Micro-Volume Microplate.

PCR Component	50 ul (1x)	7x Master Mix
H2O	30.6	214
5x Phusion HF Buffer	10	70
10 mM dNTPs	1	7.0
Template DNA	2.94	20.59

Primer A	2.5	
Primer B	2.5	
Phusion DNA Polymerase	0.5	3.50
Total	50.0	45
Phosphorylation	Amount	
DNA (from 30 ul PCR column purification)	17 ul	
T4 Buffer	2 ul	
T4 PNK (Polynucleotide Kinase)	1 ul	
DNA Ligation	Amount	
Phosphorylation Solution (from above)	9-18 ul	
T4 DNA Ligase	1-2 ul	
10 mM ATP	1-2 ul	
Transformation	Amount	
E. Coli cells	50 ul	
Ligation Solution (from above)	10-20 ul	

Table 1: Table: PCR, Phosphorylation, Ligation, and Transformation details

Screened plasmids (200-300 ng) were sent for Sanger sequencing to Sick Kids The Center for Applied Genomics. Verified plasmids were purified using Thermo Fisher Midiprep Kit. Once verified by sequencing, plasmids were expression tested by transfecting plasmids (4 ug per 6 cm well) into HT1080 cells with polyethylenimine (PEI). After 24 hours, cells were lysed using NP-40 lysis buffer and protease inhibitor. Protein lysates were resolved on 8% SDS-PAGE gels and then transferred to PVDF membrane. Membranes were blocked with 3% BSA in TBS and then incubated with antibodies overnight at 4°C and subjected to western blot. The primary antibody was anti-HA with a HRP anti-rabbit secondary.

2.6 Stable Cell Lines

HT1080 cells were transfected with a deletion mutant plasmid of interest and allowed to grow for about 24h. At which point the media was changed to include 375 ug/ml Neomycin for negative selection of cells transfected with a resistance marker in the plasmid. Media was changed regularly until cells were confluent (at least 2 weeks) at which point plasmid overexpression was verified using a western blot.

2.7 ChIP-Seq

ChIPs were performed by Peter Pelka on 4 samples (1 on HT cells pulling down RNA Pol II, 2 on HTA cells pulling down RNA Pol II and Arglu1 and IgG as a control) essentially as described in Baccus et al. 2024. Sequencing was done on an Illumina Hiseq 2500 at the Donnelly Sequencing Centre at the University of Toronto. BBTools was used to remove Illumina adaptors and trim and clip reads. The reads were aligned to the *Homo sapiens* reference genome Hg38 using the BMap aligner (Bushnell 2023). Peak calling/normalization, annotation, and motif enrichment analysis was done using Homer (Heinz et al. 2010). Peak visualization was done using Drompa+. Pausing Index was determine using PIC (Nakato and Sakata 2020). Pausing Index was calculated using software from Chen et al. (2017) and Day et al. (2016).

Software	Reference	Website
R Version 4.4.1 (Race for Your Life)		https://www.r-project.org/

DESeq2 1.46.0	Love et al. 2014	https://github.com/thelovelab/DESeq2
ClusterProfiler 4.14.3	Yu 2024	https://yulab-smu.top/contribution-knowledge-mining/
pcaExplorer 3.0.0	Ludt et al.	https://github.com/federicomarini/pcaExplorer
ggVennDiagram 1.5.2	Gao et al. 2024	https://github.com/gaospecial/ggVennDiagram
pheatmap 1.0.12		https://github.com/raivokolde/pheatmap
Drompa+	Nakato and Sakata 2020	https://drompaplus.readthedocs.io/en/latest/
Homer 5.1	Heinz et al. 2010	http://homer.ucsd.edu/homer/
BBTools 39.11		sourceforge.net/projects/bbmap/
Pausing Index Calculator 1.0	Day et al. 2016	https://github.com/MiMiroot/PIC
proDA	Ahlmann-Eltze and Anders 2019	https://github.com/const-ae/proDA
Trimmomatic	Anthony et al. 2014	http://www.usadellab.org/cms/?page=trimmomatic
Cufflinks	Trapnell et al. 2010	https://cole-trapnell-lab.github.io/cufflinks/
STAR aligner	Dobin et al.	https://github.com/alexdobin/STAR
g:Profiler 0.2.3	Kolberg et al. 2023	https://biit.cs.ut.ee/gprofiler/gost
MASCOT 2.7.0	Matrix Science	https://www.matrixscience.com/
Scaffold 5.3.0		https://www.proteomesoftware.com/products/scaffold-5
ggplot2 3.5.1		https://ggplot2.tidyverse.org/

Table 2: Software Ysed

3 Results

3.1 Adenovirus E1A RNA-Seq

Adenovirus E1A is one of the earliest expressed viral proteins during Adenovirus infection. It is an oncogene in cell culture and activates S phase through the Retinoblastoma protein (pRB) pathway. E1A also helps suppress the host immune response, among other effects. There are five isoforms of E1A, the earliest expressed of which are 13S and 12S. The difference between E1A 13S and 12S is that 12S is lacking most of the conserved region 3 (CR3) and is deficient at transactivation (Figure 3 top). Previously YCD mutants of 13S and 12S were made that deleted residues 2-11 and mutated Y47H and C124G. The YCD mutants are deficient at binding pRb and E2F, and viral growth. However, in a BrdU assay, 13S YCD was still able to drive S phase at a rate similar WT and 13S, unlike 12S and 12S YCD (Figure 3 bottom). BrdU is a thymidine analogue that is incorporated into replicating DNA and can be detected with anti-BrdU antibodies (Pelka et al. 2011). This was unexpected as most previous research has attributed cell cycle activation in E1A to CR1 and 2, both of which are in 12S (King et al. 2019). Unless mentioned otherwise, YCD below refers to the E1A 13S YCD mutant.

	Uninfected	Ad WT	E1A 13S	E1A 12S	E1A YCD
<i>Base Counts</i>	22081	22267	21736	21914	21753
<i>Fold Change > 2</i>		3059	2453	1118	3419
<i>Upregulated</i>		2072	1620	846	1893
<i>Downregulated</i>		987	833	272	1526

Table 3: RNA-Seq Samples. In total 23,289 genes were detected in at least one of the 5 samples. A total of 6385 genes had an adjusted p value (BH) of less than 0.05 and were up- or down-regulated more than 2-fold in at least one of the 4 experimental samples using uninfected as a control.

	13S/WT	12S/WT	13S/12S	YCD/WT	YCD/13S	YCD/12S
<i>Fold Change > 2</i>	81	249	282	1669	822	1735
<i>Upregulated</i>	11	8	233	679	373	869
<i>Downregulated</i>	70	241	49	990	449	866
<i>FC > 1.5</i>	362	927	973	3429	1988	3890
<i>Upregulated</i>	105	134	647	1449	906	1856
<i>Downregulated</i>	257	793	326	1980	1082	2034

Table 4: RNA-Seq Sample Comparisons

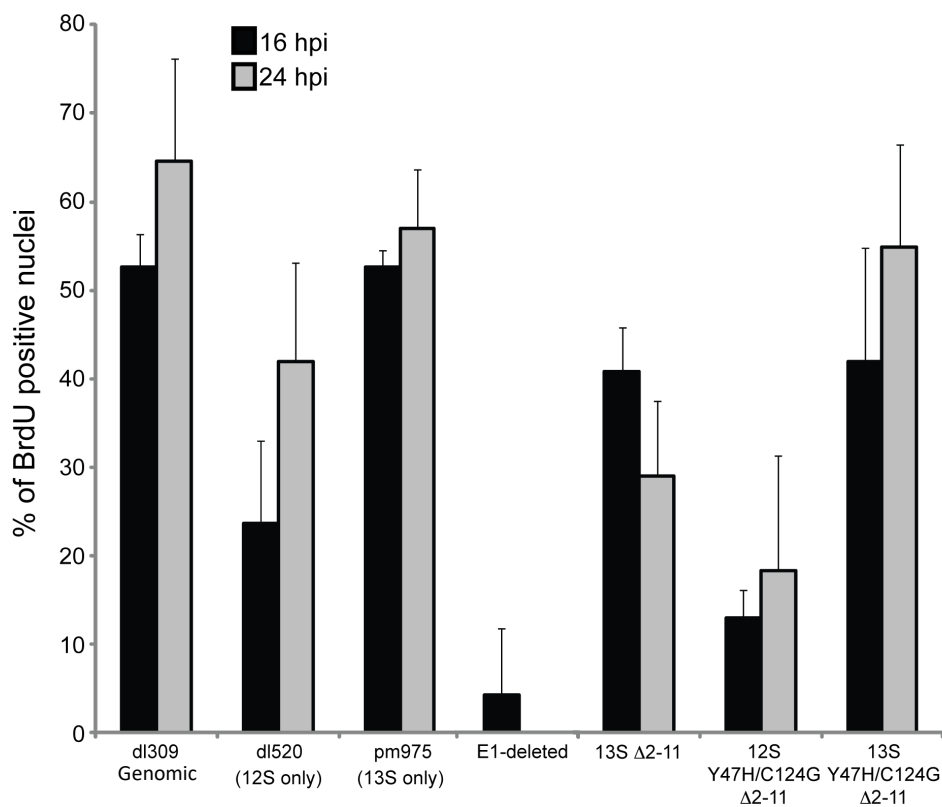
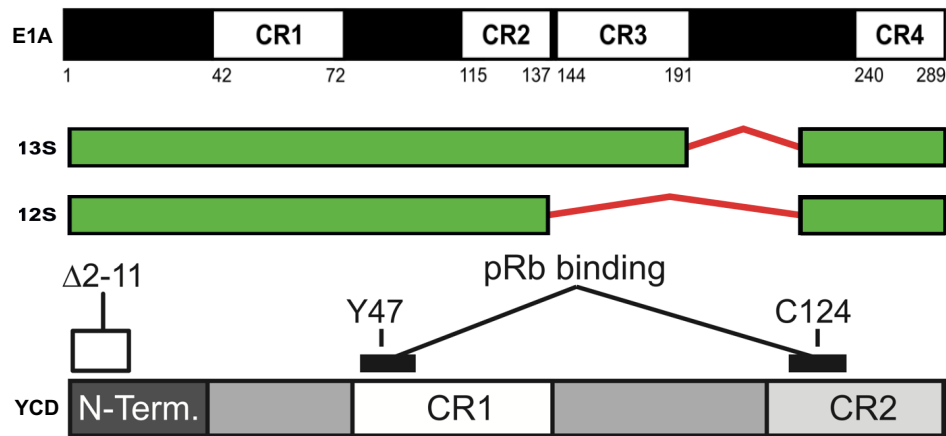


Figure 3: E1A 13S YCD mutant is able to induce S phase. (Top) Schematic representation of E1A Isoforms and YCD mutant. Numbers are nucleotides. The YCD mutant deletes residues 2-11 and mutates Y47 and C124 and as a result is deficient at binding pRb and E2F, and viral growth. (Bottom) BrdU Assay. Arrested IMR-90 cells were infected with the indicated adenoviruses, pulsed with BrdU for 1 h at the indicated time points, fixed, and stained with anti-BrdU anti-body. Results are expressed as a percentage of BrdU-positive nuclei and constitute a mean of three random fields of view at a low magnification.

Global Expression Analysis of E1A Mutants and Isoforms

To uncover how YCD can drive S phase, as well as other differences between the E1A isoforms, we performed RNA-seq on 2 Adenovirus isoforms (13S and 12S) and 1 mutant of 13S, designated YCD with 2 point mutations targeting the pRb binding residues and residues 2-11 deleted that eliminate binding to the E2F/DP1 complex. These mutations eliminate the ability of E1A to initiate S phase by E1A243R (12S derived) and decreases viral growth. Principal component analysis shows that WT, 13S, and 12S are the most similar of the strains (Figure 4 top, middle) and that YCD and Uninfected are very different from the other strains and each other. Principal component analysis is a type of clustering algorithm (specifically dimensionality reduction) that transforms the data into a series of components in descending order of the amount of variation they represent or explain (Deisenroth et al. 2021). For this data set, the first and second principal components explain 52% and 33% of the expression variation, respectively. The other principal components had low explanatory values and hence were not included in the plot, as is customary. The strain replicates are all very close to each other and not with any of the other virus types, indicating high expression differences between the samples and suggesting the samples are high quality and do not have batch effects. Similarly, the raw counts for each of the samples are similar (Figure 4 bottom). Some of the genes that contribute most to PC1, which differentiates the control from the other samples are Myc and IL1B while CXCL2 and IL11 are important for PC2, which differentiates between YCD from the more similar cluster of 13S, 12S, and WT (Figure S2). Gene ontology (GO) analysis of PC1 and PC2 suggests gene functions underlying the clustering differences (Tables S3

and S4). For the negative part of PC1 (Mock and 12S), enriched GO terms included SMAD protein signal transduction and cell migration while for PC1+ (WT, 13S, YCD) enriched GO terms included DNA replication, cell cycle, and DNA repair. PC2- (13S, 12S, WT) also showed enriched DNA repair and cell cycle while PC2+ (Mock, YCD) had increased inflammatory and cytokine responses. PCA and Heat maps with 3 samples (WT, 13S, and 12S) had similar results (Figures S4-S6).

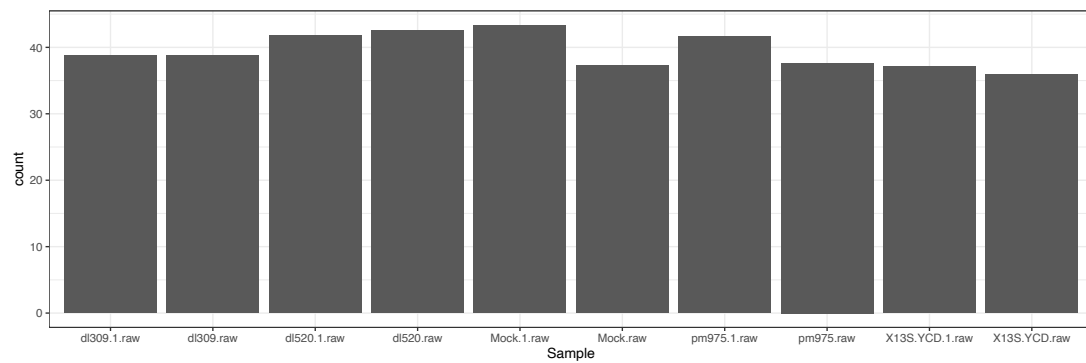
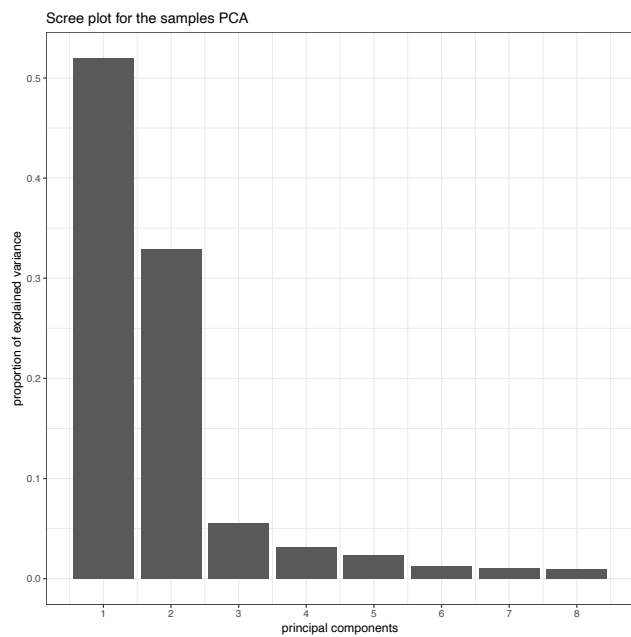
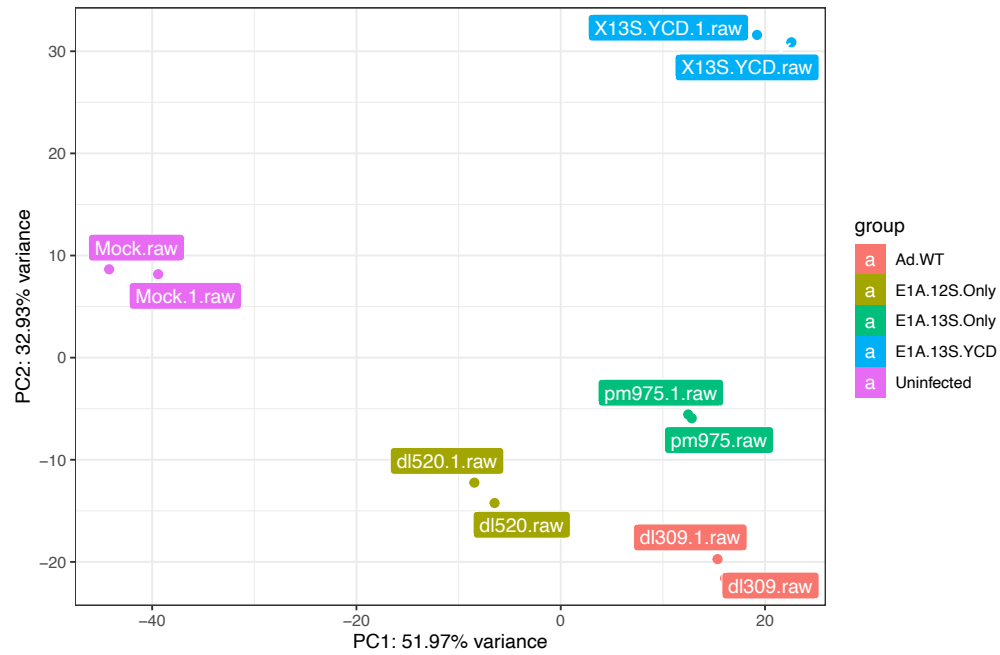


Figure 4: YCD mutant is highly distinct from other groups. (top) RNA-Seq PCA. Principal Component Analysis plot generated using pcaExplorer with scaling via DESeq2 rlog transformation and displaying the first two principal coordinates. Individual replicates are colored according to group (WT, 13S, YCD, etc) , (middle) PCA Scree Plot, displaying the 8 principle components on the x axis and the percent of variance they explain on the y axis. (middle), (bottom) Sample Counts, amount of reads for each sample.

Differential Expression Analysis

The raw counts for each strain all have about 22 thousand transcripts (Table 3). Differential expression analysis was then performed using DESeq2 (Tables S1 and S2). Note that differential expression analysis always compares two sets of samples say X/Y, where Y can be a control (ex. WT) or another experimental condition. For a comparison X/Y, transcripts are referred to as upregulated or overexpressed when their fold change is higher in X than in Y, conversely for downregulated/underexpressed. Upon filtering for significance (adjusted p value less than 0.05) and fold change, most of the samples had 2.5-3 thousand transcripts with a fold change of more than 2, except E1A 12S which only had 1118 differential transcripts. All samples had more transcripts upregulated than downregulated, with WT, 13S, and 12S having more than double the amount of upregulated compared to downregulated transcripts. In contrast, YCD had only slightly more upregulated transcripts. The comparison with the lowest differentially expressed transcripts was E1A 13S over Adenovirus Wild Type (Table 4), this makes sense because they are the most similar. The comparison with the highest amount of differentially expressed transcripts is E1A 13S YCD over WT, providing further indication how deficient the mutant is at viral function. Using the uninfected sample as a control, the YCD mutant and WT have the highest number of unique differentially expressed transcripts with 1440 and 648, respectively (Figure 5 top). There are 628 transcripts common to all four non-control strains while 13S and 12S have

much less unique transcripts, only 119 and 20 respectively. Similar results occur for analysis using WT as a control (Figure 5 bottom).

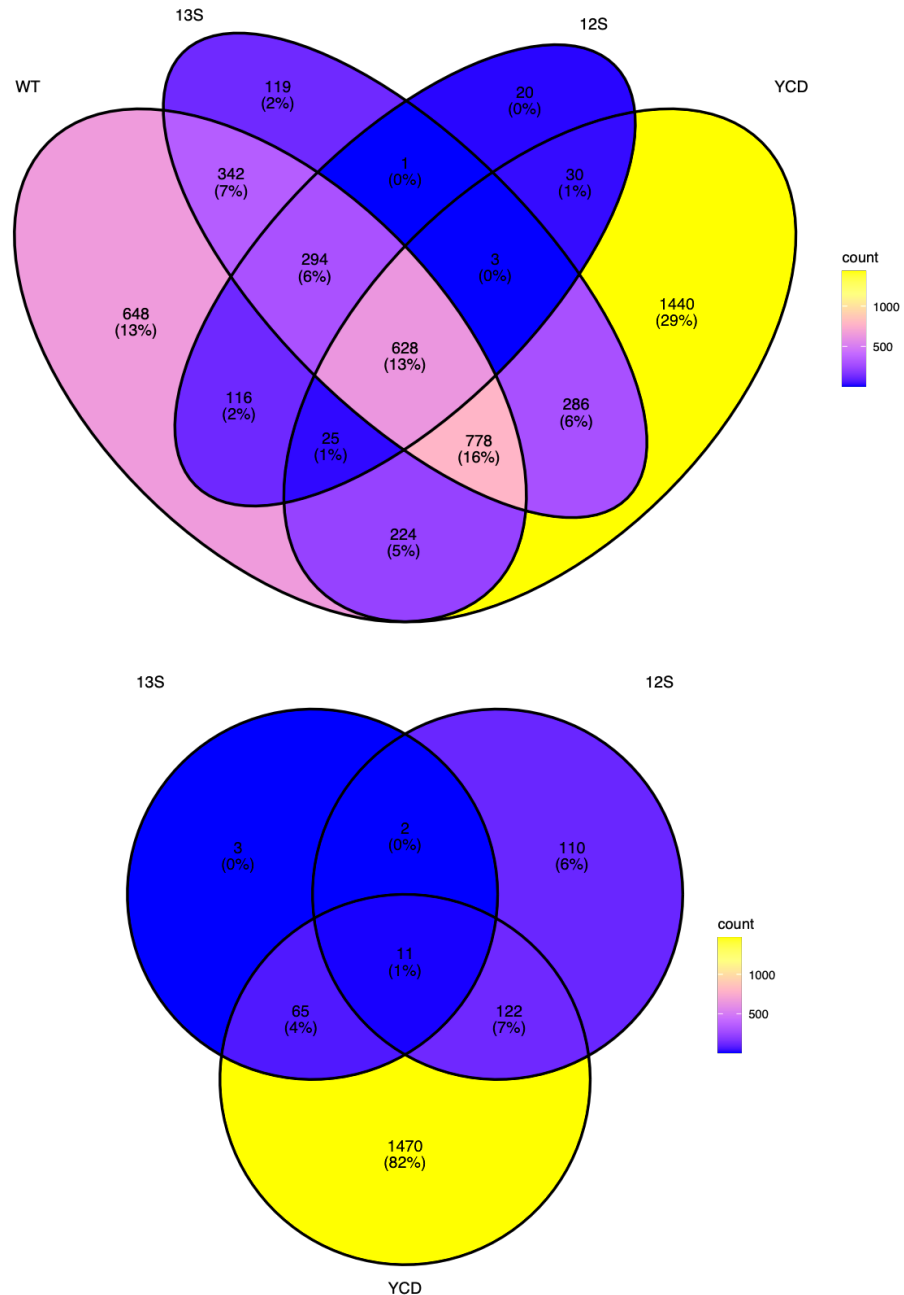


Figure 5: RNA-Seq Venn Diagrams (top) over uninfected (mock) (bottom) over HAdV WT. Venn diagrams generated using ggVennDiagram, displaying the amount of differentially expressed transcripts ($FC > 2$, $padj < 0.05$).

Clustering I

A hierarchical clustering-based heat map of the top 50 most variable transcripts across the 5 strains (Figure 6) also shows that YCD and Uninfected are the most distinct. YCD has higher expression of heat shock proteins, cytokines like IL11, and other chemokines like CYCL5. The PCA and heat map also verify that there are no batch effects or other problems since each sample is closest to its replicate. Interestingly, heat shock proteins are also some of the most differentially expressed transcripts in 13S over WT (Figure 7) while E2F2 is lower expressed, suggesting having all isoforms is important for optimal activation of transcription. Transcripts overexpressed in 12S include the cell cycle proteins CCNE2 and CDC25A as well as the tumor suppressor gene KISS1R. In YCD, transcripts involved in cell stress and alcohol metabolism are underexpressed while 13S/12S and the samples over UI (Uninfected/mock infection control) show similar results to over WT. See also the sample-to-sample heat map (Figure S1) for another indication of high sample quality and high similarity between replicates.

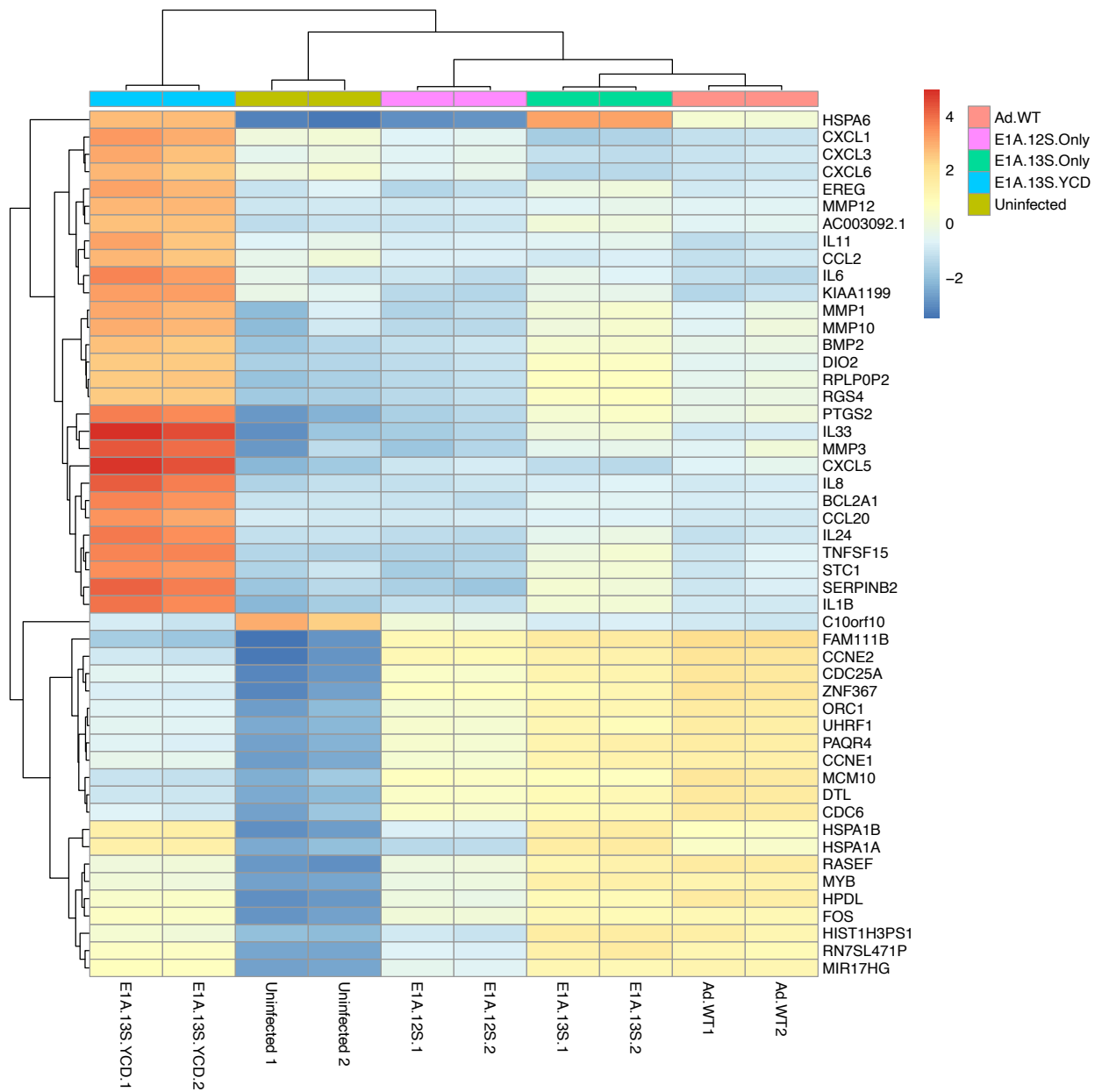
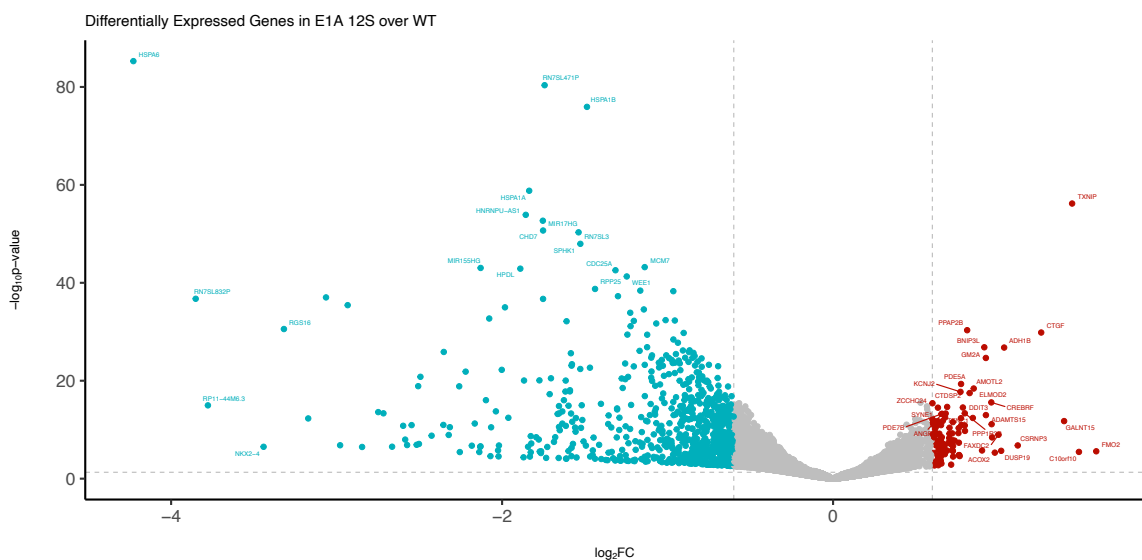
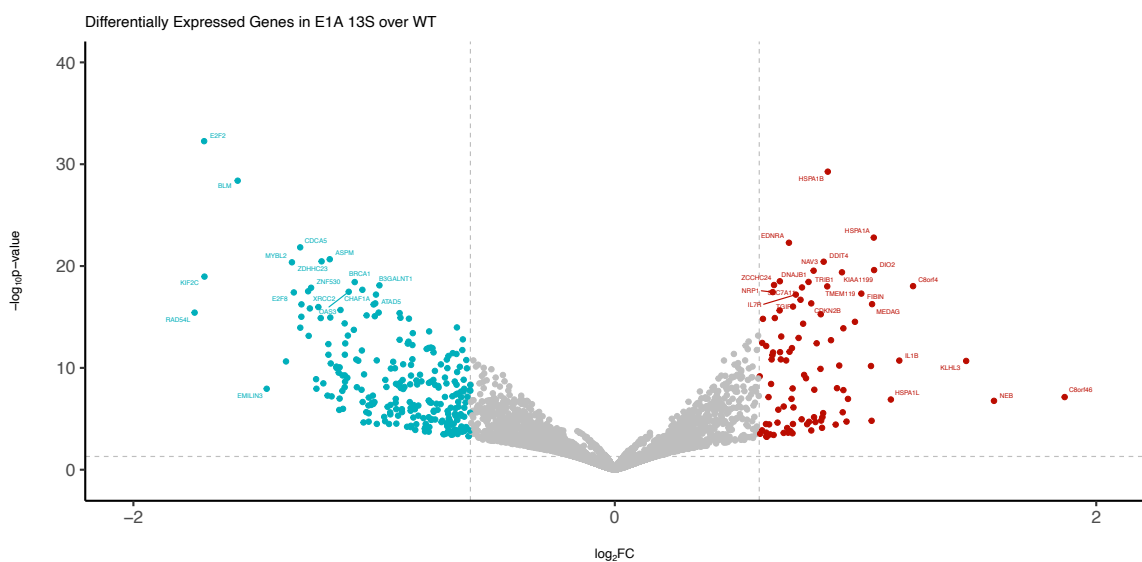
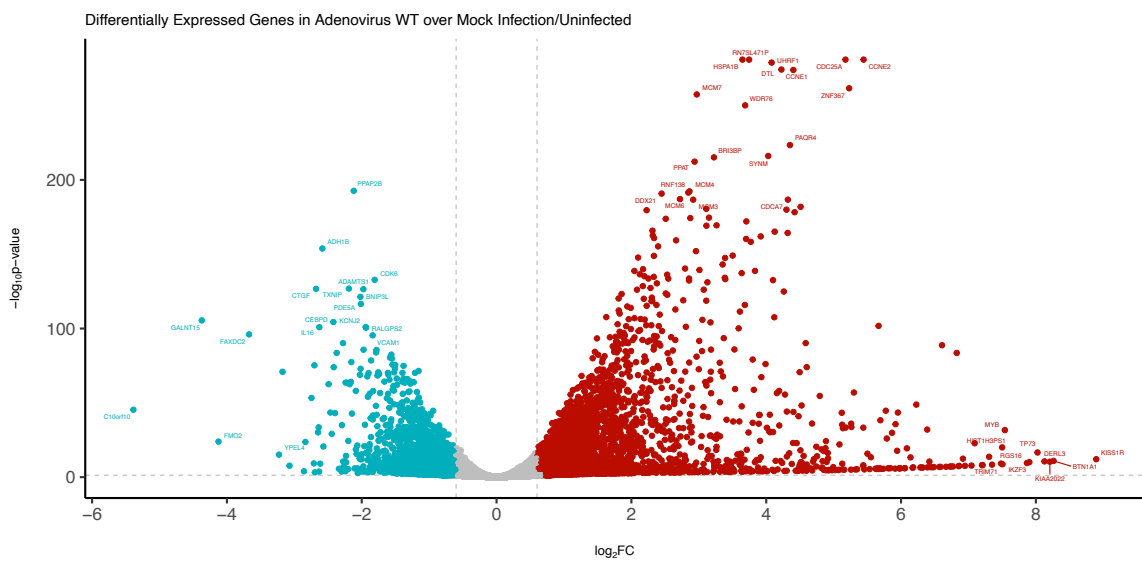
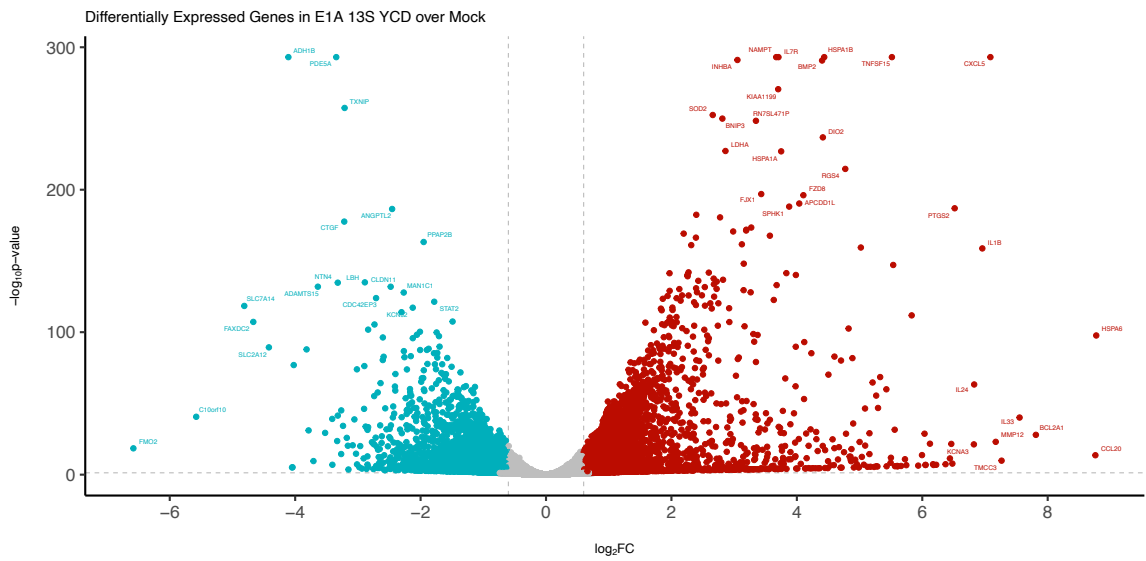
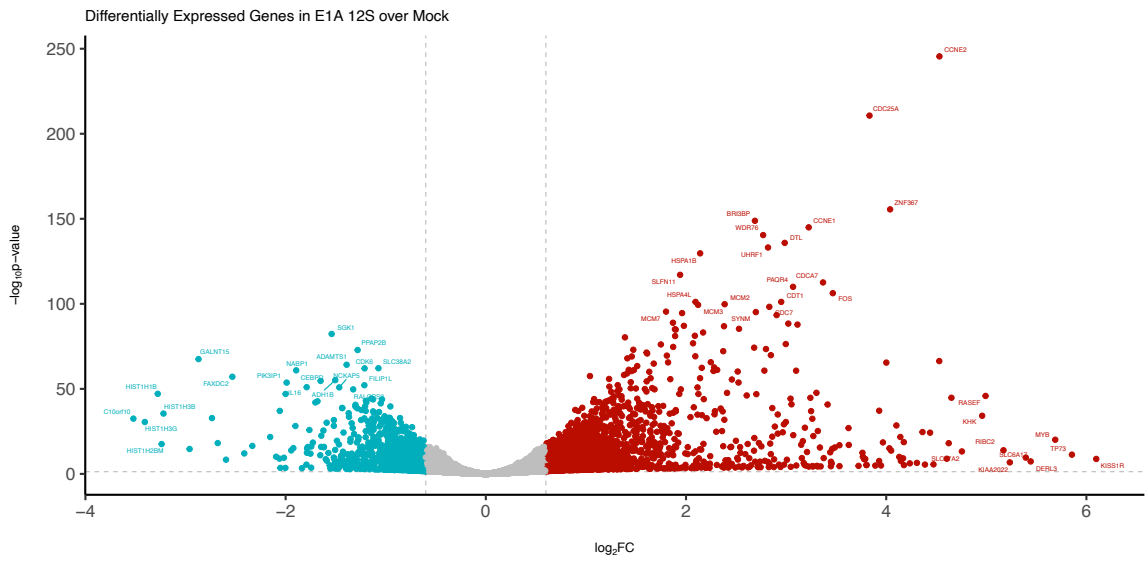
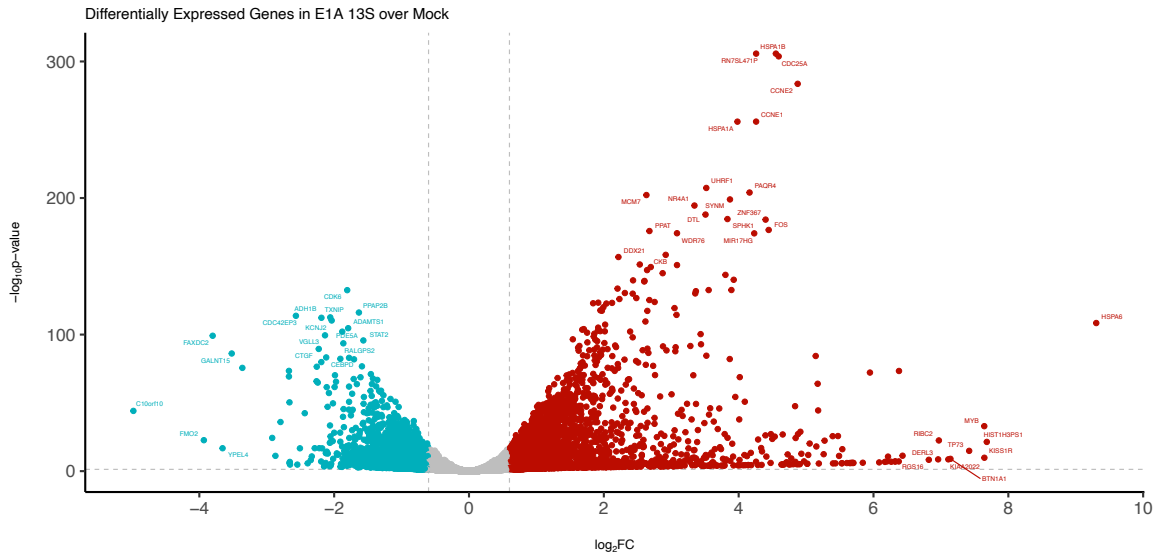


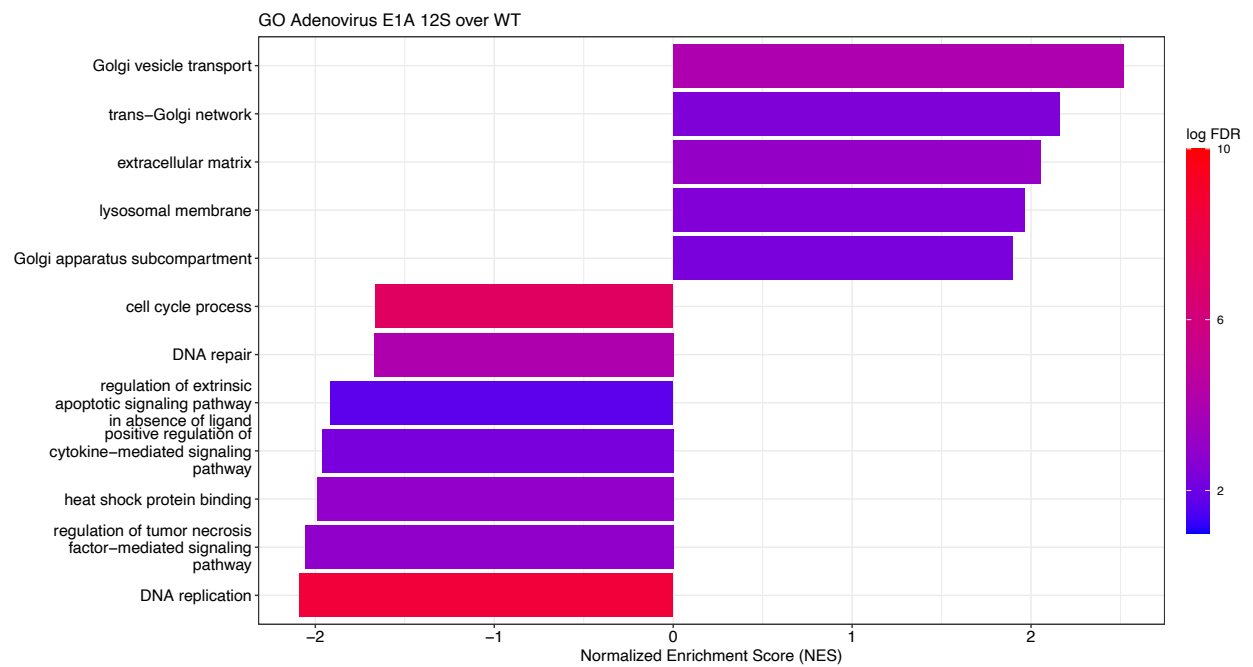
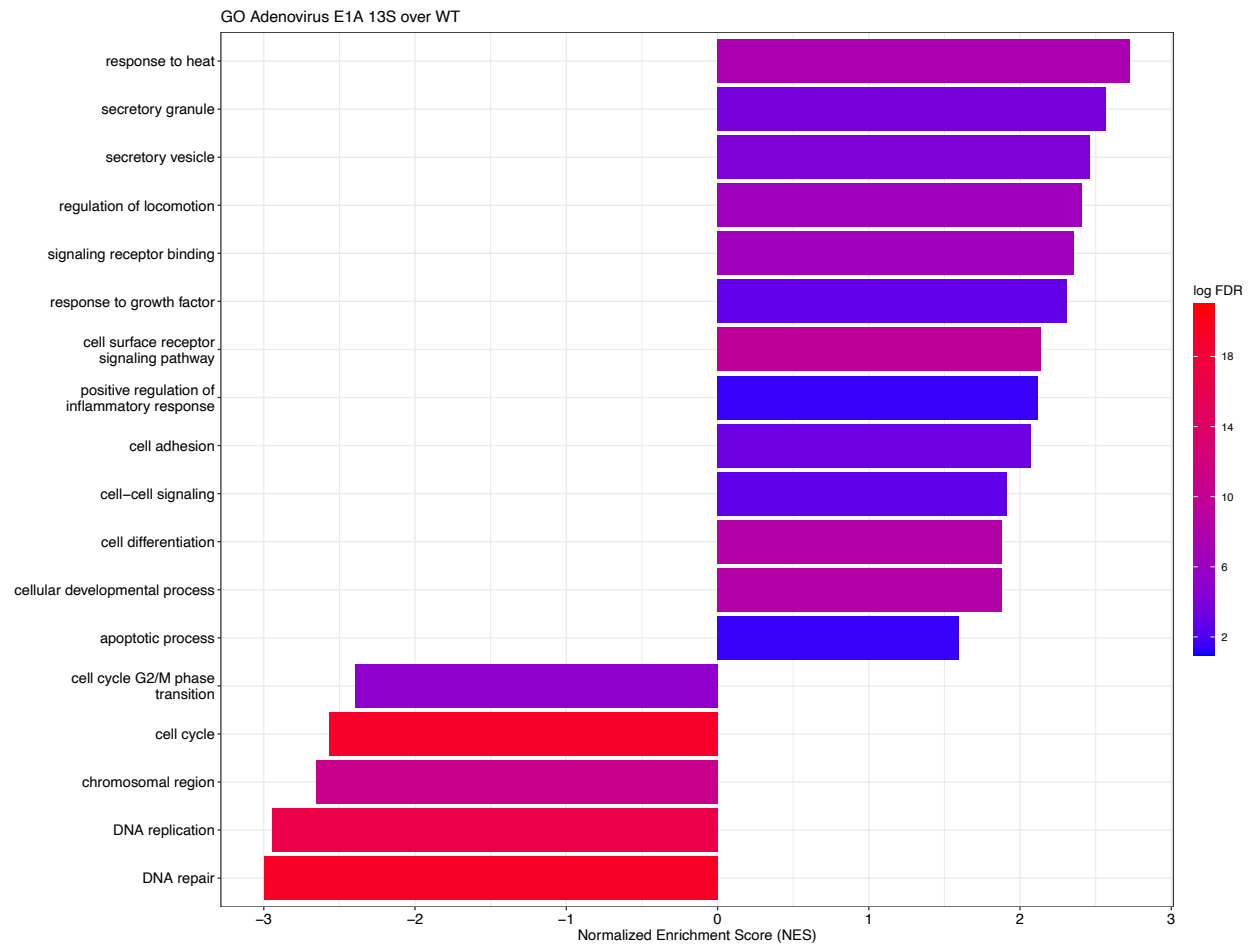
Figure 6: YCD mutant is highly distinct from other samples. RNA-Seq Heat Plot. The 50 most variable genes across all 5 samples. Heat map generated using pheatmap with hierarchical clustering (samples and transcripts) after using DESeq2 rlog transformation for scaling. Boxes in red are upregulated, blue are downregulated.

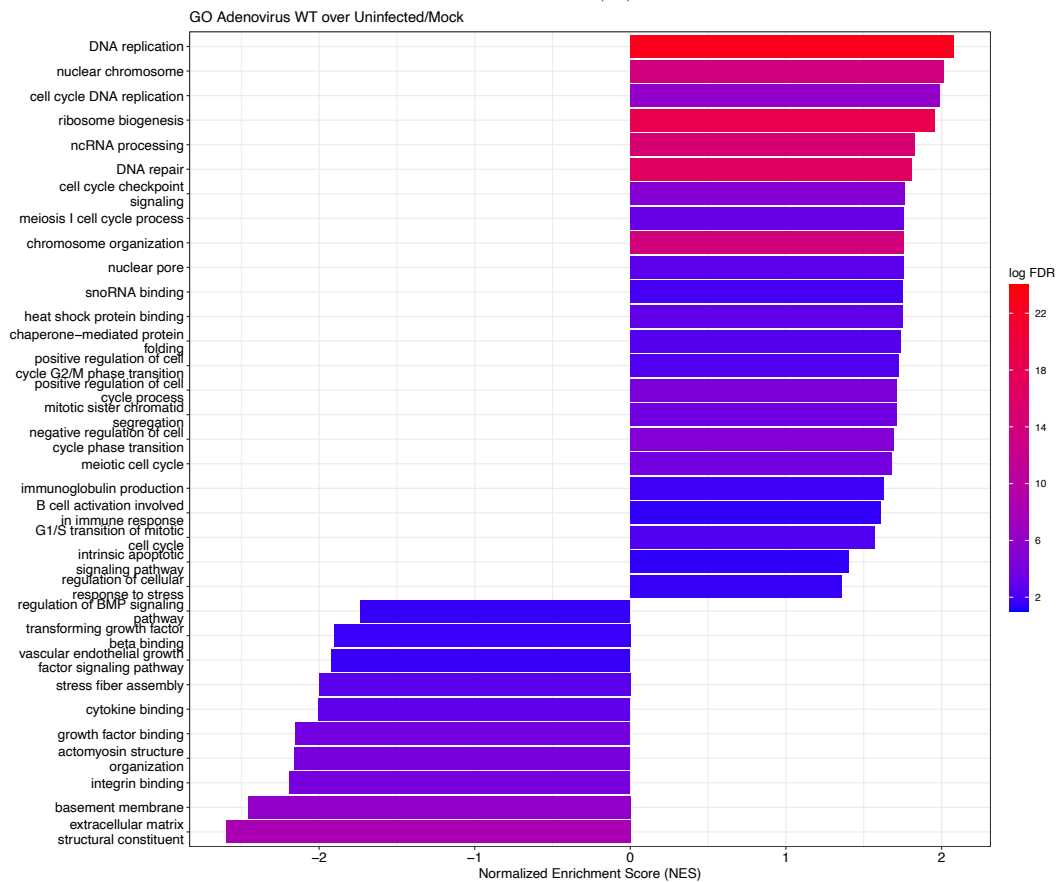
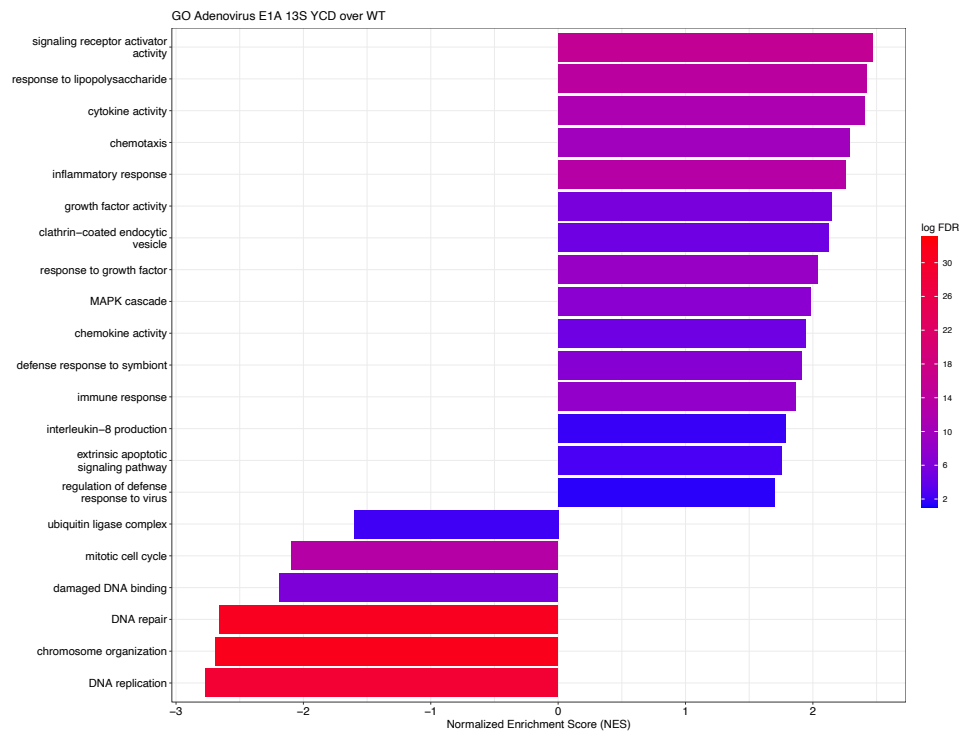


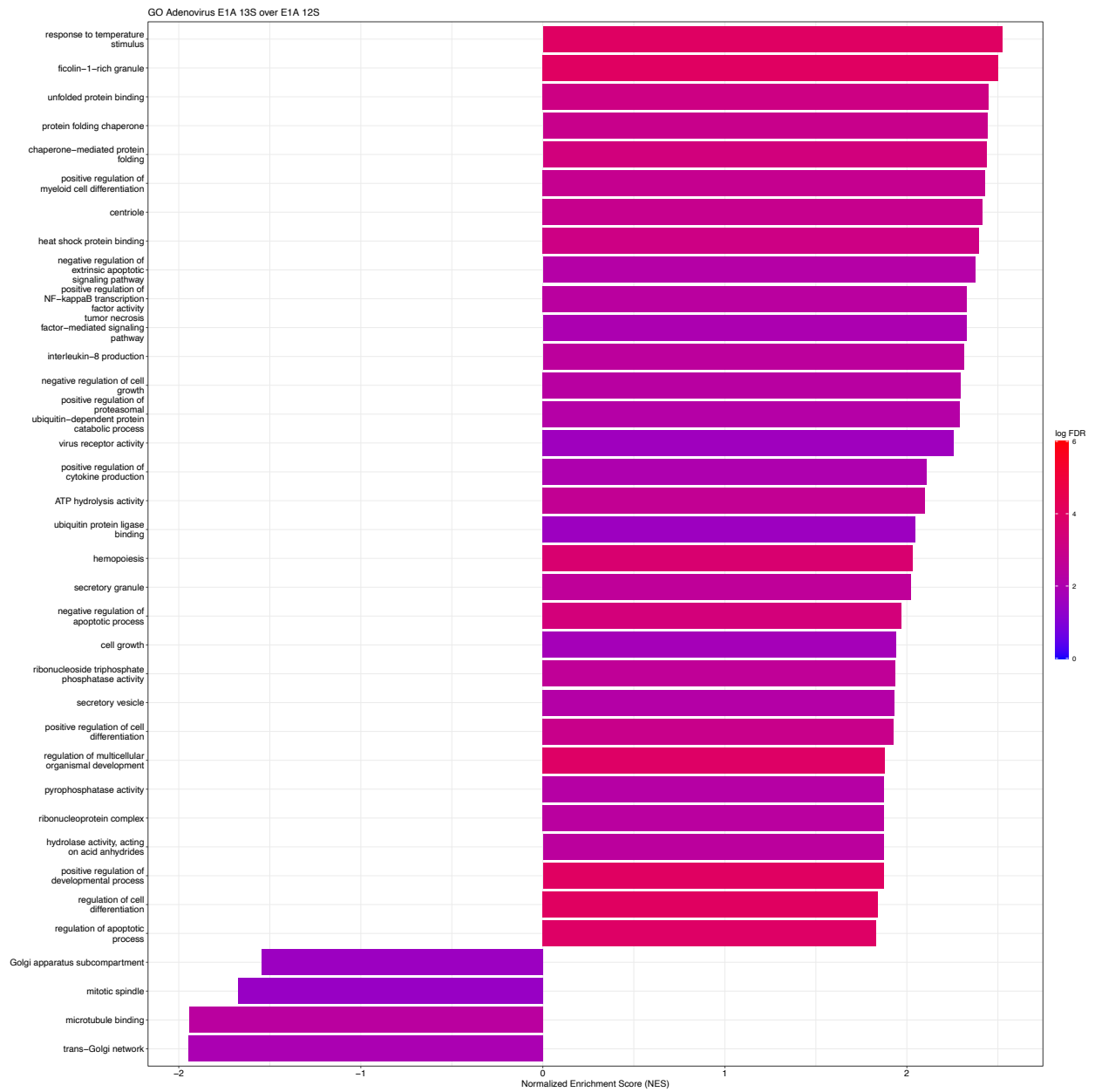


Functional Gene Analysis

To analyse the functional differences among the differentially expressed transcripts, gene set enrichment analysis (GSEA) was applied to the DESeq2 results. Because GSEA analysis uses a ranked list of genes (such as by fold change), it can compute the transcripts that contribute most to a GO term of interest. For example, in the YCD over WT comparison, the 5 transcripts that contribute most to the high enrichment of the cytokine response category are NAMPT, CXCL5, TNFSF15, INHBA, and BMP2. Upregulated GO terms for 13S include heat shock proteins, cytokines and other signaling molecules while DNA replication and cell cycle are downregulated (Figure 8). For 12S, Golgi associated transcripts and vesicles are upregulated while DNA replication, cell cycle, heat shock proteins, cytokines and other signaling molecules are downregulated. YCD in some ways looks like an intensification of the 13S GSEA terms, like 13S Heat shock proteins, cytokines and other signaling molecules are upregulated while DNA replication and cell cycle are downregulated, but YCD also has terms like immune response and defensive response to symbiont (i.e. virus). The other GSEA results are similar (Figure S7).







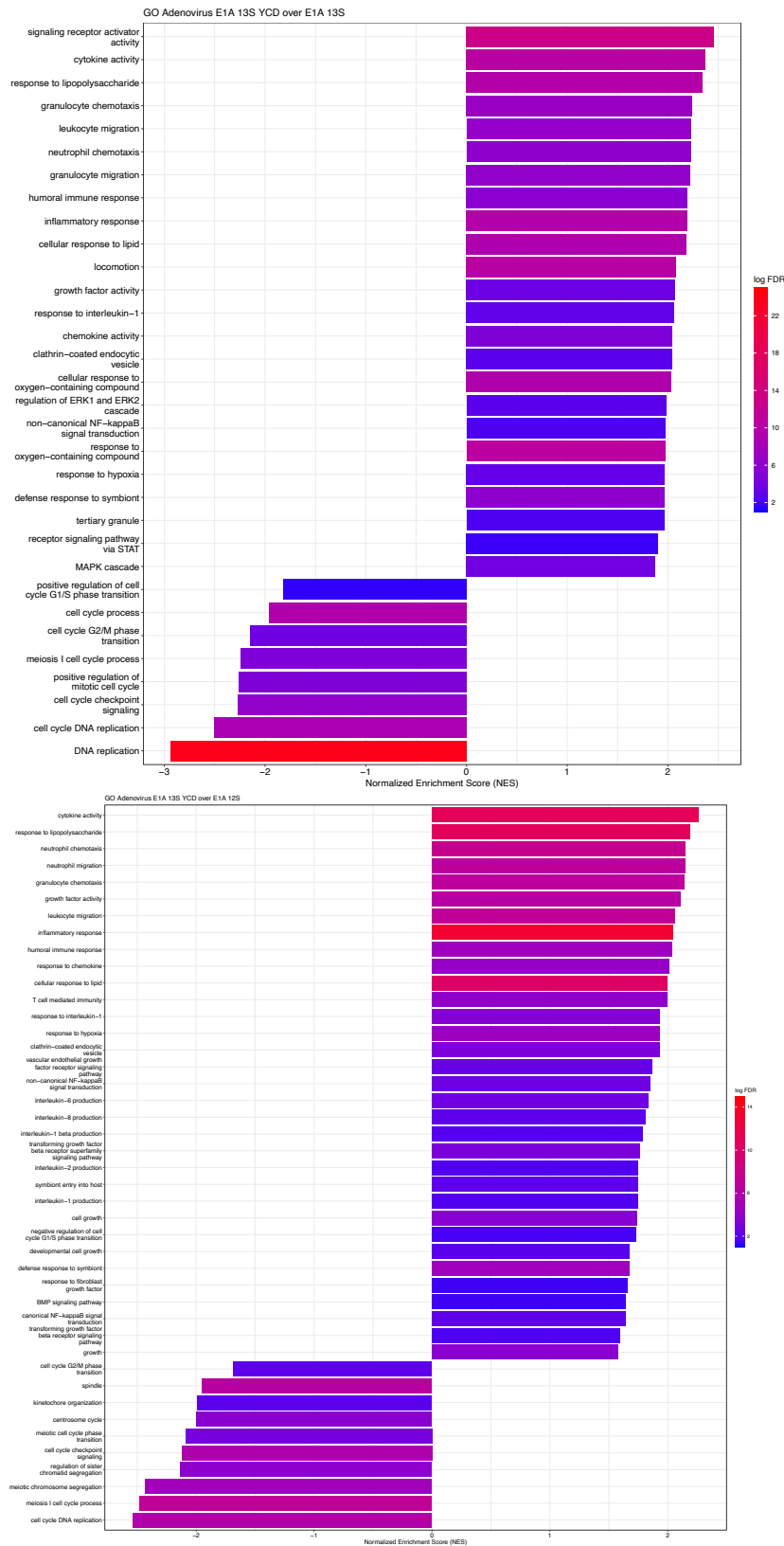


Figure 8: RNA-Seq GSEA GO Charts. Gene set enrichment analysis (GSEA) of differential expression comparisons. The x axis is the normalized expression score (NES), indicating

the extent to which the GO terms (y axis) are upregulated (NES > 0) or downregulated (NES < 0). Bars are colored based on logFDR (false discovery rate) (scale bar on side).

Clustering II

To further elucidate the impact of each of the strains, hierarchical clustering was performed on 2 sets of samples (Figures S9 and S10), and gene enrichment analysis with GO was performed to determine the function of the differentially expressed clusters (Tables S6 and S7). In the first set, comprising WT, 13S, and 12S there are 9 clusters, of size ranging from 283 to 1401 genes. WT has increased expression in clusters 5-7, associated with GO terms for the cell cycle, DNA repair, and translation, while clusters 1, 3, 4, and 9 had lower expression and are associated with cell morphogenesis, migration, stress response, inflammation and signaling. Similarly, 13S has increased protein folding, chaperone activity, signaling, and inflammatory response, but decreased cell cycle activity, DNA repair, and cytoskeleton binding. While 12S has increased cytoskeleton binding, inflammation, proliferation and MAPK cascade, but is decreased in some signaling, inflammation, chaperone activity, DNA replication and translation. In 12S, most cell cycle transcripts are down (Cluster 6, C7), but some in C8 are up, such as in centromere complex assembly (CENPI, OIP5, DLGAP5, CENPE, and CENPP). Finally, YCD has increased signaling, cytokines, growth factors, chaperone, ubiquitin activity, and apoptosis but decreased ion channel activity, cytoskeleton binding, cell cycle, and interstrand cross-link repair.

Upregulated				Downregulated			
Transcript	logFC	pvalue	padj	Transcript	logFC	pvalue	padj
<i>IL1B</i>	5.600	3.45E-190	4.48E-187	<i>FANCD2</i>	-1.338	1.04E-14	1.73E-13

<i>EREG</i>	4.445	4.06E-112	1.72E-109	<i>ERCC6</i>	-1.326	1.04E-16	2.09E-15
<i>INHBA</i>	3.829	0.00E+00	0	<i>KAT2B</i>	-1.310	3.02E-24	1.03E-22
<i>CCL2</i>	3.633	2.66E-176	2.46E-173	<i>TICRR</i>	-1.298	2.43E-13	3.49E-12
<i>IER3</i>	3.515	3.79E-200	5.28E-197	<i>CDC7</i>	-1.293	1.19E-27	4.90E-26
<i>IL1A</i>	3.304	9.29E-26	3.49E-24	<i>FBXO5</i>	-1.278	1.23E-22	3.77E-21
<i>MYC</i>	2.806	1.21E-134	7.13E-132	<i>BARD1</i>	-1.273	2.16E-27	8.82E-26
<i>SPRY2</i>	2.746	5.32E-152	3.83E-149	<i>TIPIN</i>	-1.271	1.01E-27	4.18E-26
<i>HSPA1A</i>	2.726	1.49E-128	7.84E-126	<i>FBXO6</i>	-1.218	1.33E-12	1.78E-11
<i>GJA1</i>	2.682	1.49E-132	8.54E-130	<i>DTX3L</i>	-1.205	4.49E-25	1.60E-23
<i>HSPA1B</i>	2.278	6.26E-177	6.10E-174	<i>CDC6</i>	-1.200	3.56E-06	2.09E-05
<i>SPHK1</i>	1.703	1.05E-59	1.41E-57	<i>KIF20B</i>	-1.194	7.58E-17	1.54E-15
<i>CACNB4</i>	1.639	3.89E-11	4.49E-10	<i>CDC45</i>	-1.179	4.82E-12	6.07E-11
<i>CDKN2B</i>	1.632	3.30E-61	4.55E-59	<i>CENPF</i>	-1.178	2.74E-18	6.18E-17
<i>DACT1</i>	1.604	3.06E-69	5.46E-67	<i>ATRIP</i>	-1.176	2.08E-18	4.79E-17
<i>RHOB</i>	1.335	3.28E-50	3.07E-48	<i>MCM2</i>	-1.158	6.11E-27	2.46E-25
<i>RRS1</i>	1.276	2.66E-23	8.49E-22	<i>MYBL2</i>	-1.140	4.11E-13	5.75E-12
<i>KLF4</i>	1.081	3.16E-30	1.52E-28	<i>INSR</i>	-1.133	3.57E-28	1.52E-26
<i>CCND1</i>	0.986	2.72E-38	1.79E-36	<i>CENPK</i>	-1.122	3.34E-12	4.27E-11
<i>CCNJ</i>	0.891	3.06E-25	1.11E-23	<i>MSH2</i>	-1.115	8.24E-36	5.03E-34
<i>BID</i>	0.836	3.03E-17	6.29E-16	<i>CDC25A</i>	-1.065	1.22E-26	4.79E-25
<i>DUSP1</i>	0.753	1.19E-18	2.79E-17	<i>MTBP</i>	-1.064	2.51E-10	2.65E-09
<i>PPME1</i>	0.676	3.66E-14	5.70E-13	<i>RHNO1</i>	-1.060	1.62E-18	3.78E-17
<i>BCCIP</i>	0.657	5.40E-14	8.30E-13	<i>BIRC5</i>	-1.058	2.00E-09	1.89E-08
<i>DDX3X</i>	0.635	2.50E-15	4.39E-14	<i>POLE</i>	-1.036	2.85E-25	1.03E-23
<i>SPDL1</i>	0.593	2.01E-09	1.90E-08	<i>CEP85</i>	-1.024	6.02E-16	1.12E-14

Table 5: Cell Cycle Genes Upregulated and Downregulated in YCD/12S, RNA-Seq of E1A 13S YCD mutant suggests possible transcripts involved in driving S phase without direct binding of pRb or E2F. Transcripts involved in EGF, ERK/MAPK, FSH, G1/S, and Cyclin E1

To look at transcripts associated with the cell cycle, we looked at cell cycle-relevant terms, specifically GSEA terms upregulated in 12S/UI (Table 5) and WT/UI including cell cycle, cell cycle processes, positive regulation of cell cycle and negative regulation of cell cycle.

Genes in terms of note were then cross referenced with log2FC and p values of YCD/12S to identify cell cycle-relevant transcripts that were differentially expressed for a total of 204 transcripts. GO terms (cell cycle, positive and negative regulation of cell cycle) from

overexpression analysis of YCD/12S were also analyzed similarly with a total of 413 transcripts in the cell cycle GO term and 105 transcripts in either the positive or negative regulation terms (Table S8).

3.2 Arglu1 Phenotype

Arglu1 overexpression has previously been shown to increase the cellular growth rate, DNA damage repair, and resistance to genotoxic drugs (Baccus et al. 2024). Actinomycin D is a genotoxic drug often used in chemotherapy. It works by binding to DNA at the transcription initiation complex and preventing elongation by RNA polymerase. (Sobell 1985). Previously Arglu1 overexpression was found to increase resistance to the double strand DNA-break inducing drug Bleomycin. Using a similar methodology, I found that Arglu1 increases resistance to Actinomycin D (Figure 9). I also performed flow cytometry with propidium iodide staining (Figure 10) to show that Arglu1 overexpressing cells have reduced duration of G1 and S-phase but slightly extended G2/M phase in the cell cycle compared to wild type cells. Together with proteomic analysis of HTA cells (Arglu1 overexpressing HT1080s), these two experiments were published in Baccus et al. (2024)

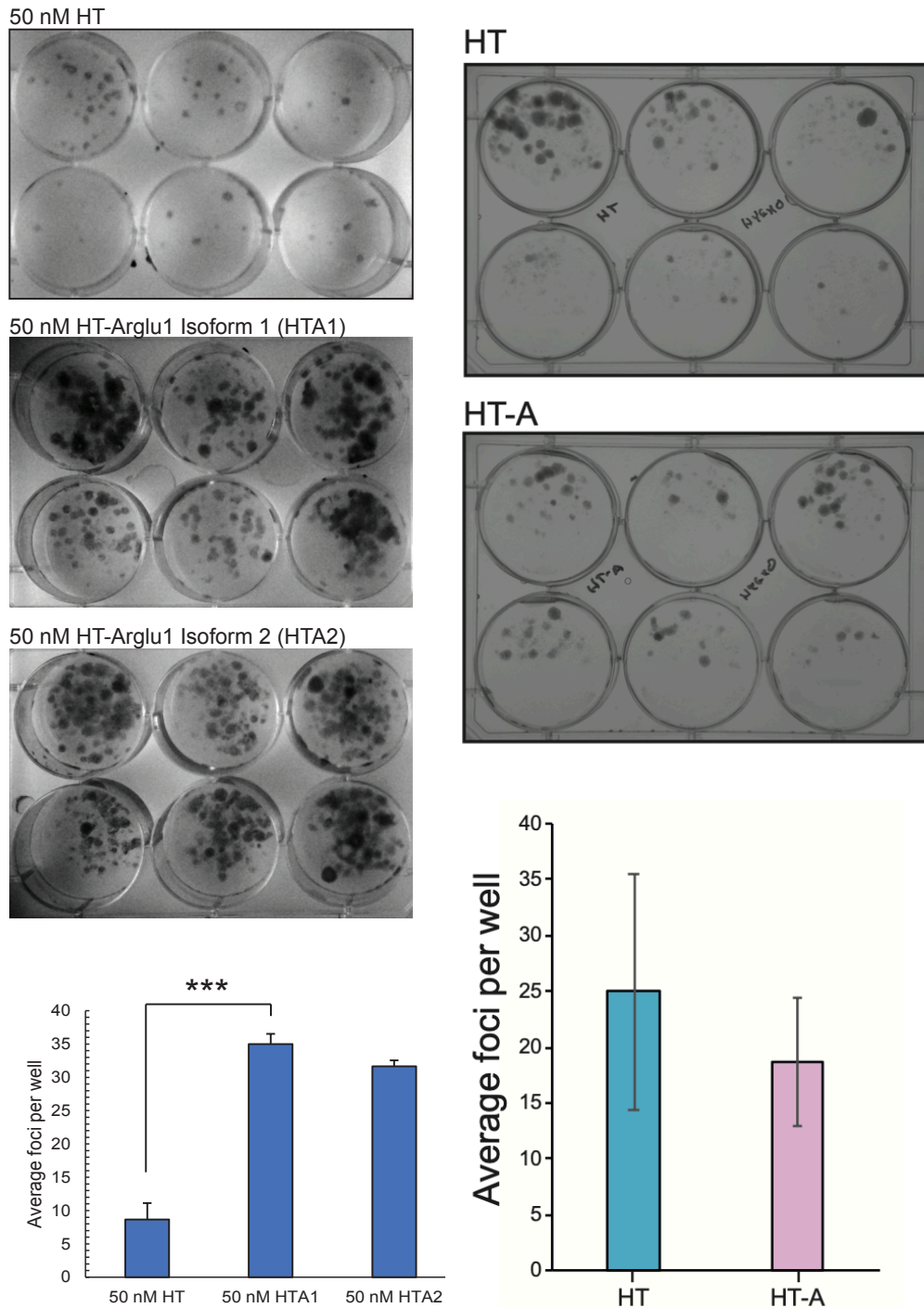


Figure 9: (Left) Actinomycin D Assay. 24h after plating, cells were treated with 50 nM Actinomycin D for 24h, then allowed to grow for 3 weeks before imaging. A negative control of Hygromycin applied to the same samples found no difference. (Right) 6 wells of HT or HT-A cells were plated at 200,000 cells/well in 6-well plates and treated with 350 $\mu\text{g/mL}$ of hygromycin B for 24 hours. The cells were then washed and allowed to grow for 21 days before colonies were visualized with crystal violet stain.

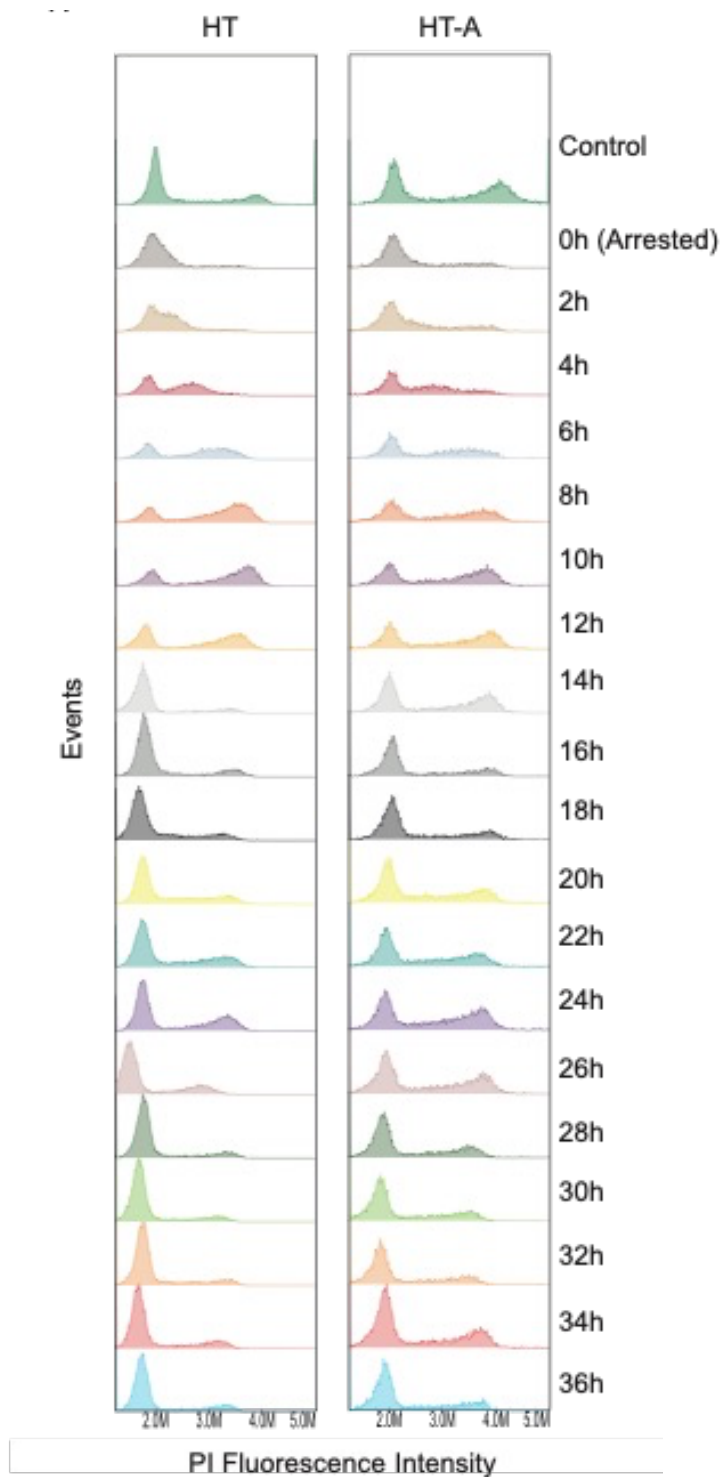


Figure 10: Flow Cytometry. DNA content profiles of HT or HT-A cells undergoing cell cycle analysis showing raw DNA content values for each time point and cell line. PI (propidium iodide)

3.3 Arglu1 Proteomics

HT-1080 cells overexpressing Arglu1 (HTA), along with HT cells as a control, were harvested for protein and total protein samples were sent to Southern Alberta Mass Spectrometry facility for shotgun 2D LC-MS/MS analysis. Proteins were identified using Mascot and label free quantification was performed by Scaffold. A total of 3543 different proteins were present in at least one of the samples, comprising 6 replicates each of HTA and HT cells. Then differential expression analysis was performed using proDA. Briefly, median normalization was performed to make the averages approximately equal (Figure S10 top), followed by the construction of a linear probabilistic dropout model that increases the statistical power by incorporating information about the missing values, and finally a Wald test was used to identify proteins that are significantly differentially expressed between the HTA and HT samples. Across all samples, 39% of values were missing (Figure S10 middle). Overall, 467 proteins were identified as differentially expressed, defined here as a p lower than 0.05 and a Log Fold Change (LogFC) greater than ± 1 . (Tables 6, S9). Additionally, 71 proteins had a p-value lower than 0.005 while 4 had a p-value lower than 0.0005. 66 proteins had a FC greater than 1 while 401 proteins had a FC less than -1.

Protein	pval	LogFC	t_stat	Std. Err.	avg intensity	n_approx	n_obs
<i>RPA2</i>	0.00028	5.79	5.44	1.06	10.52	12.00	12
<i>NOL10</i>	0.00035	-4.48	-5.30	0.85	-0.29	3.70	3
<i>PDS5B</i>	0.00035	-4.48	-5.30	0.85	-0.29	3.70	3
<i>TMEM109</i>	0.00038	-4.13	-5.24	0.79	-0.34	3.77	3
<i>CD2AP</i>	0.00056	4.30	4.97	0.87	5.96	12.00	12
<i>OPA1</i>	0.00065	-3.74	-4.87	0.77	-0.49	4.71	4
<i>RSL24D1</i>	0.00065	-4.27	-4.87	0.88	-0.46	3.89	3

<i>NFIC</i>	0.00072	-5.22	-4.80	1.09	-0.49	4.70	4
<i>PARK7</i>	0.00074	5.31	4.79	1.11	38.35	12.00	12
<i>CALB2</i>	0.00076	-5.76	-4.77	1.21	-0.48	3.70	3
<i>DOK1</i>	0.00076	-5.76	-4.77	1.21	-0.48	3.70	3
<i>TFAM</i>	0.00076	-3.86	-4.77	0.81	-0.52	3.74	3
<i>NDUFB10</i>	0.00086	-3.95	-4.69	0.84	-0.55	3.94	3
<i>NUP85</i>	0.00086	-3.95	-4.69	0.84	-0.55	3.94	3
<i>MKI67</i>	0.00086	-4.97	-4.69	1.06	-0.54	4.89	4
<i>EPB41L3</i>	0.00093	9.34	4.63	2.02	11.75	12.00	12
<i>VMP1</i>	0.00098	-4.78	-4.60	1.04	-0.58	3.96	3
<i>MRPL11</i>	0.00106	-3.59	-4.55	0.79	-0.61	3.94	3
<i>SDF2</i>	0.00106	-3.59	-4.55	0.79	-0.61	3.94	3
<i>NUP205</i>	0.00113	-4.44	-4.51	0.99	-0.63	4.02	3
<i>UQCRRS1</i>	0.00113	-4.44	-4.51	0.99	-0.63	4.02	3
<i>GATAD2B</i>	0.00123	-5.24	-4.45	1.18	-0.66	3.98	3
<i>HSD17B12</i>	0.00145	-3.69	-4.35	0.85	-0.69	4.02	3
<i>GOLGA4</i>	0.00154	-3.25	-4.31	0.75	-0.70	3.93	3
<i>IVD</i>	0.00154	-3.25	-4.31	0.75	-0.70	3.93	3

Table 6: Arglu1 Proteomics top proteins by p value, where n_obs (observed) is the number of samples that a specific protein has been detected in and n_approx is the number of replicates that the proDA algorithm has been able to extrapolate.

	ID	Description	setSize	NES	pvalue	qvalue	leading_edge_genes
BP	GO:0033554	cellular response to stress	68	2.231	0.00183	0.22341	RPA2, CD2AP, PARK7, PYCR2, ACTL6A, MNAT1, RAD23B, PDIA6, DHRS2, HTATSF1, PCNA, RPA1, ERO1A, NAE1, CDK2, MTREX, RCN3, RTCA, WDHD1, ATG7, CTH, BRD8, INHBA, NHEJ1, BAD, ATG5, CDKN2D, GSTM3, DNAJB2, TBC1D24, NAGLU, MLH3, RRAGA, BACH1, LOC102724560
BP	GO:0006281	DNA repair	24	2.129	0.00208	0.22341	RPA2, PARK7, ACTL6A, MNAT1, RAD23B, HTATSF1, PCNA, RPA1, CDK2, WDHD1, BRD8, NHEJ1, CDKN2D, MLH3, BACH1
BP	GO:0006974	DNA damage response	30	2.038	0.00539	0.22601	RPA2, PARK7, ACTL6A, MNAT1, RAD23B, HTATSF1, PCNA, RPA1, NAE1, CDK2, MTREX, WDHD1, BRD8, NHEJ1, BAD, CDKN2D, MLH3, BACH1
BP	GO:0006511	ubiquitin-dependent protein catabolic process	23	1.654	0.02075	0.35396	PARK7, RAD23B, ELOB, PSMA5, PSMD7, NAE1, CDK2, PPP2R5C, CHMP2B, DNAJB2, NAGLU, PSMB10, HECTD3, RNF25, CCDC22
BP	GO:0006950	response to stress	110	1.428	0.04728	0.48974	RPA2, CD2AP, PARK7, PDLIM1, CDC42EP4, PYCR2, ACTL6A, MNAT1, RAD23B, PDIA6, DHRS2, HTATSF1, PCNA, RPA1, ERO1A, NAE1, CDK2, FKBP10, MTREX, RCN3, RTCA, ITGA5, WDHD1, KRT16, ATG7, TIMP1, CTH, BRD8, INHBA, NHEJ1, BAD, ATG5, CHMP2B, PTPN6, CDKN2D, GSTM3, DNAJB2, TRIM4, CARMIL2, TBC1D24, NAGLU, MLH3, RRAGA, BACH1

Table 7: Arglu1 Proteomics GSEA terms of note

Table 8: Arglu1 Proteomics g:Profiler terms of note

term_name	term_id	-log10 padj	intersections
GO:BP cell cycle	GO:0007049	4.994	MNAT1, RB1CC1, MYO16, PRKCG3, PPP2R5C, THOC1, EPB41L2, NDC80, PDS5B, CHMP2B, CDC34, CDC37, RHEB, GBR1, PTPN6, BRD8, PPP2CA, NR3C1, HES1, PRM13, TRIM32, MLH3, TEX14, INHBA, CIT, CENPK, CDK2, TUBGCP3, PCID2, CHTF18, CDKN2D, PAK4, HAUS8, LLGL1, RPA1, PCNA, NASP, CDK4, TBRG4, ACT16A, TUBB2A, NUSAP1, WDR12, FANCI, AKT1, POGZ, DIS3L2, CCNA2, MKI67, CENPJ, HAUS1, ING1, PRKCA, NAE1, PCNT, PBRM1, WASHOS, IGF2, TEF2, CHD3, HCF1, SFN, FOXG1, CDC26, PDEAD1P, GAK, EXOC7, HTT, CD2AP, WDH1, SGOA1
GO:BP developmental process	GO:0032302	2.386	BAD, ITGA3, MNAT1, RB1CC1, MYO16, DSG2, NEXMIF, ATG5, YMP1, COL17A1, KDM4A, CBRF, MARK2, FRF, EIF4G3, XAB2, THOC1, EPB41L3, APOB, AAMD, C, RAB18, DHRS2, HM13, LAMA1, EMD, TIMP1, CLU6, HERC1, RPK2, SLC39A14, PPP1R13L, PLD3, NUMBL, ELL, RHEB, NCST1, PDLM1, CCDC47, MAGLU, VAT1, OCLAD1, MICH2, NUP107, PTPN6, PAK1IP1, PPP2CA, NR3C1, HES1, GNB4, ERRE1, CTH, ACT18, ECE1, CNN3, RAD23B, TRIM2, INHBA, CIT, COL10A1, ACSL3, RREB1, NUP85,
GO:BP DNA-templated transcription elongation	GO:0006354	2.305	ELOA, THOC1, H1ATS1, ELOB, ELL, PCID2, ZC3H4, MED10, MED8, PBRM1, WDR82, HMGNT, ELOA2
GO:BP mitotic cell cycle	GO:0000278	2.292	MNAT1, MYO16, PPP2R5C, NDC80, CHMP2B, CDC34, GBR1, PTPN6, PPP2CA, HES1, RPA2, TEX14, INHBA, CIT, CENPK, CDK2, TUBGCP3, PCID2, CDKN2D, NUSAP1, AKT1, DIS3L2, CCNA2, MKI67, CENPJ, PRKCA, NAE1, PCNT, PBRM1, IGF2, FOXG1, CDC26, EXOC7, HTT, WDH1
GO:BP cell death	GO:0008219	2.190	BAD, MNAT1, RB1CC1, ATG5, KDM4A, PPP2R5C, THOC1, EPB41L3, CDC34, DHRS2, TIMP1, STX4, RPK2, PPP1R13L, UBR1, TIMM50, MICH2, TMEM109, PTPN6, DAP, BRD8, PPP2CA, NR3C1, ERRE1, PARK7, CTH, FASTKD2, HS1BP3, TRIM32, INHBA, CIT, PCID2, CDKN2D, PAK4, UBE4B, ACT16A, ETV6, AKT1, ARL6IP5, DLG5, SLC25A27, ING1, TOMM70, PRKCA, RRA6A, NAE1, RRP1B, ITGA5, NCSTN, SLC25A6, ARL6IP1, NPTX1, SFN, ALDH1A3, C8ORF120, PTPN1, MYO18A, IGF2R, HTT, ATG7, ERO1A, QORCH1, OPA1, MLLT1, NDUFS3, TXNDC5, GBA
GO:BP apoptotic process	GO:0006915	2.158	BAD, MNAT1, RB1CC1, ATG5, KDM4A, PPP2R5C, THOC1, EPB41L3, CDC34, DHRS2, TIMP1, STX4, RPK2, PPP1R13L, UBR1, TIMM50, MICH2, TMEM109, PTPN6, DAP, BRD8, PPP2CA, NR3C1, ERRE1, PARK7, CTH, FASTKD2, HS1BP3, TRIM32, INHBA, CIT, PCID2, CDKN2D, PAK4, UBE4B, ACT16A, ETV6, AKT1, ARL6IP5, DLG5, SLC25A27, ING1, TOMM70, PRKCA, RRA6A, NAE1, RRP1B, ITGA5, NCSTN, SLC25A6, ARL6IP1, NPTX1, SFN, ALDH1A3, C8ORF120, PTPN1, MYO18A, IGF2R, HTT, ATG7, ERO1A, QORCH1, OPA1, MLLT1, NDUFS3, TXNDC5, GBA
GO:BP G1/S transition of mitotic cell cycle	GO:0000082	2.023	MNAT1, MYO16, PPP2R5C, CDC34, PTPN6, PPP2CA, RPA2, INHBA, CDK2, CDKN2D, CDK4, ACT16A, AKT1, CCNA2, CENPJ, PBRM1
GO:BP regulation of mitotic cell cycle	GO:0007246	1.957	MNAT1, MYO16, PPP2R5C, NDC80, GBR1, PTPN6, PPP2CA, HES1, RPA2, TEX14, INHBA, CDK2, PCID2, CDKN2D, CDK4, ACT16A, NUSAP1, AKT1, MKI67, CENPJ, PRKCA, NAE1, PBRM1, IGF2, FOXG1, CDC26
GO:BP regulation of G1/S transition of mitotic cell cycle	GO:2000045	1.466	MNAT1, MYO16, PPP2R5C, PTPN6, PPP2CA, RPA2, INHBA, CDK2, CDKN2D, ACT16A, AKT1, CENPJ, PBRM1
GO:BP cell cycle G1/S phase transition	GO:0044843	1.380	MNAT1, MYO16, PPP2R5C, CDC34, PTPN6, PPP2CA, RPA2, INHBA, CDK2, CDKN2D, CDK4, ACT16A, AKT1, CCNA2, CENPJ, PBRM1
REAC: Transcriptional Regulation by TP53	REAC:R-HSA-3709889	2.606	ELOA, MNAT1, PPP2R5C, RABGGT, A, ELOB, PPP1R13L, ELL, RHEB, PPP2CA, RPA2, CDK2, RPA1, PCNA, TACOT1, RABGGT, FANCI, AKT1, GATAO2B, GNOT9, CCNA2, CENPJ, RRA6A, CHD3, SFN, EHM1, ELOA2
REAC: Cell Cycle, Mitotic	REAC:R-HSA-69278	2.592	MNAT1, HMMR, PPP2R5C, NDC80, PDS5B, CHMP2B, EMD, PSMO7, CEP41, NUP107, NUP155, PPP2CA, RPA2, CENPK, CDK2, NUP85, TUBGCP3, CDKN2D, HAUS8, RPA1, PCNA, CDK4, TUBB2A, AKT1, PSMA5, CCNA2, CENPJ, HAUS1, PRKCA, NUP205, PCNT, LEMD2, CDC26, PSM10

Proteomics PCA and Functional Gene Analysis

Principal component analysis found that the HTA and HT samples clustered separately (Figure 11). Interestingly, unlike the RNA-seq PCA there is high variation within the two sample types although the samples are still in distinct clusters. This could be because of batch effects between the two sets of samples from 2021 and 2023 (Voß et al. 2022) or because cancer cells have more variation in gene expression compared to normal cells (Zhao et al. 2020). Some of the most variable genes between samples include ANPEP, PCNA, NASP, and KHSRP (Figure 12). Other genes like RPA2, NOL10, and PDS5B had lower fold changes but among the highest p values (Figure 13). Finally, GSEA found that categories including cellular response to stress, DNA repair, DNA damage response, and ubiquitin-dependent protein catabolic process were upregulated (Figure 14). The proteins most responsible for DNA damage response being upregulated are RPA2, PARK7, ACTL6A, MNAT1, RAD23B, HTATSF1, PCNA, RPA1, NAE1, CDK2, MTREX, WDHD1, BRD8, NHEJ1, BAD, CDKN2D, MLH3, and BACH1. (Tables 7, S10). Overexpression analysis (OEA) also identified GO terms including cell cycle, developmental process, cell death, apoptotic process, G1/S transition of mitotic cell cycle, and Transcriptional Regulation of PT53 (Figure S11, Tables 8 and S11). Arglu1 proteins that are particularly important for the cell cycle, DDR, or transcription are noted in (Table 9) (see Discussion).

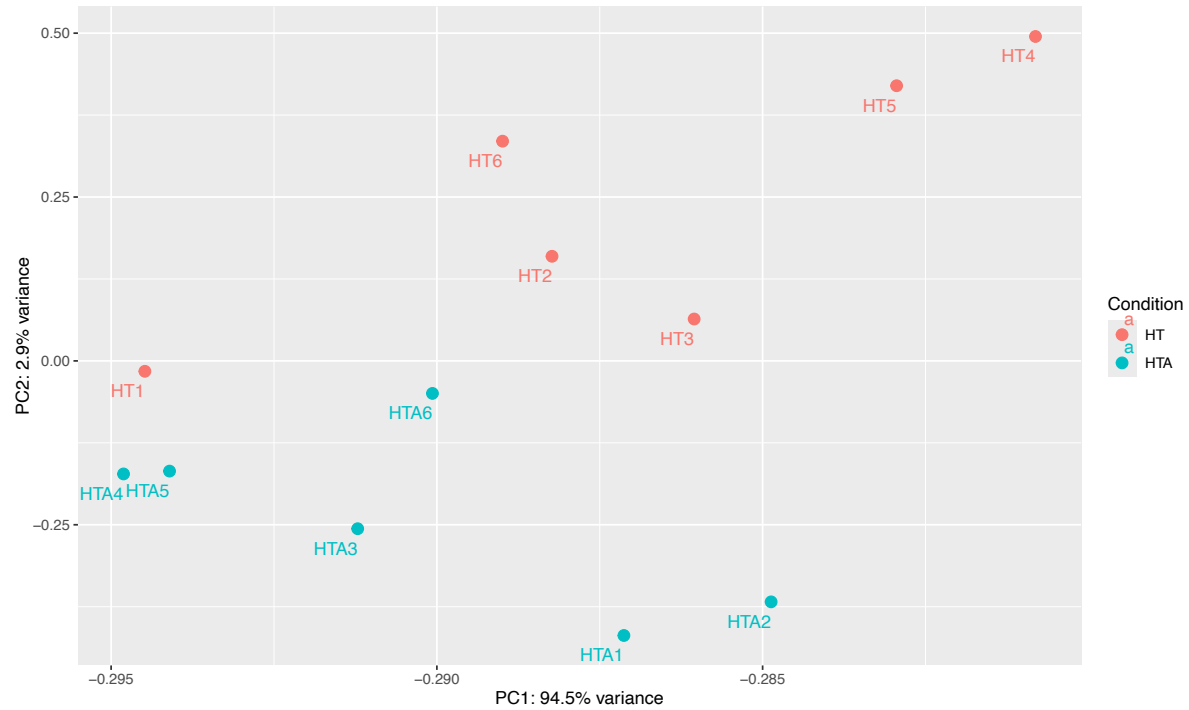


Figure 11: Arglu1 Proteomics PCA, the top 50 proteins by pvalue (about 10% of the differentially expressed proteins). Principal Component Analysis plot generated using R with zscore normalization for scaling and displaying the first two principal coordinates. Individual replicates are colored according to group (HTA or HT).

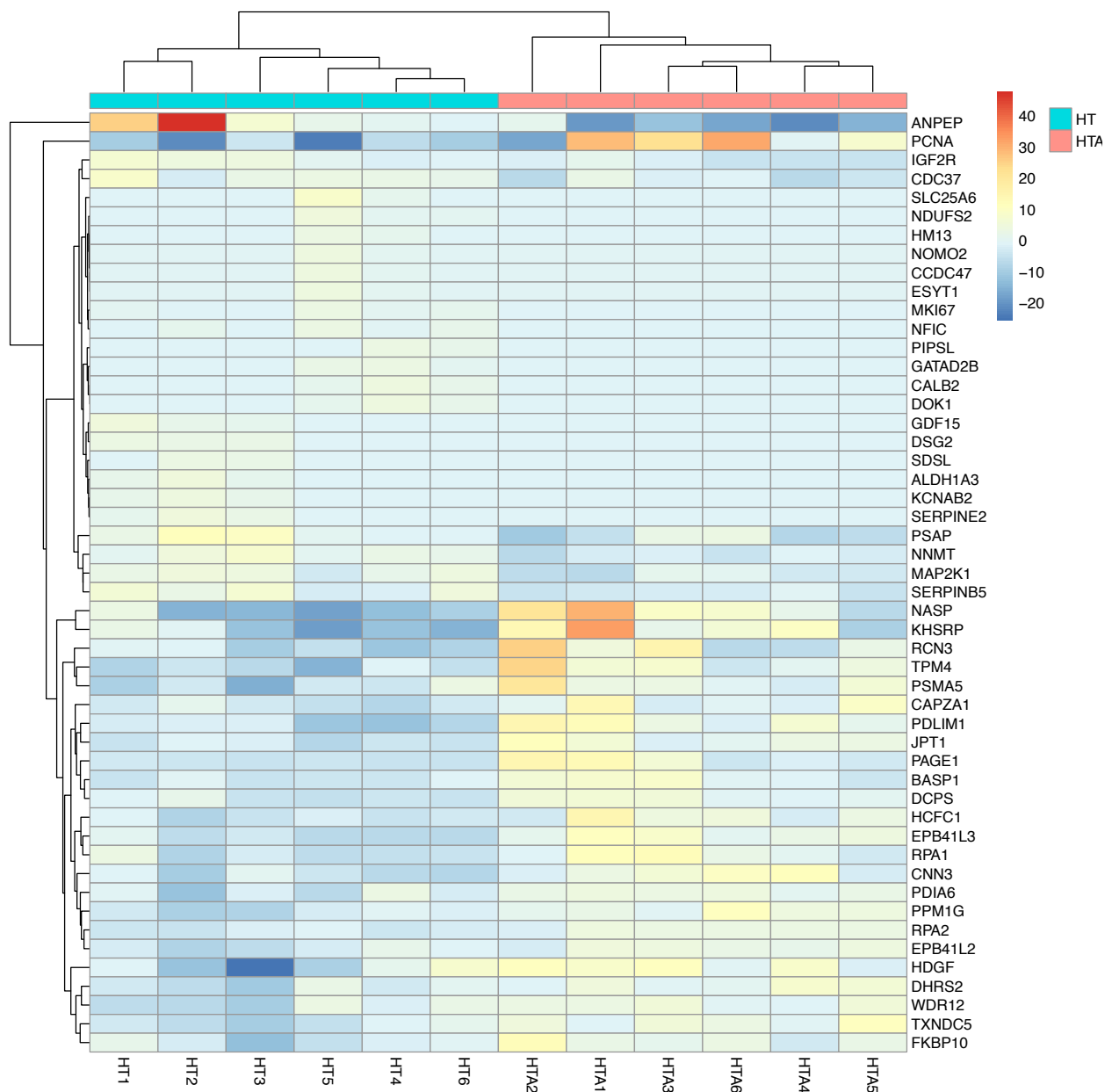


Figure 12: Arglu1 Proteomics Top Fold Change Heat Map The 50 most variable genes across all 6 replicates of the 2 samples HT and HTA . Heat map generated using pheatmap with hierarchical clustering (samples and transcripts) after normalization. Boxes in red are upregulated, blue are downregulated.

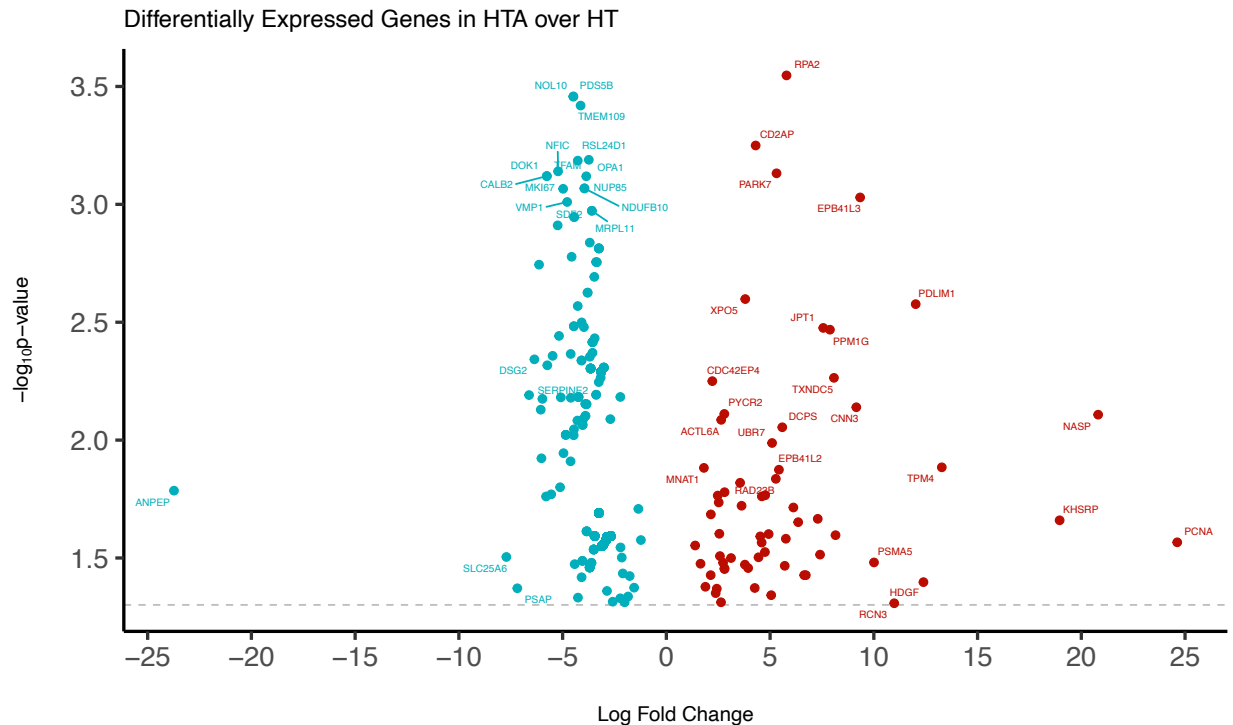


Figure 13: Arglu1 Proteomics Volcano Plot. Red indicates upregulated (FC > 1, pvalue < 0.05), blue indicates downregulated (FC < -1, pvalue < 0.05). Volcano plot generated using ggplot2.

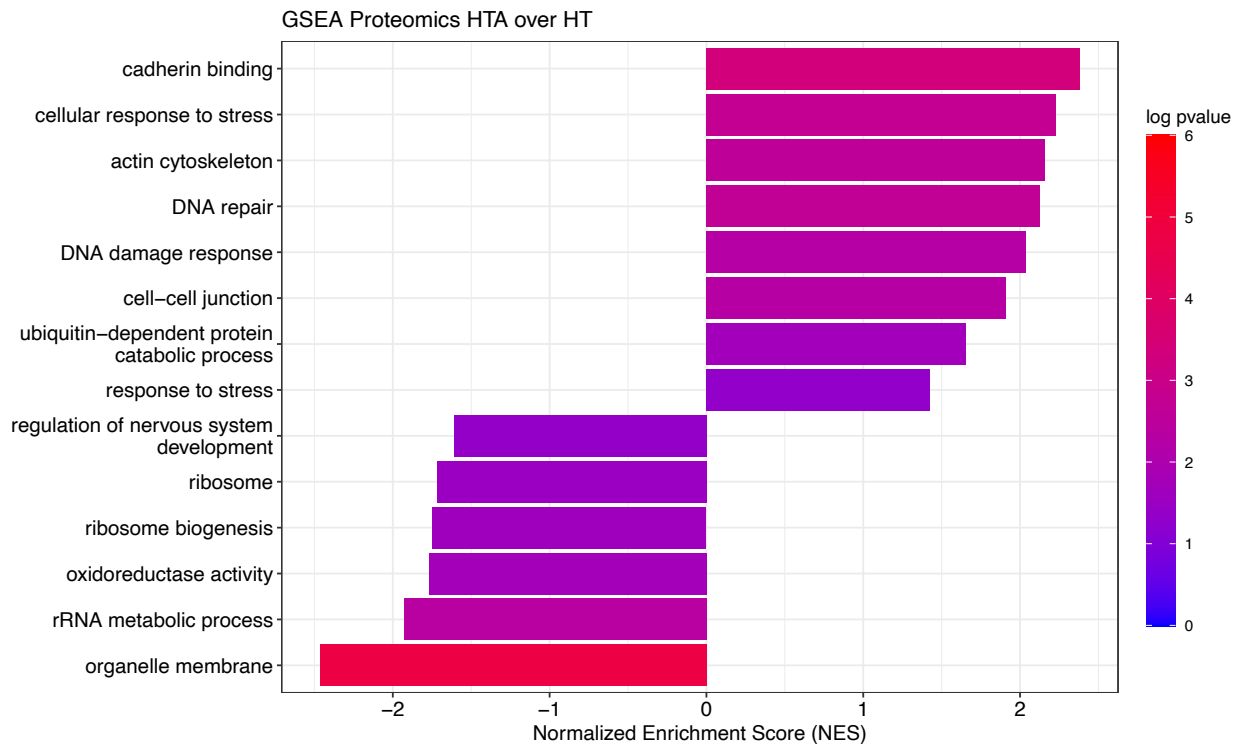


Figure 14: Arglu1 Proteomics GSEA Bar Chart. Gene set enrichment analysis (GSEA) of differential expression comparisons. The x axis is the normalized expression score (NES),

indicating the extent to which the GO terms (y axis) are upregulated (NES > 0) or downregulated (NES < 0). Bars are colored based on logFDR (false discovery rate) (scale bar on side).

Proteins	pval	logFC
<i>PCNA</i>	2.71E-02	24.624
<i>PPM1G</i>	0.0034	7.886
<i>RPA2</i>	2.84E-04	5.792
<i>EPB41L2</i>	0.01336	5.427
<i>PARK7</i>	7.39E-04	5.311
<i>RAD23B</i>	1.46E-02	5.277
<i>HTATSF1</i>	2.56E-02	4.527
<i>NAE1</i>	3.49E-02	3.942
<i>ACTL6A</i>	8.20E-03	2.644
<i>ELOB</i>	1.72E-02	2.476
<i>MNAT1</i>	1.31E-02	1.811
<i>ELL</i>	2.55E-02	-2.683
<i>MED8</i>	2.55E-02	-2.683
<i>ELOA</i>	2.55E-02	-2.683
<i>PBRM1</i>	2.55E-02	-2.683
<i>THOC1</i>	2.55E-02	-2.683
<i>ELOA2</i>	2.58E-02	-2.866
<i>CCDC22</i>	2.58E-02	-2.866
<i>HECTD3</i>	2.66E-02	-2.914
<i>RNF25</i>	2.66E-02	-2.914
<i>MLH3</i>	2.77E-02	-3.029
<i>PPP2R5C</i>	2.82E-02	-3.111
<i>MYO16</i>	5.12E-03	-3.146
<i>CDC34</i>	5.12E-03	-3.146
<i>CENPJ</i>	5.12E-03	-3.146
<i>NHEJ1</i>	2.91E-02	-3.499
<i>PCID2</i>	4.43E-03	-3.686
<i>CCNA2</i>	3.32E-03	-3.972
<i>PPP2CA</i>	8.62E-03	-4.035
<i>MED10</i>	3.36E-02	-4.418
<i>PDS5B</i>	3.49E-04	-4.482
<i>CDC37</i>	1.73E-02	-5.798

Table 9: Arglu1 Notable Proteins

3.4 Arglu1 Deletion Mutant Plasmids and Stable Cell Lines

To further elucidate the mechanisms of Arglu1 I started functionally characterizing Arglu1 by making 29 mutated plasmids of Arglu1 missing 10 to 50 amino acids across the entire sequence of Arglu1 as well as 21 stable cell lines overexpressing these plasmids (Table S8). However, the effects on cellular growth and binding to E1A, Brd4, and Jmjd6 were inconclusive.

3.5 Arglu1 ChIP-Seq

ChIPs were performed by Peter Pelka on 4 samples (1 IgG control in HT cells, 1 on HT cells pulling down RNA Pol II, 2 on HTA cells pulling down RNA Pol II and Arglu1) and sequenced at the University of Toronto. I then edited, aligned the reads to the Hg38 reference genome and normalized and visualized the reads. Normalization/Peak Calling with Homer produced 6552 peaks for Arglu1, 19, 532 peaks for HTA RNA Pol II, and 19, 502 peaks for HT RNA Pol II (Tables 10 and 11). Homer was also used to search for super enhancers (the algorithm uses regions of the genome that contain several enhancers in close proximity as an operational definition). Two super enhancers were found to interact with HTA RNA Pol II, IFL3/IGLF3 (Figure 16) and PANDAR (not shown). ILF3 is associated with IL-2 and mTOR (Yan et al. 2023). PANDAR is a lncRNA that binds p53 and contributes to tumourigenesis and progression of cancer and is a potential therapeutic target (Li et al. 2017).

	Arglu1/IgG	HTA RNA Pol II/IgG	HT RNA Pol II /IgG	Arglu1/HTA Pol II	HTA/HT RNA Pol II
Peak Calling	6552	19532	19502	21	4
Super Enhancers	0	2			
Motifs Enrichment (de novo)	40	40	38		
Motifs (known)	75	147	160		

Table 10: ChIP-Seq Sample Peaks/Motifs Normalization/Peak Calling and Motif Enrichment performed using Homer.

Gene Name	Peak Score	Log 2 FC	pval	padj	Annotation	Distance to TSS	Nearest Promoter ID	Gene Description	Gene Type	Chromosome
RNA5-8SN4	996.7	3.708	1.544E-35	2.020E-31	Intergenic	1191	NR_146120	RNA, 5.8S ribosomal N4	rRNA	chr22_K1270733v1_random
RNA28SN1	455.1	4.224	5.631E-32	2.456E-28	promoter-TSS (NR_003287).3	-135	NR_145822	RNA, 28S ribosomal N1	rRNA	chr22_K1270733v1_random
MIR10396B	352.6	4.048	4.953E-32	2.456E-28	promoter-TSS (NR_162111).3	-613	NR_162111	microRNA 10396b	ncRNA	chr22_K1270733v1_random
RNA5-8SN2	979.6	3.931	1.770E-31	5.791E-28	promoter-TSS (NR_003285)	-725	NR_146147	RNA, 5.8S ribosomal N2	rRNA	chr22_K1270733v1_random
MIR10396B	357.7	3.467	2.500E-29	6.182E-26	Intergenic	-1550	NR_162111	microRNA 10396b	ncRNA	chr22_K1270733v1_random
RNA28SN4	1099.9	3.385	2.835E-29	6.182E-26	TTS (NR_046235).2	5544	NR_146118	RNA, 28S ribosomal N4	rRNA	chr22_K1270733v1_random
MIR10396B	291.7	4.602	3.504E-29	6.550E-26	promoter-TSS (NR_003286)	-50	NR_162111	microRNA 10396b	ncRNA	chr22_K1270733v1_random
RNA5-8SN2	267	4.056	1.548E-27	2.531E-24	promoter-TSS (NR_145821).5	-328	NR_146147	RNA, 5.8S ribosomal N2	rRNA	chr22_K1270733v1_random
MIR10396B	317.9	3.173	3.601E-24	5.234E-21	promoter-TSS (NR_162111).4	-486	NR_162111	microRNA 10396b	ncRNA	chr22_K1270733v1_random
RNA28SN5	196.1	3.890	3.162E-22	4.137E-19	TTS (NR_046235).3	5512	NR_003287	RNA, 28S ribosomal N5	rRNA	chrUn_GL000220v1
RNA5-8SN2	168.8	4.004	6.248E-22	7.431E-19	promoter-TSS (NR_145821).5	-891	NR_146147	RNA, 5.8S ribosomal N2	rRNA	chr22_K1270733v1_random
RNA28SN4	988.2	3.511	2.160E-21	2.355E-18	non-coding (NR_046235, exon 1 of 1).2	4915	NR_146118	RNA, 28S ribosomal N4	rRNA	chr22_K1270733v1_random
RNA28SN5	162.5	4.653	1.634E-19	1.645E-16	promoter-TSS (NR_146148)	-546	NR_003287	RNA, 28S ribosomal N5	rRNA	chr21
MIR10396A	118	4.611	2.936E-17	2.744E-14	TTS (NR_037421).2	524	NR_162105	microRNA 10396a	ncRNA	chr21
RNA28SN2	160.9	2.771	3.802E-15	3.316E-12	TTS (NR_146144)	6017	NR_146148	RNA, 28S ribosomal N2	rRNA	chr21
MIR3648-1	85.7	4.279	3.659E-14	2.992E-11	promoter-TSS (NR_003286).3	513	NR_037421	microRNA 3648-1	ncRNA	chr21
RNA28SN3	95.5	5.443	8.949E-13	6.887E-10	promoter-TSS (NR_146154)	-189	NR_146154	RNA, 28S ribosomal N3	rRNA	chr21
MIR3648-1	35.6	3.735	1.160E-07	8.432E-05	TTS (NR_037421)	130	NR_037421	microRNA 3648-1	ncRNA	chr21
IRF2BP2	11.5	5.865	1.314E-04	3.029E-02	exon (NM_182972, exon 1 of 2)	952	NM_001077397	interferon regulatory factor 2 binding protein 2	protein	chr1

									coding	
MGC70870	11	5.854	4.520E-03	3.029E-02	Intergenic	-82454	NR_003682	C-terminal binding protein 2 pseudogene	pseudogene	chr17_GL000205v2_random
PHLDA1	11.5	5.854	1.366E-04	3.029E-02	intron (NM_007350, intron 1 of 1)	1540	NM_007350	pleckstrin homology like domain family A member 1	protein-coding	chr12
SH3BP5	10.8	5.805	2.000E-04	3.029E-02	5' UTR (NM_004844, exon 1 of 9)	176	NM_004844	SH3 domain binding protein 5	protein-coding	chr3
TRIO	11.2	5.798	1.646E-04	3.029E-02	intron (NM_007118, intron 1 of 56)	1593	NR_134469	trio Rho guanine nucleotide exchange factor	protein-coding	chr5
GAS6	10.1	5.769	2.595E-04	3.029E-02	promoter-TSS (NR_044993)	261	NM_000820	growth arrest specific 6	protein-coding	chr13
PIGX	10.7	5.741	2.127E-04	3.029E-02	promoter-TSS (NM_032898)	176	NM_017861	phosphatidylinositol glycan anchor biosynthesis class X	protein-coding	chr3

Table 11: ChIP-Seq Arglu1 (/IgG) Peaks, top by p value

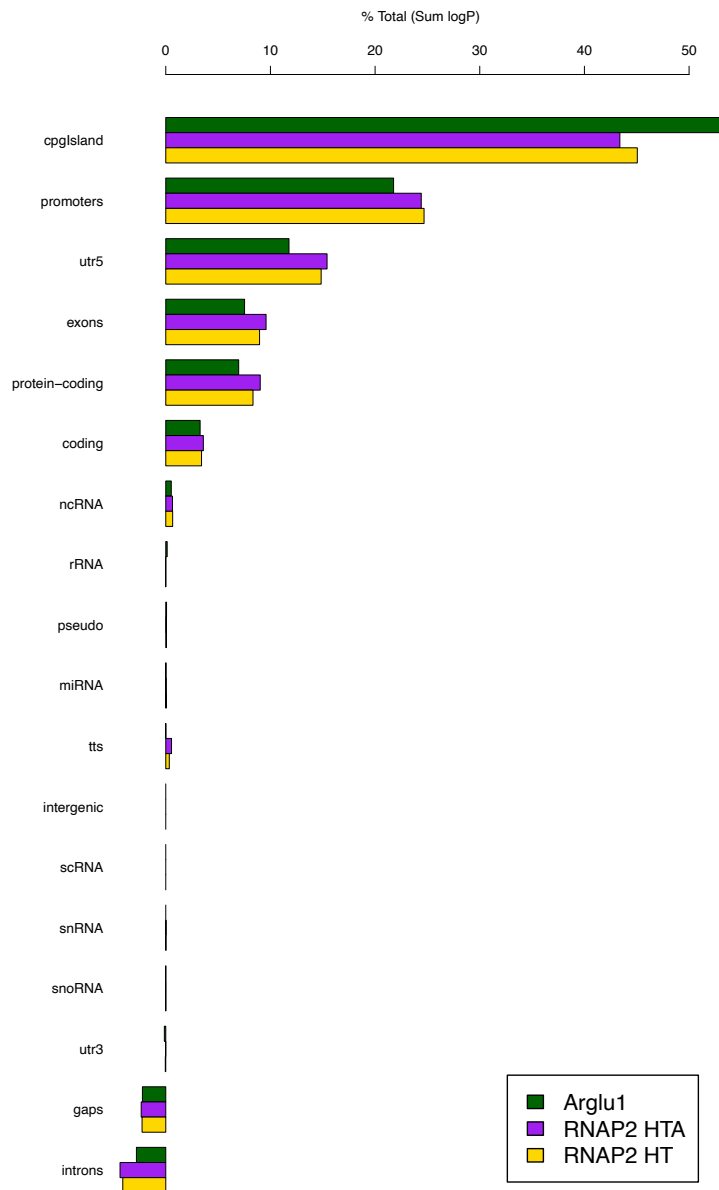


Figure 15: ChIP-Seq Genome Ontology Comparison. For each sample, enrichment p-values for each genomic region were calculated, then log transformed and displayed as a percentage over the sum of all log enrichment values.

Genome Annotation

Genome Ontology of the samples (Figure 15) found that Arglu1 is enriched for CpG islands, promoters, 5' UTR and mildly negatively enriched for 3' UTR, gaps, and introns. This suggests that Arglu1 mainly interacts with promoters not the protein coding region of the genome. HT and HTA RNA Pol II are similar. GO analysis of Arlgu1 (Figure 17, Tables 12-14)

also shows enrichment for genes involved in cancer like Ras GTPase binding and phosphatase activity among others. Interestingly, DNA-binding transcription repressor activity has a lower p value ($1.72\text{e-}7$, 91/242 gene ratio) than transcription activator activity ($4.81\text{e-}4$, 131/440 gene ratio).

TermID	Term	logP	Genes in Term	Target Genes in Term	Targets / Term
GO:0044424	<i>intracellular part</i>	158.899	15067	3622	0.862
GO:0005622	<i>intracellular</i>	158.899	15067	3622	0.862
GO:0005737	<i>cytoplasm</i>	138.626	11626	2936	0.699
GO:0044444	<i>cytoplasmic part</i>	125.660	9736	2523	0.601
GO:0043229	<i>intracellular organelle</i>	123.731	13309	3244	0.772
GO:0043226	<i>organelle</i>	117.940	14157	3397	0.809
GO:0043227	<i>membrane-bounded organelle</i>	92.689	13015	3131	0.745
GO:0005829	<i>cytosol</i>	89.081	5085	1420	0.338
GO:0043231	<i>intracellular membrane-bounded organelle</i>	87.147	11509	2816	0.670
GO:0044422	<i>organelle part</i>	75.189	10114	2499	0.595
GO:0044446	<i>intracellular organelle part</i>	73.764	9813	2432	0.579
GO:0005623	<i>cell</i>	66.872	17586	3938	0.938
GO:0044464	<i>cell part</i>	62.532	17517	3920	0.933
GO:0005654	<i>nucleoplasm</i>	49.328	3515	966	0.230
GO:0012505	<i>endomembrane system</i>	48.978	4441	1181	0.281
GO:0042995	<i>cell projection</i>	36.175	2183	617	0.147
GO:0120025	<i>plasma membrane bounded cell projection</i>	34.948	2108	596	0.142
GO:0031090	<i>organelle membrane</i>	34.326	3484	920	0.219

GO:0005634	<i>nucleus</i>	34.276	7952	1919	0.457
GO:0044463	<i>cell projection part</i>	28.267	1470	425	0.101
GO:0120038	<i>plasma membrane bounded cell projection part</i>	28.267	1470	425	0.101
GO:0031252	<i>cell leading edge</i>	27.127	404	147	0.035
GO:0005856	<i>cytoskeleton</i>	25.599	2198	594	0.141
GO:0030027	<i>lamellipodium</i>	25.483	193	83	0.020
GO:0005794	<i>Golgi apparatus</i>	25.249	1568	442	0.105

Table 12: ChIP-Seq Arglu1 Cellular Component

TermID	Term	logP	Genes in Term	Target Genes in Term	Targets / Term
GO:0005515	<i>protein binding</i>	85.260	11810	3042	0.749
GO:0019899	<i>enzyme binding</i>	36.691	2220	666	0.164
GO:0005488	<i>binding</i>	35.287	15281	3657	0.900
GO:0016740	<i>transferase activity</i>	23.064	2298	650	0.160
GO:0051020	<i>GTPase binding</i>	22.620	544	189	0.047
GO:1990837	<i>sequence-specific double-stranded DNA binding</i>	21.514	889	282	0.069
GO:0003690	<i>double-stranded DNA binding</i>	21.427	983	307	0.076
GO:0043167	<i>ion binding</i>	21.377	6234	1595	0.393
GO:0031267	<i>small GTPase binding</i>	21.356	446	159	0.039
GO:0000976	<i>transcription regulatory region sequence-specific DNA binding</i>	20.776	839	267	0.066
GO:0001067	<i>regulatory region nucleic acid binding</i>	20.604	956	298	0.073
GO:0043168	<i>anion binding</i>	20.429	2790	764	0.188
GO:0001012	<i>RNA polymerase II regulatory region DNA binding</i>	20.365	794	254	0.063
GO:0044212	<i>transcription regulatory region DNA binding</i>	20.271	955	297	0.073
GO:0017016	<i>Ras GTPase binding</i>	19.808	430	152	0.037
GO:0000977	<i>RNA polymerase II regulatory region sequence-specific DNA binding</i>	19.692	788	251	0.062

GO:0043565	<i>sequence-specific DNA binding</i>	17.985	1139	340	0.084
GO:0030695	<i>GTPase regulator activity</i>	17.579	307	113	0.028
GO:0003824	<i>catalytic activity</i>	17.372	5733	1459	0.359
GO:0140096	<i>catalytic activity, acting on a protein</i>	17.275	2192	605	0.149
GO:0000166	<i>nucleotide binding</i>	16.890	2133	589	0.145
GO:1901265	<i>nucleoside phosphate binding</i>	16.817	2134	589	0.145
GO:0019900	<i>kinase binding</i>	16.542	731	229	0.056
GO:0060589	<i>nucleoside-triphosphatase regulator activity</i>	16.497	347	123	0.030
GO:0017076	<i>purine nucleotide binding</i>	16.416	1895	528	0.130

Table 13: ChIP-Seq Arglu1 GO MF

TermID	Term	logP	Genes in Term	Target Genes in Term	Targets / Term
GO:0016043	<i>cellular component organization</i>	64.041	5651	1516	0.374
GO:0071840	<i>cellular component organization or biogenesis</i>	57.709	5841	1544	0.381
GO:0007275	<i>multicellular organism development</i>	56.333	4984	1343	0.331
GO:0023051	<i>regulation of signaling</i>	56.203	3592	1014	0.250
GO:0010646	<i>regulation of cell communication</i>	55.691	3551	1003	0.247
GO:0048856	<i>anatomical structure development</i>	55.069	5378	1431	0.353
GO:0032502	<i>developmental process</i>	54.208	5723	1508	0.372
GO:0048731	<i>system development</i>	53.453	4381	1195	0.295
GO:0007399	<i>nervous system development</i>	52.679	2336	699	0.172
GO:0050794	<i>regulation of cellular process</i>	51.478	11023	2663	0.657
GO:0006996	<i>organelle organization</i>	51.142	3386	953	0.235
GO:0031323	<i>regulation of cellular metabolic process</i>	49.422	6303	1627	0.401
GO:0009966	<i>regulation of signal transduction</i>	48.055	3179	896	0.221
GO:0044260	<i>cellular macromolecule metabolic process</i>	47.774	5036	1333	0.329

GO:0006464	<i>cellular protein modification process</i>	47.517	3060	866	0.214
GO:0036211	<i>protein modification process</i>	47.517	3060	866	0.214
GO:0048522	<i>positive regulation of cellular process</i>	47.040	5449	1426	0.352
GO:0048518	<i>positive regulation of biological process</i>	46.323	6218	1599	0.394
GO:0009653	<i>anatomical structure morphogenesis</i>	45.113	2125	631	0.156
GO:0120036	<i>plasma membrane bounded cell projection organization</i>	44.729	1092	364	0.090
GO:0022008	<i>neurogenesis</i>	44.507	1624	503	0.124
GO:0030030	<i>cell projection organization</i>	44.406	1135	375	0.093
GO:0048523	<i>negative regulation of cellular process</i>	44.332	4843	1279	0.315
GO:0051128	<i>regulation of cellular component organization</i>	44.049	2473	715	0.176
GO:0043412	<i>macromolecule modification</i>	43.498	3270	907	0.224

Table 14: ChIP-Seq Arglu1 GO BP

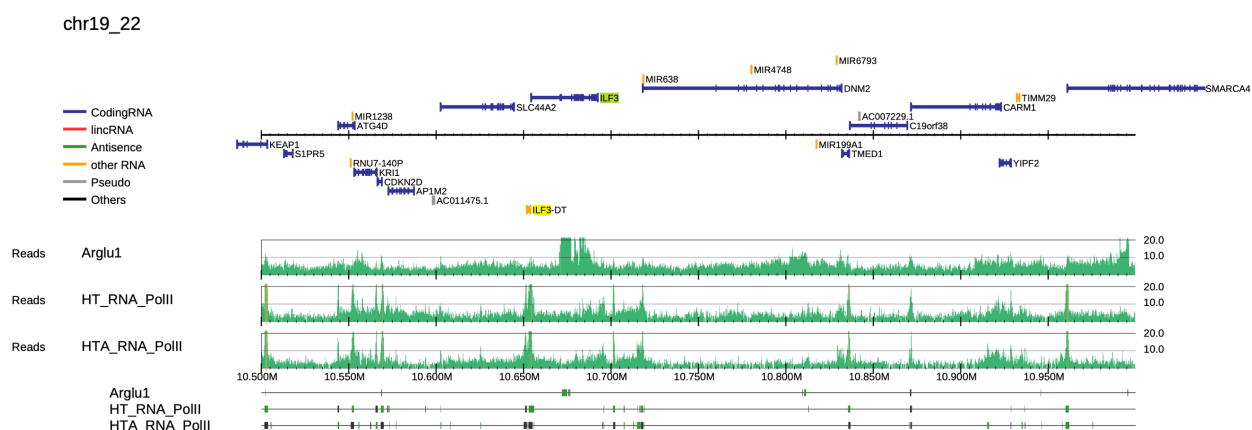


Figure 16: ChIP-Seq of ILF3 for three samples Arglu1 (HTA), HT RNAP2 (RNA Polymerase II), and HTA RNAP2. Yellow overlaps of peaks and bottom three tracks denote significantly enriched peaks. Generated using Drompa+. The top half shows genes colored according to type (CodingRNA, lincRNA, etc.).

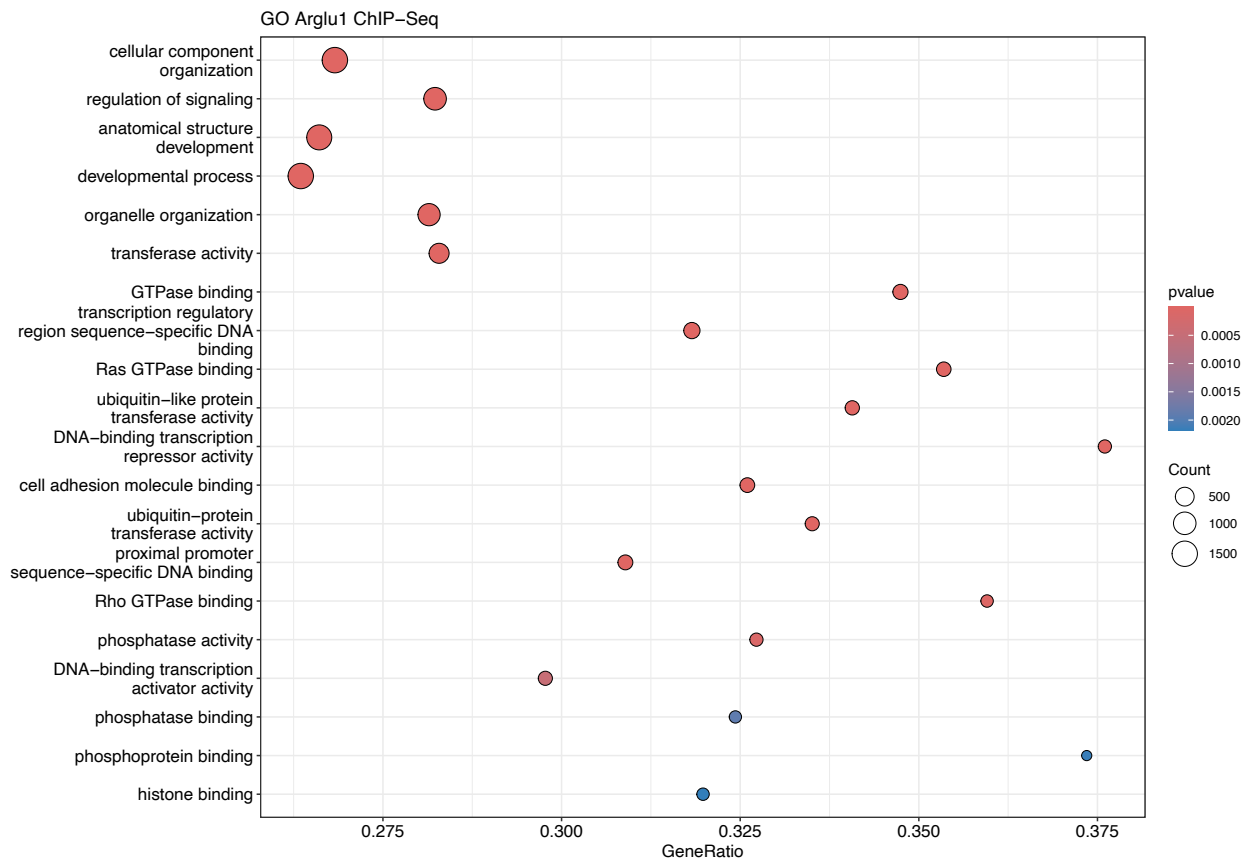


Figure 17: ChIP-Seq Arglu1 (HTA) GO Dot Plot. Gene Ratio is operationally defined as the number of sample (Arglu1) genes in the GO category (Count) divided by the total number of genes in the category. Dots are colored by p value and sized by number of genes in the sample in each GO term (y axis).

Promotor Proximal Pausing

Although my lab had previously identified 5 genes as having increased promoter proximal pausing in HTA cells (Bachus et al. 2024) I wanted more general characterization of pausing in HTA cells as well as genes occupied by Arglu1. I used the pausing index calculator (PIC) algorithm, developed by Day et al. (2016), to identify genes that had higher rates of pausing. The algorithm calculates a pausing index (PI) using the formula $PI = \frac{\text{ReadCount(TSSR)}}{L1} / \frac{\text{ReadCount(GeneBody)}}{L2}$ where TSSR is the transcription start site region, and L1 and L2 are used to normalize the lengths. The higher the PI the more pausing is predicted. Note

that the PI gives information on net values of occupying transcripts but does not give information on the stability and dynamics of pausing (i.e. it does not indicate whether peaks are from paused or terminated polymerase, see discussion), though it is correlated with gene expression (Chen et al. 2018). I applied the PIC to the ChIP-Seq results of HTA RNAP2 with HT RNAP2 as a control. Overall, 10690 transcripts were paused and 441 were unpaused. The average log pausing index over all genes was 5, the average logPI over just genes occupied by Arglu1 was 5.77 (Figure 18). Some of the genes with the highest amount of pausing were ARMC10 and CD244 (Table 15). GSEA found that GO terms involved in DNA repair, cell cycle, growth, development, transcription, chromatin, histones, and ubiquitin were upregulated (Figure 19, Tables 16 and 17). There are over 400 transcripts each associated with the cell cycle, DNA repair, and ubiquitin. Additionally, there are 207, 123, and 31 transcripts associated with double-strand break repair, recombinational repair, and the transcription elongation factor complex.

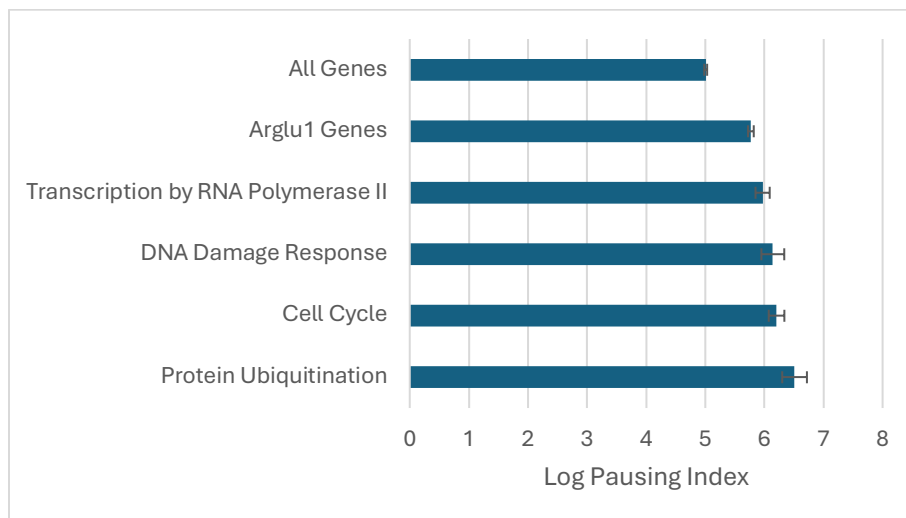


Figure 18: Average log Pausing Index for All Genes compared to Arglu1 occupied genes and GO terms from Arglu1 occupied genes. Pausing Index calculated using Pausing Index Calculator (PIC). Bottom 4 columns are from GO terms from HTA Arglu1.

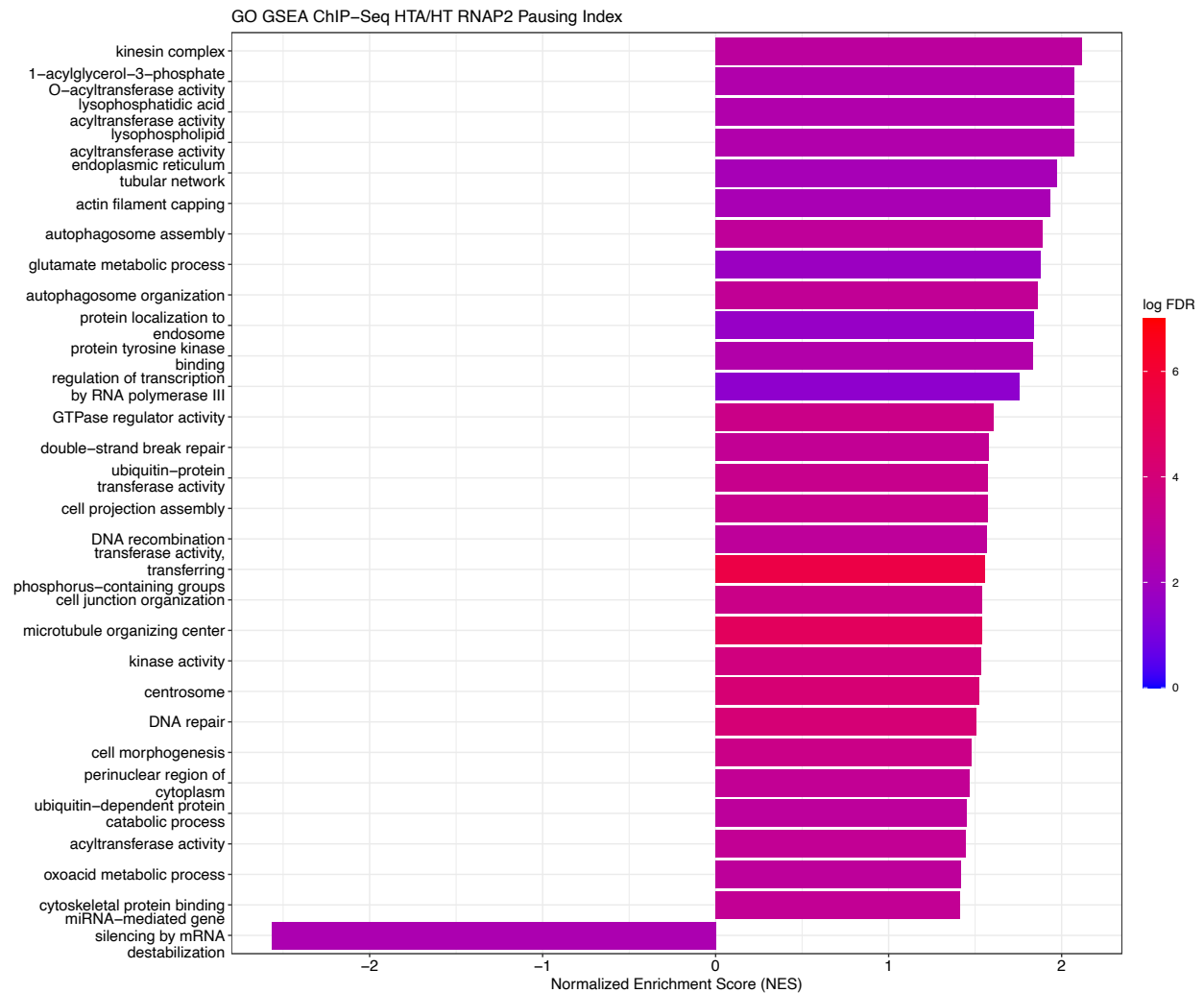


Figure 19: Pausing Index GSEA GO Bar Chart. Gene set enrichment analysis (GSEA) of differential expression comparisons. The x axis is the normalized expression score (NES), indicating the extent to which the GO terms (y axis) are upregulated (NES > 0) or downregulated (NES < 0). Bars are colored based on logFDR (false discovery rate) (scale bar on side). Note that this GSEA was calculated using pausing index instead of fold change and p value.

Gene	Log2PI	Chromosome	TSSstart	TSSend	GBstart	GBend
<i>ARMC10</i>	18.121	chr7	103074832	103075180	103075180	103102762
<i>DCUN1D4</i>	16.894	chr4	51843104	51843452	51843452	51919836
<i>YY1AP1</i>	16.624	chr1	155688475	155688823	155656443	155688475
<i>CD244</i>	16.090	chr1	160862588	160862936	160827161	160862588
<i>CLINT1</i>	15.970	chr5	157858846	157859194	157782748	157858846
<i>NBPF1</i>	15.921	chr1	16613265	16613613	16559324	16613265
<i>ERC1</i>	15.680	chr12	991174	991522	991522	1498930

LINC01163	15.616	chr10	128285901	128286249	128286249	128320725
SLC25A16	15.522	chr10	68527224	68527572	68479345	68527224
MAN2A2	15.310	chr15	90902616	90902964	90902964	90925578
EIF4ENIF1	15.295	chr22	31489543	31489891	31436368	31489543
TMCO4	15.181	chr1	19799617	19799965	19679214	19799617
PFKM	15.040	chr12	48105824	48106172	48106172	48149403
SLC20A2	15.006	chr8	42541213	42541561	42413463	42541213
SCARB2	14.984	chr4	76213525	76213873	76155740	76213525
LINC02085	14.762	chr3	101940810	101941158	101941158	102000925
NECTIN1	14.589	chr11	119728901	119729249	119671390	119728901
CDR2	14.524	chr16	22374320	22374668	22342937	22374320
PYROXD1	14.523	chr12	21437866	21438214	21438214	21474251
ZNF346	14.498	chr5	177022631	177022979	177022979	177084188
HSPH1	14.467	chr13	31162089	31162437	31131975	31162089
USP17L11	14.459	chr4	9248581	9248929	9248929	9253221
USP17L18	14.459	chr4	9248581	9248929	9248929	9253221
VTI1B	14.454	chr14	67674333	67674681	67644087	67674333
LINC01624	14.414	chr6	170272425	170272773	170272773	170282465

Table 15: Pausing Index (PI) top most paused genes (HTA/HT), top 25

Ontology	Term ID	Description	Set Size	NES	pvalue	qvalue
BP	GO:0044087	regulation of cellular component biogenesis	482	1.552	7.955E-10	2.976E-06
MF	GO:0016772	transferase activity, transferring phosphorus-containing groups	453	1.554	1.281E-09	2.976E-06
CC	GO:0005815	microtubule organizing center	468	1.539	8.460E-09	1.311E-05
CC	GO:0005813	centrosome	410	1.522	6.030E-08	7.006E-05
BP	GO:0006281	DNA repair	415	1.506	8.337E-08	7.749E-05
MF	GO:0016301	kinase activity	349	1.534	1.950E-07	1.510E-04
BP	GO:0000902	cell morphogenesis	431	1.480	4.896E-07	2.908E-04
MF	GO:0030695	GTPase regulator activity	231	1.606	6.003E-07	2.908E-04
MF	GO:0060589	nucleoside-triphosphatase regulator activity	231	1.606	6.003E-07	2.908E-04
BP	GO:0034330	cell junction organization	298	1.541	6.258E-07	2.908E-04
BP	GO:0030031	cell projection assembly	263	1.571	1.018E-06	4.076E-04

MF	GO:0008047	<i>enzyme activator activity</i>	290	1.560	1.052E-06	4.076E-04
BP	GO:0120031	<i>plasma membrane bounded cell projection assembly</i>	257	1.591	1.172E-06	4.118E-04
MF	GO:0004842	<i>ubiquitin-protein transferase activity</i>	249	1.574	1.241E-06	4.118E-04
BP	GO:0031344	<i>regulation of cell projection organization</i>	298	1.521	1.463E-06	4.123E-04
BP	GO:0048666	<i>neuron development</i>	497	1.425	1.497E-06	4.123E-04
MF	GO:0016773	<i>phosphotransferase activity, alcohol group as acceptor</i>	319	1.512	1.532E-06	4.123E-04
BP	GO:0051129	<i>negative regulation of cellular component organization</i>	379	1.469	1.634E-06	4.123E-04
BP	GO:0032989	<i>cellular anatomical entity morphogenesis</i>	328	1.510	1.686E-06	4.123E-04
BP	GO:0120035	<i>regulation of plasma membrane bounded cell projection organization</i>	290	1.547	2.184E-06	5.075E-04
BP	GO:0048812	<i>neuron projection morphogenesis</i>	277	1.524	2.425E-06	5.366E-04
CC	GO:0048471	<i>perinuclear region of cytoplasm</i>	362	1.469	3.422E-06	7.162E-04
MF	GO:0019787	<i>ubiquitin-like protein transferase activity</i>	270	1.547	3.629E-06	7.162E-04
BP	GO:0048858	<i>cell projection morphogenesis</i>	285	1.533	3.698E-06	7.162E-04
BP	GO:0006302	<i>double-strand break repair</i>	207	1.577	4.251E-06	7.228E-04

Table 16: Pausing Index GSEA (HTA/HT). Top GO terms by p value.

	Ont	Description	NES	pvalue	qvalue
DNA Repair					
GO:0006281	BP	<i>DNA repair</i>	1.506	8.340E-08	7.750E-05
GO:0006302	BP	<i>double-strand break repair</i>	1.577	4.250E-06	0.001
GO:0000724	BP	<i>double-strand break repair via homologous recombination</i>	1.588	1.350E-04	0.005
GO:0000725	BP	<i>recombinational repair</i>	1.567	2.210E-04	0.007
GO:0045739	BP	<i>positive regulation of DNA repair</i>	1.572	7.560E-04	0.017
Cell Cycle/Growth/Development					
GO:1903047	BP	<i>mitotic cell cycle process</i>	1.331	2.529E-04	0.008
GO:0044770	BP	<i>cell cycle phase transition</i>	1.312	2.867E-03	0.038
GO:0016049	BP	<i>cell growth</i>	1.347	2.571E-03	0.035
Other					
GO:0008023	CC	<i>transcription elongation factor complex</i>	1.720	1.997E-03	0.031

GO:0016567	BP	<i>protein ubiquitination</i>	1.376	4.238E-05	0.003
------------	----	-------------------------------	-------	-----------	-------

Table 17: Pausing Index GSEA important Terms. Interesting Terms with relevance to DNA Damage Response, Cell Cycle, elongation, and ubiquitin.

Motif Enrichment

Motif enrichment analysis of Arglu1 occupied regions (Table 18) found that the most enriched motif is an Atf3 motif. Atf3 is a ‘hub’ transcription factor involved in inflammation, stress response, regeneration, DNA damage, and cancer (Ku and Cheng 2020). As far as I am aware, this is the first time Atf3 have been directly associated with Arglu1 functionally.

Homer *de novo* Motif Results Arglu1 over IgG

Total target sequences = 6555

Total background sequences = 42880

* - possible false positive

Rank	Motif	P-value	log P-value	% of Targets	% of Background	STD(Bg STD)	Best Match/Details
1		1e-59	-1.376e+02	4.26%	1.34%	52.6bp (54.9bp)	Atf3(bZIP)/GBM-ATF3-ChIP-Seq(GSE33912)/Homer(0.955)
2		1e-47	-1.103e+02	0.40%	0.00%	60.3bp (9.0bp)	RUNX2(Runt)/PCa-RUNX2-ChIP-Seq(GSE33889)/Homer(0.570)
3		1e-46	-1.060e+02	0.44%	0.01%	38.9bp (21.9bp)	p53(p53)/mES-cMyc-ChIP-Seq(GSE11431)/Homer(0.673)
4		1e-42	-9.694e+01	0.41%	0.01%	49.4bp (43.1bp)	Pit1(Homeobox)/GCrat-Pit1-ChIP-Seq(GSE58009)/Homer(0.651)
5		1e-41	-9.502e+01	0.35%	0.00%	50.8bp (0.0bp)	PU.1(ETS)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer(0.648)
6		1e-37	-8.621e+01	0.44%	0.01%	43.6bp (55.0bp)	Nr2e3/MA0164.1/Jaspar(0.622)
7		1e-34	-7.923e+01	0.35%	0.01%	46.7bp (32.3bp)	RELA/MA0107.1/Jaspar(0.604)
8		1e-31	-7.265e+01	0.50%	0.02%	55.3bp (41.1bp)	PB0126.1_Gata5_2/Jaspar(0.665)
9		1e-30	-6.908e+01	2.79%	1.05%	51.8bp (57.3bp)	BORIS(Zf)/K562-CTCF-ChIP-Seq(GSE32465)/Homer(0.907)
10		1e-29	-6.725e+01	0.37%	0.01%	39.8bp (39.8bp)	GCM2/MA0767.1/Jaspar(0.609)
11		1e-28	-6.638e+01	0.31%	0.01%	60.1bp (5.3bp)	Sp5(Zf)/mES-Sp5.Flag-ChIP-Seq(GSE72989)/Homer(0.600)
12		1e-27	-6.403e+01	0.49%	0.03%	52.3bp (51.9bp)	PB0055.1_Rfx4_1/Jaspar(0.579)
13		1e-23	-5.398e+01	0.26%	0.01%	48.9bp (35.5bp)	PB0173.1_Sox21_2/Jaspar(0.654)
14		1e-23	-5.377e+01	0.64%	0.08%	47.7bp (50.7bp)	ETV6/MA0645.1/Jaspar(0.675)
15		1e-21	-4.875e+01	11.57%	8.15%	53.2bp (58.5bp)	HIC1(Zf)/Treg-ZBTB29-ChIP-Seq(GSE99889)/Homer(0.627)
16		1e-20	-4.684e+01	41.14%	35.55%	55.7bp (59.4bp)	RFX7/MA1554.1/Jaspar(0.740)
17		1e-18	-4.309e+01	27.16%	22.44%	56.5bp (58.8bp)	PU.1(ETS)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer(0.838)
18		1e-18	-4.269e+01	0.18%	0.00%	53.3bp (14.6bp)	TEAD1(TEAD)/HepG2-TEAD1-ChIP-Seq(Encode)/Homer(0.738)
19		1e-18	-4.244e+01	3.54%	1.87%	53.3bp (59.4bp)	TEAD3/MA0808.1/Jaspar(0.753)
20		1e-16	-3.827e+01	0.20%	0.01%	63.4bp (9.8bp)	ZNF354C/MA0130.1/Jaspar(0.645)

21		1e-14	-3.320e+01	16.53%	13.17%	55.9bp (58.9bp)	NFIC/MA0161.2/Jaspar(0.802)
22		1e-12	-2.804e+01	0.37%	0.05%	48.6bp (53.4bp)	SD0003.1_at_AC_acceptor/Jaspar(0.700)
23 *		1e-11	-2.669e+01	11.43%	8.90%	55.3bp (60.4bp)	ETS2/MA1484.1/Jaspar(0.728)
24 *		1e-11	-2.621e+01	1.01%	0.38%	59.3bp (57.0bp)	IRF4(IRF)/GM12878-IRF4-ChIP-Seq(GSE32465)/Homer(0.828)
25 *		1e-11	-2.586e+01	17.55%	14.52%	56.0bp (61.8bp)	POL010.1_DCE_S_III/Jaspar(0.676)
26 *		1e-11	-2.578e+01	0.12%	0.00%	55.8bp (37.8bp)	PSE(SNAPc)/K562-mStart-Seq/Homer(0.629)
27 *		1e-10	-2.443e+01	10.26%	7.97%	54.5bp (60.8bp)	OSR2/MA1646.1/Jaspar(0.660)
28 *		1e-10	-2.334e+01	0.37%	0.07%	55.0bp (54.5bp)	GMEB2/MA0862.1/Jaspar(0.830)
29 *		1e-10	-2.325e+01	9.43%	7.29%	55.7bp (58.0bp)	POL002.1_INR/Jaspar(0.713)
30 *		1e-9	-2.110e+01	8.46%	6.53%	55.3bp (56.9bp)	Zic(Zf)/Cerebellum-ZIC1.2-ChIP-Seq(GSE60731)/Homer(0.680)
31 *		1e-8	-2.030e+01	14.84%	12.36%	56.2bp (58.8bp)	ZFX(Zf)/mES-Zfx-ChIP-Seq(GSE11431)/Homer(0.726)
32 *		1e-8	-1.855e+01	0.31%	0.06%	57.4bp (66.5bp)	Eomes(T-box)/H9-Eomes-ChIP-Seq(GSE26097)/Homer(0.650)
33 *		1e-8	-1.848e+01	4.58%	3.27%	56.9bp (59.0bp)	CREB1/MA0018.4/Jaspar(0.790)
34 *		1e-7	-1.776e+01	5.21%	3.83%	60.1bp (63.6bp)	PB0035.1_Irf5_1/Jaspar(0.704)
35 *		1e-7	-1.738e+01	1.48%	0.81%	52.8bp (58.9bp)	OVOL2/MA1545.1/Jaspar(0.920)
36 *		1e-7	-1.737e+01	2.15%	1.32%	55.8bp (61.4bp)	HIF-1a(bHLH)/MCF7-HIF1a-ChIP-Seq(GSE28352)/Homer(0.729)
37 *		1e-7	-1.647e+01	3.14%	2.14%	55.2bp (61.7bp)	PB0099.1_Zfp691_1/Jaspar(0.662)
38 *		1e-6	-1.573e+01	0.21%	0.04%	52.2bp (47.2bp)	ZBTB33/MA0527.1/Jaspar(0.766)
39 *		1e-5	-1.156e+01	1.16%	0.68%	62.0bp (58.7bp)	PB0143.1_Klf7_2/Jaspar(0.622)
40 *		1e-2	-5.579e+00	0.05%	0.00%	56.2bp (10.2bp)	Ptf1a/MA1618.1/Jaspar(0.684)
41 *		1e-1	-3.723e+00	0.05%	0.01%	34.8bp (45.6bp)	PB0064.1_Sox14_1/Jaspar(0.788)

Table 18: Arglu1 ChIP-Seq de novo Motif Enrichment using Homer

Comparison of Paused Genes and Arglu1-Occupied Genes

Next the three ChIP-Seq samples, HT RNAP2, HTA RNAP2, and HTA Arglu1, were compared. It is difficult to compare peak to peak because they can be in slightly differently locations. For an initial comparative approach, the peaks for each gene were summed. This

discards information about where peaks are located. Future investigations could compare ex. peaks in the promotor area or the gene body. The number of peaks in HT RNAP2, HTA RNAP2, and HTA Arglu1 are 19443, 19475, and 6524 respectively and the number of genes with peaks are 11921, 11813, and 5354. There are 3637 genes with peaks in all three samples (Figure 20 top). Not surprisingly, since the two RNAP2 samples had about double the genes, they also had about double the unique genes. Next the genes were filtered for having both Arglu1 peaks (I used a peak score of greater than 2 as a filter) and a pausing index above 0, which resulted in 2470 genes (Figure 20 bottom). Some genes with high Arglu1 peak scores and HTA/HT RNAP2 pausing indices were CDR2 and ZBTB38. Numerous genes were in GO terms for cell cycle, G1/S phase transition and DNA damage response (Table 19). For the cell cycle, G1/S, DNA repair, DNA damage response, and site of DNA damage GO terms there are 275, 46, 101, 146, and 27 genes, respectively. There are also 127 genes associated with protein ubiquitination.

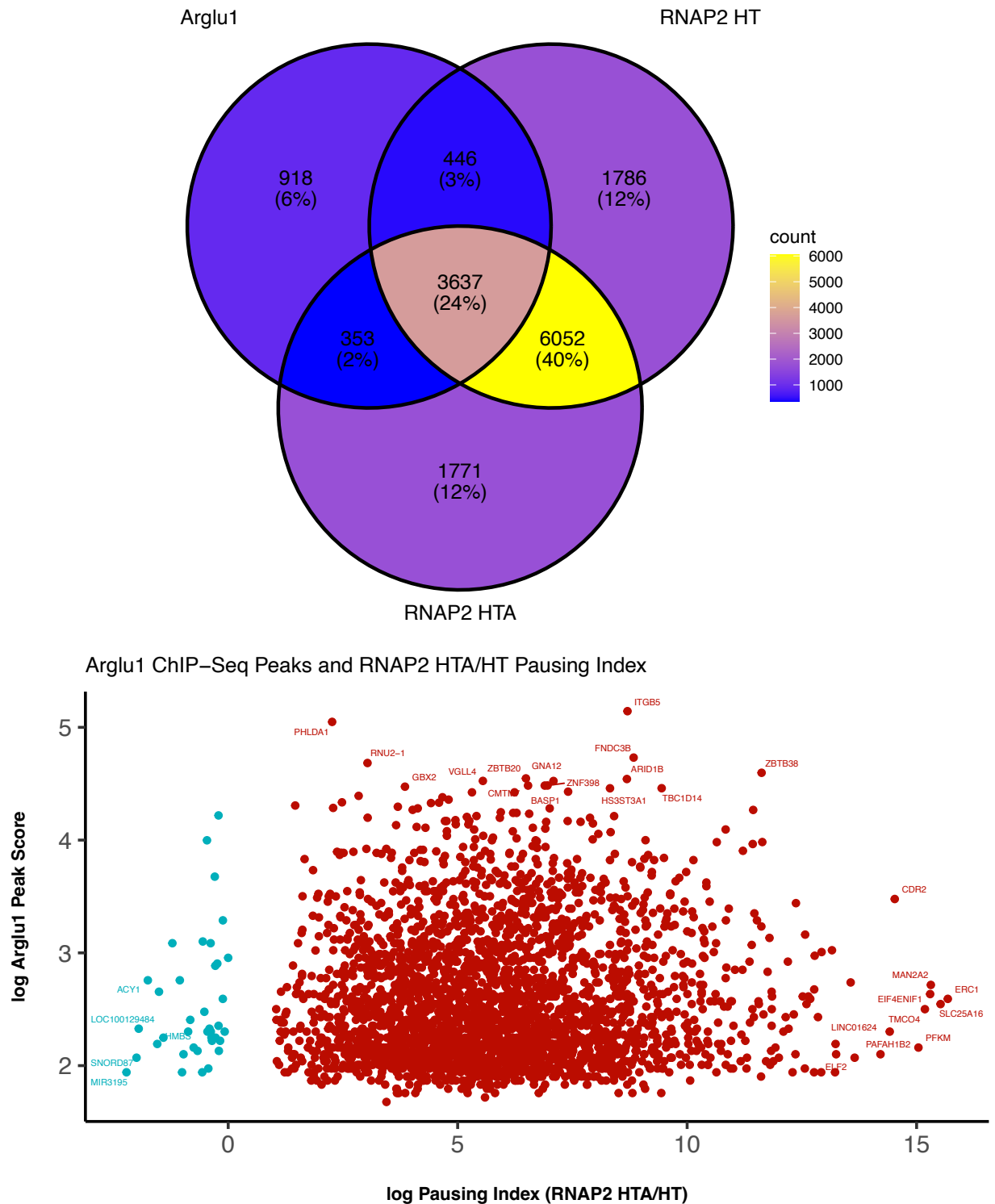


Figure 20-: Arglu1 Peaks and RNAP2 Pausing Index: (top) Venn diagram showing gene peak overlap between RNAP2 HT, RNAP2 HTA, and Arglu1. (bottom) Volcano plot with the log pausing index (PI) for RNAP2 HTA/HT on the x axis and log peak score (PS) for Arglu1. Paused genes (logPI greater than 0) are colored red, unpaused genes are colored blue. Venn diagram generated using ggVennDiagram, Volcano plot generated using ggplot2.

cell cycle	logPI_cc	cell cycle G1/S phase transition	logPI_G1.S	DNA damage response	logPI_DDR
<i>MELK</i>	12.683	<i>ITGB1</i>	11.754	<i>NEDD4</i>	12.769
<i>INO80D</i>	12.518	<i>DLG1</i>	10.908	<i>INO80D</i>	12.518
<i>ITGB1</i>	11.754	<i>APBB1</i>	9.954	<i>TRAF6</i>	11.841
<i>EP400</i>	11.664	<i>APPL2</i>	9.149	<i>EMSY</i>	11.81
<i>PKP4</i>	11.638	<i>ARID1B</i>	8.688	<i>EP400</i>	11.664
<i>RAD18</i>	11.408	<i>CUL1</i>	8.429	<i>ZBTB38</i>	11.622
<i>DLG1</i>	10.908	<i>PPP2R5C</i>	8.075	<i>PIK3R1</i>	11.43
<i>SEN6</i>	10.841	<i>E2F3</i>	7.893	<i>RAD18</i>	11.408
<i>EPB41</i>	10.646	<i>CUL2</i>	7.883	<i>CHD1L</i>	10.631
<i>RAD1</i>	10.508	<i>EZH2</i>	7.769	<i>RAD1</i>	10.508
<i>ODF2</i>	10.453	<i>KIF14</i>	7.498	<i>APBB1</i>	9.954
<i>SEPTIN8</i>	10.379	<i>CCND3</i>	7.352	<i>NPAS2</i>	9.914
<i>CDC42</i>	10.205	<i>RDX</i>	7.274	<i>UBE2N</i>	9.583
<i>FZD3</i>	10.091	<i>BCL7B</i>	7.248	<i>UBR5</i>	9.473
<i>AVEN</i>	10.015	<i>JADE1</i>	7.138	<i>USP7</i>	9.349
<i>SEN5</i>	9.999	<i>ACVR1</i>	6.783	<i>NDRG1</i>	9.152
<i>APBB1</i>	9.954	<i>BRD7</i>	6.677	<i>BRD4</i>	9.022
<i>FGF2</i>	9.946	<i>RPTOR</i>	6.523	<i>ARID1B</i>	8.688
<i>KIF2A</i>	9.722	<i>GIGYF2</i>	6.423	<i>DYRK1A</i>	8.658
<i>TFDP2</i>	9.713	<i>INO80</i>	6.288	<i>PTTG1</i>	8.615
<i>KLHL13</i>	9.666	<i>EIF4E</i>	6.287	<i>POLI</i>	8.586
<i>MND1</i>	9.368	<i>UBE2E2</i>	6.062	<i>RNF138</i>	8.428
<i>CDC25C</i>	9.349	<i>DDR2</i>	5.983	<i>TLK1</i>	8.338
<i>TACC2</i>	9.281	<i>CDK6</i>	5.723	<i>SPIDR</i>	8.132
<i>ING5</i>	9.254	<i>ANKRD17</i>	5.69	<i>PSMD14</i>	8.129
<i>LIN9</i>	9.246	<i>FBXO31</i>	5.682	<i>UBE2W</i>	8.106
<i>C11ORF80</i>	9.244	<i>APPL1</i>	5.556	<i>PPP2R5C</i>	8.075
<i>APPL2</i>	9.149	<i>BCL7C</i>	5.472	<i>FANCB</i>	7.955
<i>RALA</i>	9.068	<i>HINFP</i>	5.443	<i>PPP4R2</i>	7.864
<i>BRD4</i>	9.022	<i>ADAM17</i>	5.302	<i>CIB1</i>	7.852
<i>TACC1</i>	8.987	<i>EGFR</i>	5.124	<i>TIGAR</i>	7.786
<i>PARD3B</i>	8.98	<i>CCNE1</i>	5.041	<i>IMMP2L</i>	7.66
<i>CDK5RAP2</i>	8.813	<i>PRMT2</i>	4.984	<i>ABRAXAS1</i>	7.574
<i>SPTBN1</i>	8.805	<i>E2F7</i>	4.973	<i>PSEN1</i>	7.452
<i>CSPP1</i>	8.78	<i>RRM2</i>	4.676	<i>NSD2</i>	7.42

<i>MAP3K8</i>	8.753	<i>PML</i>	4.521	<i>REV1</i>	7.403
<i>PKN2</i>	8.741	<i>TCF3</i>	4.208	<i>POLQ</i>	7.285
<i>ARID1B</i>	8.688	<i>WEE1</i>	3.98	<i>SETX</i>	7.278
<i>CROCC</i>	8.662	<i>CCNA2</i>	3.334	<i>BCL7B</i>	7.248
<i>STRADB</i>	8.638	<i>PHF10</i>	3.148	<i>ZMAT3</i>	7.159
<i>PTTG1</i>	8.615	<i>DGKZ</i>	3.108	<i>PMS2</i>	7.13
<i>ROCK2</i>	8.608	<i>BCL7A</i>	2.894	<i>HMCES</i>	6.788
<i>C2CD3</i>	8.54	<i>CDKN2C</i>	2.814	<i>BCL2L11</i>	6.781
<i>MYH10</i>	8.498	<i>CDKN2D</i>	2.783	<i>KDM2A</i>	6.759
<i>SLC16A1</i>	8.447	<i>PLK2</i>	2.728	<i>TEX264</i>	6.715
<i>CUL1</i>	8.429	<i>SMARCC2</i>	2.158	<i>FANCA</i>	6.702

Table 19: Arglu1 Peaks-RNAP2 Pausing Index GO terms

Comparison of Proteomics, Pausing, and Arglu1-occupied Genes

Finally, I wanted to analyse genes that were paused, occupied by Arglu1, and differentially expressed as proteins. To that end, the genes with pausing and Arglu1 peaks were intersected with differentially expressed proteins in HTA/HT (Figure 21 top). There are 82 genes in both sets. Some genes with a high protein fold change and pausing index include CDC42EP4, BASP1, RAD23B, and EPB41L3 (Figure 20 bottom). There were also 12 genes in the GO term for cell cycle, 3 for G1/S phase transition, 6 for DNA damage response, 11 for transcription by RNA polymerase II (Table 21). There are 12 genes associated with the cell cycle that were in the overlap: CCNA2, CDC37, CDKN2D, MKI67, NR3C1, PAK4, PCNT, PPM1G, PPP2R5C, RHEB, TET2, and TUBB2A (Table 20). Of these, only the PCNT and PPM1G proteins were upregulated. Three of these, CCNA2, CDKN2D, and PPP2R5C are also associated with the G1/S phase transition. There are also 6 genes associated with DNA damage response, BACH1, BAD, CDKN2D, PPP2R5C, RAD23B, and UBR5; and one gene, ELOA, associated with sites of DNA damage. Of these only the RAD23B protein was upregulated.

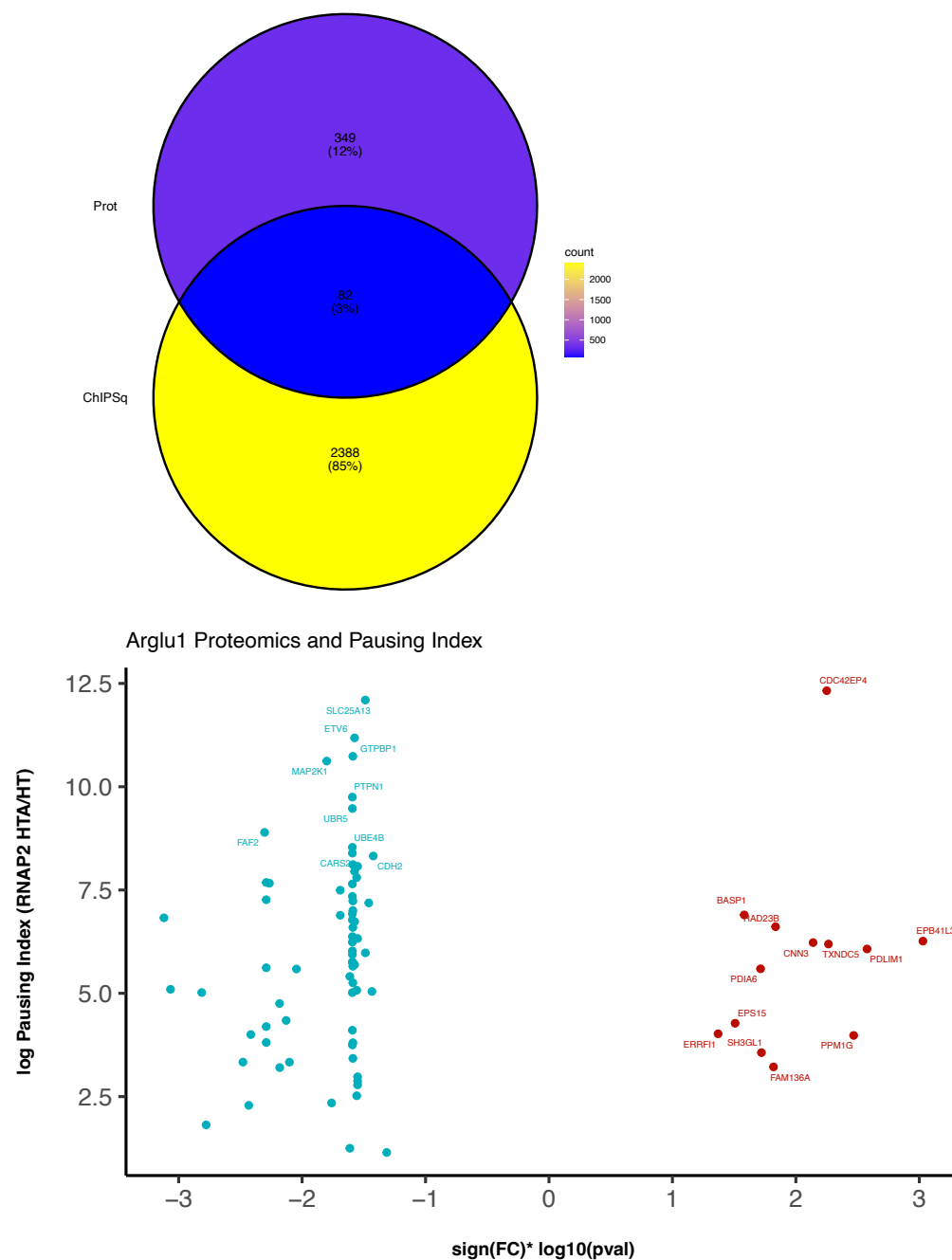


Figure 21: Arglu1 Proteomics and ChIP-Seq Comparison. (top) Venn diagram showing overlap between proteins differentially expressed in HTA/HT (Prot) and Paused Arglu1 peak genes (ChIPSq). (bottom) Volcano plot with proteomics HTA/HT $\text{sign(Fold Change)} * \log_{10}(\text{pvalue})$ on the x axis and the log pausing index (RNAP2 HTA/HT) on the y axis. Upregulated genes ($\text{FC} > 0$, $\text{pvalue} < 0.05$) are red, downregulated genes ($\text{FC} < 0$, $\text{pvalue} < 0.05$) are blue. Venn diagram generated using ggVennDiagram, Volcano plot generated using ggplot2.

Cell Cycle					
Gene	Prot_pval	logFC	logPS	logPI	
CCNA2	0.00331671	-3.9724817	2.77610399	3.33402753	
CDC37	0.01733995	-5.7981582	3.50080205	2.3461864	
CDKN2D	0.02816542	-3.1105555	2.40599236	2.78299291	
MKI67	0.00085928	-4.9741804	2.07038933	5.09629549	
NR3C1	0.02554489	-2.6828211	2.54596837	6.23881155	
PAK4	0.0051236	-3.1462881	2.71589337	7.26435093	
PCNT	0.02037441	-3.2518348	2.22032995	6.88890147	
PPM1G	0.00340145	7.88564694	2.54596837	3.97979452	
PPP2R5C	0.02816542	-3.1105555	2.32768736	8.07478677	
RHEB	0.00900075	-4.4526508	2.63459327	5.58796834	
TET2	0.02554489	-2.6828211	2.50080205	6.02475249	
TUBB2A	0.02816542	-3.1105555	2.07038933	6.327124	
G1/S Phase Transition					
Gene	Prot_pval	logFC	logPS	logPI	
CCNA2	0.00331671	-3.9724817	2.77610399	3.33402753	
CDKN2D	0.02816542	-3.1105555	2.40599236	2.78299291	
PPP2R5C	0.02816542	-3.1105555	2.32768736	8.07478677	
DNA Damage Response					
Gene	Prot_pval	logFC	logPS	logPI	
BACH1	0.02770102	-3.0294524	2.07038933	5.07362171	
BAD	0.02816542	-3.1105555	2.69599381	2.87339592	
CDKN2D	0.02816542	-3.1105555	2.40599236	2.78299291	
PPP2R5C	0.02816542	-3.1105555	2.32768736	8.07478677	
RAD23B	0.01459727	5.27735063	3.56924803	6.61147774	
UBR5	0.02554489	-2.6828211	2.63459327	9.47328395	
Site of DNA Damage					
Gene	Prot_pval	logFC	logPS	logPI	
ELOA	0.02554489	-2.6828211	3.62410089	6.37168678	
Protein Ubiquitination					
Gene	Prot_pval	logFC	logPS	logPI	
TRIM4	0.028	-3.111	2.43	2.979	
UBE4B	0.026	-2.683	2.614	8.534	
UBR5	0.026	-2.683	2.635	9.473	
transcription by RNA polymerase II					
Gene	Prot_pval	logFC	logPS	logPI	regulation
BACH1	0.028	-3.029	2.07	5.074	+/-

<i>ELOA</i>	0.026	-2.683	3.624	6.372	
<i>ETV6</i>	0.027	-2.914	2.275	11.183	+/-
<i>NR3C1</i>	0.026	-2.683	2.546	6.239	+/-
<i>PDLIM1</i>	0.003	12.018	2.478	6.073	
<i>PKIG</i>	0.008	-3.901	3.442	3.333	-
<i>SMAD1</i>	0.02	-3.252	2.22	7.495	+
<i>TAGLN3</i>	0.037	-2.098	2.635	5.045	-
<i>TET2</i>	0.026	-2.683	2.501	6.025	+
<i>TSHZ3</i>	0.035	-3.694	3.417	7.188	
<i>ZNF219</i>	0.028	-3.029	2.16	2.521	+/-

Table 20: Arglu1 Proteomics and Arglu1 ChIP-Seq GO terms. The last column indicates whether the genes activate transcription, inhibit it, or both.

Results summary

In summary, I found with RNA-Seq that Adenovirus without 13S and 12S has nonoptimal effects on human transcription particularly with respect to viral growth and immune suppression and I identified cell cycle associated transcripts differentially expressed in YCD/12S. I found that HT1080 cells overexpressing Arglu1 (HTAs) have increased resistance to the genotoxic drug Actinomycin D and have altered cell cycle phase duration. I made 29 Arglu1 deletion mutant plasmids and 21 stable cell lines overexpressing these mutants but effects on protein binding and cell growth were inconclusive. I analysed ChIP-Seq data pulling down RNA Polymerase II and Arglu1 in HTA cells and found that HTA RNAP2 has increased occupancy of coding genomic regions compared to HT RNAP2, that Arglu1 is enriched for transcription repressor and activator activity, and identified Atf3, ILF3, and PANDAR as functional interaction partners of Arglu1 or HTA RNAP2. I also identified genes associated with the cell cycle, G1/S transition, DDR, ubiquitin, and transcription in Arglu1 data sets for proteomics, genes occupied by Arglu1 and paused HTA genes.

4 Discussion

4.1 Adenovirus E1A Isoforms and YCD Mutant

Adenovirus E1A is one of the earliest expressed viral proteins during Adenovirus infection and an oncogene in cell culture. It activates S phase through the Retinoblastoma protein (pRB) pathway and helps suppress the host immune response, among other effects. After infecting IMR-90s (human primary lung fibroblasts), we performed bulk RNA-seq on a mock/uninfected control, wild type human adenovirus (WT), 2 mutated strains of Adenovirus expressing only the E1A 13S and 12S isoforms, respectively and a 13S YCD mutant with residues 2-11 deleted and Y47 and C124 mutated (Figure 3 top), resulting in a loss of binding to pRb, deficient S phase activation, and reduced viral growth. However, unlike E1A 12S, the 13S YCD mutant was able to drive S phase in a BrdU assay (Figure 3 bottom). PCA and hierarchical clustering showed that the YCD mutant has a distinct gene expression pattern (Figures 4 and 6). Differential expression analysis (Figure 7) and GSEA (Figure 8) further identified differences between the virus strains. Compared to wild type adenovirus (WT), E1A 13S had upregulated heat shock proteins, cytokines and other signaling molecules and downregulated DNA replication and cell cycle transcripts. E1A 12S/WT upregulated golgi associated transcripts, and vesicles while downregulating DNA replication, cell cycle, heat shock proteins, cytokines and other signaling molecules. E1A 13S YCD/WT upregulated heat shock proteins, cytokines and other signaling molecules while downregulating DNA replication and cell cycle transcripts. Some of the most differentially expressed transcripts were heat shock proteins and cytokines like HSPA6,

CXCL6, and IL8. Currently, E1A is mainly thought to affect the immune system through CR1 and the C-terminus (Hendrickx et al. 2014, Graves et al. 2024) so it is not clear why E1A 12S would have less cytokine expression. Note that while cell cycle associated transcripts in 13S, 12S, and YCD are downregulated as compared to WT, they are upregulated when compared to uninfected (UI or Mock) IMR-90 cells (Figure S7, 13S/UI, 12S/UI, YCD/UI).

There are several transcripts that could contribute to driving S phase induction in YCD. Gene set enrichment analysis (GSEA) and over-expression analysis (OEA) identified numerous cell cycle-associated upregulated transcripts in YCD/12S including MYC, Cyclin D1 (CCND), Cyclin J (CCNJ), DDX3X, and PPME1. Cyclin J binds and activates CDK2 in *Drosophila* (Kolonin and Finley 2000). Its human binding partner does not seem to have been identified but it promotes cancer proliferation (Zhang et al. 2019). Cyclin D promotes returning to interphase after cytokinesis (Matthews et al. 2021). DDX3X promotes S phase induction by repressing KLF4 (Cannizzaro et al. 2018). PPME1 regulates phosphorylation of ERK1/2 (MAPK1/2) and PP2A both involved in the cell cycle (Labuzan et al. 2020). PP2A dephosphorylates pRB, inhibiting cell cycle entry, along with PP1 in competition with CDKs (Wlodarchak and Xing, 2016). PPME1 inhibits PP2A, promoting cell cycle entry, cell proliferation, and increasing tumor grade when overexpressed (Guffens et al. 2023). It is frequently over expressed in cancers particularly glioma.

4.2 Arglu1 Cell Cycle

Arglu1 is a human protein that increases the cellular growth rate, pausing of RNA polymerase II, and localizes to sites of DNA damage. We found that Arglu1 overexpressing cells had increased cell growth and decreased G1 and S phases of the cell cycle but increased G2/M phase (Figure 10). Proteomics (HTA vs HT) and ChIP-seq (pulling down RNAP2 in HTA and HT and Arglu1 in HTA) were performed to provide further insight (Figures 11-21). There are 276 cell cycle associated paused genes occupied by Arglu1 including cyclins, phosphatases, kinases and other signaling molecules, chromosome remodelers and separators, histone methylators and acetylators, as well as genes involved in the cytoskeleton, cell differentiation and ubiquitin. Twelve of these genes are also differentially expressed proteins (Tables 8, 20).

Several paused and Arglu1-occupied transcripts could be involved with how Arglu1 overexpression alters the phase duration of the cell cycle including CyclinA2/CCNA2, CyclinD3/CCND3, CDKN2D, CyclinE1/CCNE1, WEE1, PPP2R5C and PPM1G (Matthews et al. 2021, Kim and Sharpless 2006). Among the cyclins, CCNA2 binds and activates CDK2 promoting the G1/S and G2/M transitions, while CDKN2D binds CDK4 and 6 and represses the G1/S transition. CCND3 regulates CDK4 and 6, CCNE1 regulates CDK2, and the WEE1 checkpoint kinase inhibits CDK1. There are two phosphatases, PPP2R5C and PPM1G, both of which regulate the cell cycle. Ectopic expression of PPM1G delayed the G1/S phase transition while knockdown of PPM1G increased the rate of G1/S transition and increased cell resistance to cell cycle arrest caused by serum deprivation (Sun et al. 2016). PPP2R5C

is a regulatory subunit of protein phosphatase 2A and together with PPP2R5(A–E) is degraded by the HIV-1 Vif protein to induce G2/M cell cycle arrest (Hu et al. 2024).

Overview of Cyclins in the Cell Cycle

The role of cyclins in the cell cycle can be simplified as follows (Matthews et al. 2021). The MAPK pathway (RAS–RAF–MEK–ERK) activates MYC which activates various growth signals and mitogens which then block CDK inhibitors and activate CyclinD which activates E2F-dependent transcription. E2F is inhibited by binding to RB but RB can be inactivated by CDK based phosphorylation. E2F is also regulated by 2 other pocket proteins, p107 and p130. E2F transcription increases Cyclins E and A which creates a decision window to enter S phase (G1/S transition). CyclinE-CDK2 activity further increases E2F transcription which, in a positive feedback loop, increases CyclinE-CDK2 and CyclinA-CDK2. The A/B-CDK2 complexes inactivate APC/C-CDH1 and initiate S phase. S phase results in an increase in CyclinA/B-CDK1 (not 2) which opens a G2/M mitotic entry decision window and activates APC/C-CDC20 which then causes mitotic exit and ubiquitin tagging of cyclins. Finally, 'favorable conditions' and an increase in CyclinD-CDK4/6 initiate cell cycle reentry to G1 (Matthews et al. 2021).

4.3 Arglu1 in Transcription and Pausing

Gene over expression analysis (OEA) of Arglu1 proteomics data found several enriched GO terms associated with transcription and pausing including DNA-templated transcription elongation and transcription regulation by TP53 (Table 8) while GSEA GO terms of paused

RNAP2 HTA genes included transcription elongation factor complex (Table 17). DICER, which cleaves dsRNA and is involved with miRNA and siRNA pathways (Foulkes et al. 2014), was among the paused transcripts.

Why are some genes more paused upon Arglu1 overexpression than others? What determines which genes Arglu1 occupies? Why does Arglu1 overexpression increase the amount of paused DDR genes when this might seem like it would decrease expression? In fact, pausing is only weakly correlated to expression and pausing can have other effects including coordinating gene expression, integrating signaling pathways, regulating cell development and cell fate, and possibly increasing open chromatin (Abuhashem et al. 2022). Sorting this out might require a very fine-grained analysis of gene expression in time. scRNA-seq can model some aspects of the temporal dynamics of genes expression but with considerable limitations (Ding, et al. 2023).

Some studies suggest pausing increases open chromatin and hence transcription, though all the methods for measuring this have confounding factors (Chen et al. 2012), though paused RNAP2 also allows H3K4me3 and H3K9ac histone modifications, which can then recruit the SEC to promote elongation (Gates et al. 2017). Conversely, histone deacetylases such as SIRT6 can remove the H3K9ac mark to increase pausing and decrease gene expression though controlling P-TEFb and SEC assembly (Etchegaray et al. 2019, Vaid et al. 2020). Studies in *Drosophila* have found that there is more pausing in early development compared to adult tissue specific genes. This early pausing could be used for

coordinating expression in early development. Unfortunately, the role of pausing in mammalian development is less understood compared to *Drosophila* (Abuhashem et al. 2022), though in the number of paused genes varies during mouse embryo development (Liu et al. 2020). Interestingly, several GSEA GO terms for RNAP2 HTA paused genes were related to development and cell growth, particularly development of neurons (Table 16).

Pausing is often altered in cancers. In human breast cancer cells (MCF-7), genes that respond to signaling pathways are also enriched for pausing. (Abuhashem et al. 2022). Retinoic acid and estrogen signaling can also control pausing and hence transcription, furthermore this control is disrupted in cancer. ERα (Estrogen receptor alpha) directly interacts with NELFB and p-TEFb and is likely involved in their recruitment to RNAP2. In mice, NELF binding to RNAP2 controls inflammation in macrophages (Sacta MA et al. 2018, Yu et al. 2020). In a case of promotor non-proximal pausing on the MYB gene in MCF-7 cells, ERα drives pause release at the SL-dT region, a regulatory region ~1.7 kb after the TSS (Mitra et. a. 2012). In mES cells, MYC can regulate gene expression by pause release via p-TEFb and other elongation factors (Rahl et al. 2010, Cartwright et al. 2005). Similarly, TRIM33 in zebrafish and TRIM28 in humans have also been found to mediate pausing and gene expression (Bai et al. 2010, Bunch et al. 2014).

4.4 Arglu1 DNA Damage

DNA damage has a complex relationship with cancer. On the one hand DNA damage often contributes to cancer. On the other hand, too much DNA damage will kill cancer cells even if they're more resistant than regular cells. One effect of DDR is to shut down transcription to allow for repair of double strand breaks (DSB) (Gregersen and Svejstrup 2018). Arglu1 could cause more DSB by accelerated cell division or pause RNAP2 to allow for more repair of DSB (Rothkamm et al. 2015) or both. Either could contribute to cancer. Previously my lab found that Arglu1 is associated with anti-pause enhancers of RNAP2. RNAP2 Promotor Proximal Pausing (RNAP2PPP) regulates the transcription of many genes (Adelman and Lis 2012). The main factor in releasing paused RNAP2 is the P-TEFb complex (Wada et al. 1998). Shortly after transcription initiation around 20-60 nucleotides, DSIF and NELF interact with RNAP2 to pause transcription. Pausing can be relieved (and transcription and elongation resumed) by P-TEFb phosphorylating DSIF and NELF. However, P-TEFb can be inhibited by a complex including the snRNA 7SK and proteins HEXIM1 or 2 (Nguyen et al. 2001; Yang et al. 2001). BRD4 has been found to release P-TEFb from its inhibitory complex (Jang et al. 2005; Yang et al. 2005; Krueger et al. 2010).

The 7SK snRNA interacts with H4R3me2 repressive histone markers and the 7SK/HEXIM complex holds the P-TEFb complex at such markers. BRD4 and JMJD6 bind P-TEFb through CDK9 and CCNT1. Jmjd6 then demethylates H4R3me2 and 7SK which results in the disassembly of the complex inhibiting P-TEFb which subsequently releases RNAP2 (Liu et al. 2013). This activity of BRD4, JMJD6, and HEXIM (BJH complex) is associated with their binding to distal anti-pause enhancers. Arglu1 disrupts the ability of BJH to bind to distal

anti-pause enhancers by inhibiting the Brd4-Jmjd6 interaction (Baccus et al. 2024). This prevents the release of paused RNAP2. This suggests Arglu1 could induce transcription disruption possibly leading to a stress response induced by replication stalling.

In addition to BRD4, many other bromodomain proteins were paused including BRD1, BRD2 (observed), BRD3, BRD7, BRD9, CYBRD1, BRD3OS, and LMBRD1. Notably, 3 out of 4 of the Bromo- and Extra-Terminal domain (BET) family (BRD2-4 and BRDT) were paused. BET family proteins are transcription activators and are targeted by the BET inhibitor class of chemotherapy drugs (Wang et al. 2023). BRDT was not occupied by RNAP2 (HT or HTA) or Arglu1. Brd4, Jmjd6, and Hexim1 (together the BJH complex) have all been associated with the DNA damage response pathway (Floyd et al. 2013; Zhang et al. 2018, Bowry et al. 2018; Huo et al. 2020). Brd4 (Bromodomain-containing protein 4) is a mitotic bookmark (epigenetic control of cell identity during division) and transcriptional activator of genes associated with M/G1 growth. Brd4 dysfunction is associated with cancer and up regulation of oncogenes such as Myc. (Drumond-Bock and Bieniasz 2021; Lovén et al. 2013; Rhyasen et al. 2018; Wang et al. 2020). HTA cells also displayed increased resistance to bleomycin (Baccus et al. 2024) and actinomycin D (Figure 9). Bleomycin causes DSBs, while actinomycin D inhibits RNA polymerases eventually leading to single strand breaks by stalled replication forks. Further experiments in Baccus et al. (2024) have shown that this resistance is due to enhanced DDR by Arglu1, probably through Brd4 and Jmjd6 interaction.

GSEA (Gene Set Enrichment Analysis) of the Arglu1 proteomics identified numerous upregulated GO terms associated with DDR including cellular response to stress, DNA repair, and DNA damage response (Table 7). For Arglu1 ChIP-seq pausing some of the DDR GO terms included DNA repair, DSB repair, and recombinational repair (Tables 16, 17). There are numerous paused transcripts in the PARP family including PARP1, PARP2, PARP3, PARP8, PARP9, PARP10, TIPARP, PARPBP. PARP enzymes are mainly responsible for detecting single strand breaks, especially PARP1 and 2 (Jubin et al. 2016). PARP1 appears paused on a genome viewer but it is not a peak in HT RNAP2 so a pausing index was not calculated for it. PARP inhibitors have been used as a form of chemotherapy (Groelly et al. 2023).

Transcriptionally Coupled Repair and Atf3

Atf3 is a transcription factor that can act as a transcriptional activator or repressor (Ku et al. 2020). The top enriched motif in Arglu1 ChIP-seq peaks is for ATF3 (Table 18). Transcription coupled repair (TCR) includes homologous recombination (HR), non-homologous end joining (NHEJ), and nucleotide excision repair (NER) and initiates when RNAP2 stalls at a DNA lesion (van den Heuvel et al. 2021 and Min et al. 2022). The stalling of RNAP2 is detected by ERCC6 at which point additional proteins including ERCC8 and the DDB-1-CUL4A-RBX1 complex (CRL4) are recruited (Epanchintsev et al. 2017, Cheng et al. 2024). Once DNA repair is complete, before transcription can resume, ERCC6 and ERCC8 mediate removal of Atf3 from promoters (Epanchintsev et al. 2017). ERCC6 has a logPI of

7.8, though it is not occupied by Arglu1 or a differentially expressed protein in HTA cells.

This suggests Atf3 could be one way in which Arglu1 mediates responses to DNA damage.

Arglu1 and Neurodegenerative Diseases

There are several differentially expressed proteins associated with Alzheimer's disease and Parkinson's disease including APBB1/NAE1 and PARK7/DJ-1. (Chen et al. 2000, Dolgacheva et al. 2019). PARK7 is also paused and occupied by Arglu1. APBB1 binds to the beta-amyloid precursor protein associated with Alzheimer's disease. It is also involved in NEDDylation and the S/M checkpoint. PARK7 interacts with PARP1 and promotes HR and NHEJ repair (Wang et al. 2023). There are several studies that have linked Arglu1 to ALS and disorders of development (Daneshafrooz et al. 2022, Yao et al. 2023). DNA damage is a common cause of many neurodegenerative diseases (Welch and Tsai 2022). However, since Arglu1 can also co-activate glucocorticoid receptor (GR) and elevated glucocorticoid levels are also associated with neurodegenerative diseases, further investigation is needed as to whether Arglu1 helps or hinders neurodegeneration. Or like with different cancers, it could cause both in different contexts. There are currently no articles on Pubmed linking Arglu1 with APBB1 or PARK7.

4.5 Limitations and Looking Ahead

Arglu1 and Adenovirus

E1A binds Arglu1. But why? Arglu1 increases cell growth but paradoxically it inhibits viral growth and gene expression (Baccus et al. 2024). The fact that adenovirus replicates better under cell cycle arrest (Arglu1 knockdown) suggests that the virus does not need Arglu1's ability to increase cell growth. E1A is also coimmunoprecipitated with Jmjd6 and partially restores the Brd4-Jmjd6 interaction, likely decreasing RNAP2 pausing. Further research on how Arglu1 interacts with E1A and adenovirus infection and replication is needed, such as whether Arglu1 overexpression affects localization of viral proteins and RNA.

Proteomics and Missing Values

There are many causes of missing values in proteomics, some are entirely random but they occur more often at low intensities (Ahlmann-Eltze and Anders 2019). The average amount of missing values in a proteomics sample is 40%, which is also about the average missing values in the HT and HTA samples. The most common way of dealing with missing values is imputation i.e. replacing the missing values with a 'realistic' value. One downside of imputation is that to downstream algorithms, a missing value looks as certain as a normal value. Suppose you have two samples or more for a given condition, it would be preferable to differentiate between samples with no missing values, 1 missing value, and all missing values. proDA incorporates this information by shifting the prediction based upon the assumption that missing values are more likely to have low intensities. This assumption also allows it to predict values only present in one sample.

RNA-Seq and ChIP-Seq Limitations

Sequencing and differential expression analysis involve correction using statistical models and tests, in particular normalizing for sequencing depth, library size, and library composition (Love et al. 2014). DESeq2 uses a negative binomial model to correct for these and a Wald test for determining differentially expressed transcripts. Differential expression analysis also assumes that genes with higher or lower expression are more important, but this is not always the case. One way to search for important genes that are not differentially expressed is weighted gene co-expression network analysis (WGCNA) though this can require a high number of biological samples (Zitnik et al. 2024). HOMER also uses DESeq2 for calculating differential peak counts for ChIP-Seq (Heinz et al. 2014).

Paused genes were determined the pausing index calculator (Day et al. 2016). The PI gives information on net values of occupying transcripts but does not give information on the stability and dynamics of pausing (i.e. it does not indicate whether peaks are from paused or terminated polymerase) though it is correlated with gene expression (Chen et al. 2018). This is not a limitation of the software but rather of the ChIP-seq method itself. This can be accounted for by using nucleotide run-on methods such as GRO-seq (global run-on sequencing) or PRO-seq (precision run on sequencing) although comparisons have found that the vast majority of Pol II peaks near promoters are from active polymerases engaged in elongation and capable of resuming synthesis (Core and Adelman, 2019). ChIP-Seq also does not acquire information about 3D chromatin conformation, unlike HiC, ChIA-PET, or HiChIP (Hou et al. 2024). A further difficulty for studying pausing is that drugs that can knockout pausing require 24 hours for complete effects but by this time have caused

massive side effects (Abuhashem et al. 2022). But recently, a new technique of using rapid protein degradation was used to show SPT5 is required for RNAP2 pausing (Aoi Y et al. 2021, Hu et al. 2021).

Arglu1 Further Experiments

There are several more general paths forward for studying Arglu1. One would be RNA-seq experiments in various human cells including neural cells. Single cell RNA-seq (scRNA-seq) could also determine if Arglu1 over expressing cells have a uniform expression pattern or can be clustered into distinct groups. Currently there has been no scRNA-seq study of Arglu1. It would also be interesting to investigate whether Arglu1 overexpression affects non-cancerous cells the same way as cancer cells. Determining the structure of Arglu1 could also increase knowledge of its mechanisms. It could also be informative to use a technique such as circular dichroism to determine with higher certainty the percentage of Arglu1 that is disordered (Miles et al. 2023). This could suggest that a part of Arglu1 does have a structure, as predicted by AlphaFold and others, suggesting part of it might be solvable by crystallization. Either way, knowing the structure of Arglu1 could lead to increasing knowledge of 3D confirmations it uses to bind and interact with proteins and RNA using molecular docking simulations (Holehouse and Kragelund 2024) or binding assays such as with NMR or isothermal titration calorimetry (González-Foutel et al. 2022).

Although I have made numerous cell lines overexpressing Arglu1 mutants, I have not made cell lines recapitulating the division of Magomedova et al. (2019) of Arglu1 into two parts,

an arginine rich beginning (3-74 aa) and a glutamate rich end (97-251 aa). Since smaller deletions often had minimal effects, it would be worth studying whether these larger deletions have larger effects on growth, DNA damage, pausing, etc. This could then be used to narrow in on testing smaller regions. However, it is possible that even these deletions might not have much effect. In which case, another path could be investigating the effect of sequence level features such as charge, etc., which are often more important for IDR function than the sequence itself (Holehouse and Kragelund 2024).

References

- Ablack, Jailal N. G., Michael Cohen, Gobi Thillainadesan, et al. 2012 Cellular GCN5 Is a Novel Regulator of Human Adenovirus E1A-Conserved Region 3 Transactivation. *Journal of Virology* 86(15): 8198–8209.
- Abuhashem A, Garg V, Hadjantonakis AK. RNA polymerase II pausing in development: orchestrating transcription. *Open Biol.* 2022 Jan;12(1):210220. doi: 10.1098/rsob.210220. Epub 2022 Jan 5. PMID: 34982944; PMCID: PMC8727152.
- Ackrill, A.M., Foster, G.R., Laxton, C.D., et al. (1991). Inhibition of the cellular response to interferons by products of the adenovirus type 5 E1A oncogene. *Nucleic Acids Res.* 19, 4387–4393.
- Adelman K, Lis JT. 2012 Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* 13, 720–731. (doi:10.1038/nrg3293)
- Ahlmann-Eltze and Anders 2019: proDA: Probabilistic Dropout Analysis for Identifying Differentially Abundant Proteins in Label-Free Mass Spectrometry. *bioRxiv* 661496 (Jun 2019)
- Alexander Dobin, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, Thomas R. Gingeras, STAR: ultrafast universal RNA-seq aligner, *Bioinformatics*, Volume 29, Issue 1, January 2013, Pages 15–21, <https://doi.org/10.1093/bioinformatics/bts635>
- Anthony M. Bolger, Marc Lohse, Bjoern Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, *Bioinformatics*, Volume 30, Issue 15, August 2014, Pages 2114–2120, <https://doi.org/10.1093/bioinformatics/btu170>
- Aoi et al. 2021 SPT5 stabilization of promoter-proximal RNA polymerase II. *Mol. Cell* 81, 4413–4424. doi:10.1016/j.molcel.2021.08.006
- Ashcroft M, Vousden KH 1999 Regulation of p53 stability. *Oncogene* 18:7637-7643
- Avvakumov, N., A. E. Kajon, R. C. Hoeben, and J. S. Mymryk 2004 Comprehensive Sequence Analysis of the E1A Proteins of Human and Simian Adenoviruses. *Virology* 329(2): 477–492.
- Avvakumov, Nikita, Majdina Sahbegovic, Zhiying Zhang, Michael Shuen, and Joe S. Mymryk 2002 Analysis of DNA Binding by the Adenovirus Type 5 E1A Oncoprotein. *The Journal of General Virology* 83(Pt 3): 517–524.
- Bacon and D’Orso 2019 CDK9: a signaling hub for transcriptional control. *Transcription* 10, 57–75. (doi:10.1080/21541264.2018.1523668)
- Bai et al. 2010 TIF1gamma controls erythroid cell fate by regulating transcription elongation. *Cell* 142, 133–143. (doi:10.1016/j.cell.2010.05.028)
- Berget, Susan M., Claire Moore, and Phillip A. Sharp 1977 Spliced Segments at the 5' Terminus of Adenovirus 2 Late MRNA. *Proceedings of the National Academy of Sciences* 74(8): 3171–3175.
- Berhane et al. 2011. Adenovirus E1A interacts directly with, and regulates the level of expression of, the immunoproteasome component MECL1. *Virology* 421:149 – 158. <https://doi.org/10.1016/j.virol.2011.09.025>.

- Berk, A. J., F. Lee, T. Harrison, J. Williams, and P. A. Sharp 1979 Pre-Early Adenovirus 5 Gene Product Regulates Synthesis of Early Viral Messenger RNAs. *Cell* 17(4): 935–944.
- Bowry, Akhil, Ann Liza Piberger, Patricia Rojas, Marco Saponaro, and Eva Petermann 2018 BET Inhibition Induces HEXIM1- and RAD51-Dependent Conflicts between Transcription and Replication. *Cell Reports* 25(8): 2061-2069.e4.
- Boyd, J. M., T. Subramanian, U. Schaeper, et al. 1993 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. *The EMBO Journal* 12(2): 469–478
- Bunch et al. 2014 TRIM28 regulates RNA polymerase II promoter-proximal pausing and pause release. *Nat. Struct. Mol. Biol.* 21, 876–883. (doi:10.1038/nsmb.2878)
- Burgert, H.G., Ruzsics, Z., Obermeier, S., et al. (2002). Subversion of host defense mechanisms by adenoviruses. *Curr. Top. Microbiol. Immunol.* 269, 273–318.
- Bushnell B BMap sourceforge.net/projects/bbmap/
- Cannizzaro et al. 2018 DDX3X RNA helicase affects breast cancer cell cycle progression by regulating expression of KLF4. *FEBS Lett.* 2018 Jul;592(13):2308-2322. doi: 10.1002/1873-3468.13106. Epub 2018 Jun 21. PMID: 29782654; PMCID: PMC6100109.
- Cartwright et al. 2005 LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. *Development* 132, 885–896. (doi:10.1242/dev.01670)
- Chan and Pek 2022 Distinct biogenesis pathways may have led to functional divergence of the human and Drosophila Arglu1 sisRNA, *EMBO Reports*
- Charman, Matthew, Christin Herrmann, and Matthew D. Weitzman 2019 Viral and Cellular Interactions During Adenovirus DNA Replication. *FEBS Letters* 593(24): 3531–3550.
- Chen et al. 2000, The amyloid precursor protein-binding protein APP-BP1 drives the cell cycle through the S-M checkpoint and causes apoptosis in neurons. *J Biol Chem.* 2000 Mar 24;275(12):8929-35. doi: 10.1074/jbc.275.12.8929.
- Chen FX, Smith ER, Shilatifard A. 2018 Born to run: control of transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* 19, 464–478. (doi:10.1038/s41580-018-0010-5)
- Chen G, Deng X. Cell Synchronization by Double Thymidine Block. *Bio Protoc.* 2018 Sep 5;8(17):e2994. doi: 10.21769/BioProtoc.2994
- Chow, Louise T., James M. Roberts, James B. Lewis, and Thomas R. Broker 1977 A Map of Cytoplasmic RNA Transcripts from Lytic Adenovirus Type 2, Determined by Electron Microscopy of RNA:DNA Hybrids. *Cell* 11(4): 819–836.
- Core and Adelman 2019, Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev.* 2019 Aug 1;33(15-16):960-982. doi: 10.1101/gad.325142.119. Epub 2019 May 23. PMID: 31123063
- Cramer P. 2019 Organization and regulation of gene transcription. *Nature* 573, 45–54. doi:10.1038/s41586-019-1517-4

- Crisostomo L, Soriano AM, Mendez M, Graves D, Pelka P. Temporal dynamics of human adenovirus 5 gene expression in normal human cells. *PLOS One*, 14(1): e0211192
- Crisostomo, Leandro, Andrea Michelle Soriano, Megan Mendez, Drayson Graves, and Peter Pelka 2019 Temporal Dynamics of Adenovirus 5 Gene Expression in Normal Human Cells. Michael Nevels, ed. *PLOS ONE* 14(1): e0211192.
- Cuconati, Andrea, and Eileen White 2002 Viral Homologs of BCL-2: Role of Apoptosis in the Regulation of Virus Infection. *Genes & Development* 16(19): 2465–2478.
- Cuellar et al. 2009, Identification and localization of novel genes preferentially expressed in human kidney glomerulus, *Nephrology*
- Culp, J. S., L. C. Webster, D. J. Friedman, et al. 1988 The 289-Amino Acid E1A Protein of Adenovirus Binds Zinc in a Region That Is Important for Trans-Activation. *Proceedings of the National Academy of Sciences of the United States of America* 85(17): 6450–6454.
- Daneshafrooz, N., Bagherzadeh Cham, M., Majidi, M. et al. Identification of potentially functional modules and diagnostic genes related to amyotrophic lateral sclerosis based on the WGCNA and LASSO algorithms. *Sci Rep* 12, 20144 (2022). <https://doi.org/10.1038/s41598-022-24306-2>
- Davey, Norman E., Gilles Travé, and Toby J. Gibson 2011 How Viruses Hijack Cell Regulation. *Trends in Biochemical Sciences* 36(3): 159–169.
- Davey, Norman E., Martha S. Cyert, and Alan M. Moses 2015 Short Linear Motifs - Ex Nihilo Evolution of Protein Regulation. *Cell Communication and Signaling: CCS* 13: 43.
- Davison, Andrew J., Mária Benkő, and Balázs Harrach 2003 Genetic Content and Evolution of Adenoviruses. *Journal of General Virology* 84(11): 2895–2908.
- Day, D. S., Zhang, B., Stevens, S. M., Ferrari, F., Larschan, E. N., Park, P. J., & Pu, W. T. (2016). Comprehensive analysis of promoter-proximal RNA polymerase II pausing across mammalian cell types. *Genome biology*, 17(1), 120.
- Dela Cruz, Charles S., Susan Pasnick, Jane E. Gross, et al. 2019 Adenovirus Infection and Outbreaks: What You Need to Know. *American Journal of Respiratory and Critical Care Medicine* 199(7): P13–P14.
- Dhingra, Akshay, Elias Hage, Tina Ganzenmueller, et al. 2019 Molecular Evolution of Human Adenovirus (HAdV) Species C. *Scientific Reports* 9(1): 1039.
- Ding, et al. 2023 Temporal modelling using single-cell transcriptomics. *Nat Rev Genet* **23**, 355–368 (2022). <https://doi.org/10.1038/s41576-021-00444-7>
- Doerfler, Walter, and Petra Böhm, eds. 1995 The Molecular Repertoire of Adenoviruses III: Biology and Pathogenesis, vol.199/3. *Current Topics in Microbiology and Immunology*. Berlin, Heidelberg: Springer
- Dolgacheva et al. 2019 Role of DJ-1 in the mechanism of pathogenesis of Parkinson's disease. *J Bioenerg Biomembr* 51, 175–188 (2019). <https://doi.org/10.1007/s10863-019-09798-4>

- Drumond-Bock, Ana Luiza, and Magdalena Bieniasz 2021 The Role of Distinct BRD4 Isoforms and Their Contribution to High-Grade Serous Ovarian Carcinoma Pathogenesis. *Molecular Cancer* 20: 145.
- Erickson et al. 2018 Dynamic turnover of paused Pol II complexes at human promoters. *Genes Dev.* 32, 1215–1225. doi:10.1101/gad.316810.118
- Etchegaray et al. 2019 The histone deacetylase SIRT6 restrains transcription elongation via promoter-proximal pausing. *Mol. Cell* 75, 683–699.e7. (doi:10.1016/j.molcel.2019.06.034)
- Fant et al. 2020 TFIID enables RNA polymerase II promoter-proximal pausing. *Mol. Cell* 78, 785–793.e8. (doi:10.1016/j.molcel.2020.03.008)
- Ferrari et al. 2008. Epigenetic reprogramming by adenovirus e1a. *Science* 321:1086–1088. <https://doi.org/10.1126/science.1155546>.
- Floyd, Scott R., Michael E. Pacold, Qiuying Huang, et al. 2013 The Bromodomain Protein Brd4 Insulates Chromatin from DNA Damage Signalling. *Nature* 498(7453): 246–250.
- Foulkes et al. 2014 DICER1: mutations, microRNAs and mechanisms. *Nat Rev Cancer* 14, 662–672 (2014). <https://doi.org/10.1038/nrc3802>
- Frost, Jasmine Rae, Megan Mendez, Andrea Michelle Soriano, et al. 2018 Adenovirus 5 E1A-Mediated Suppression of P53 via FUBP1. *Journal of Virology* 92(14): e00439-18.
- Gambin, Tomasz, Bo Yuan, Weimin Bi, et al. 2017 Identification of Novel Candidate Disease Genes from de Novo Exonic Copy Number Variants. *Genome Medicine* 9(1): 83.
- Gao, C.-H., Chen, C., Akyol, T., Duşa, A., Yu, G., Cao, B., and Cai, P. (2024). ggVennDiagram: intuitive Venn diagram software extended. *iMeta* 3, 69. doi: 10.1002/imt2.177.
- Gates et al. 2017 Acetylation on histone H3 lysine 9 mediates a switch from transcription initiation to elongation. *J. Biol. Chem.* 292, 14 456–14 472. (doi:10.1074/jbc.M117.802074)
- González-Foutel et al. 2022, Conformational buffering underlies functional selection in intrinsically disordered protein regions. *Nat Struct Mol Biol.* 2022 Aug;29(8):781-790. doi: 10.1038/s41594-022-00811-w. Epub 2022 Aug 10. PMID: 35948766
- Graves D, Akkerman N, Fulham L, Helwer R, Pelka P. Molecular insights into type I interferon suppression and enhanced pathogenicity by species B human adenoviruses B7 and B14. *mBio.* 2024 Aug 14;15(8):e0103824. doi: 10.1128/mbio.01038-24.
- Gregersen, Lea H., and Jesper Q. Svejstrup 2018 The Cellular Response to Transcription-Blocking DNA Damage. *Trends in Biochemical Sciences* 43(5): 327–341.
- Gressel et al. 2019 The pause-initiation limit restricts transcription activation in human cells. *Nat. Commun.* 10, 3603. doi:10.1038/s41467-019-11536-8

- Groelly et al. 2023, Targeting DNA damage response pathways in cancer. *Nat Rev Cancer*. Feb;23(2):78-94. doi: 10.1038/s41568-022-00535-5. Epub 2022 Dec 5. PMID: 36471053.
- Guffens et al. 2023, PME-1 sensitizes glioblastoma cells to oxidative stress-induced cell death by attenuating PP2A-B55 α -mediated inactivation of MAPKAPK2-RIPK1 signaling. *Cell Death Discov*. 9, 265 (2023). <https://doi.org/10.1038/s41420-023-01572-1>
- Gutch, M.J., and Reich, N.C. (1991). Repression of the interferon signal transduction pathway by the adenovirus E1A oncogene. *Proc. Natl. Acad. Sci. USA* 88, 7913–7917.
- Hale and Braithwaite 1999, The adenovirus oncoprotein E1a stimulates binding of transcription factor ETF to transcriptionally activate the p53 gene. *J Biol Chem* 274:23777-23786
- Hearing, 2021, *Adenoviridae: Their Viruses and Their Replication* in Howley and Knipe, 2021, *Fields Virology* 7th Ed. Vol 2 DNA Viruses, Wolters Kluwer Health
- Heim, Albert, Carmen Ebnet, Gabi Harste, and Patricia Pring-Åkerblom 2003 Rapid and Quantitative Detection of Human Adenovirus DNA by Real-Time PCR. *Journal of Medical Virology* 70(2): 228–239.
- Heinz S, Benner C, Spann N, Bertolino E et al. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol Cell* 2010 May 28;38(4):576-589. PMID: 20513432
- Hendrickx et al. 2014, Innate Immunity to Adenovirus, *Human Gene Therapy* 25:265–284 DOI: 10.1089/hum.2014.001
- Herlihy and de Bruin 2017 The role of the transcriptional response to DNA replication stress. *Genes* 8, 92
- Hilleman, M. R., and J. H. Werner 1954 Recovery of New Agent from Patients with Acute Respiratory Illness. *Experimental Biology and Medicine* 85(1): 183–188.
- Holehouse and Kragelund 2024, The molecular basis for cellular function of intrinsically disordered protein regions. *Nat Rev Mol Cell Biol*. 2024 Mar;25(3):187-211. doi: 10.1038/s41580-023-00673-0. Epub 2023 Nov 13. PMID: 37957331
- Holehouse and Kragelund 2024, The molecular basis for cellular function of intrinsically disordered protein regions. *Nat Rev Mol Cell Biol*. 2024 Mar;25(3):187-211. doi: 10.1038/s41580-023-00673-0. Epub 2023 Nov 13. PMID: 37957331; PMCID: PMC11459374.
- Horwitz et al. 2008. Adenovirus small e1a alters global patterns of histone modification. *Science* 321:1084 –1085.
- Hou et al. 2024, Best practices for ChIP-seq and its data analysis in Dluzen Schmidt eds. 2024, *Rigor and Reproducibility in Genetics and Genomics*, Academic Press, <https://doi.org/10.1016/C2018-0-01901-5>
- Howitt, J., C. W. Anderson, and P. Freimuth 2001 Adenovirus Interaction with its Cellular Receptor CAR. BNL-68839; 400412000. Brookhaven National Laboratory, Upton, NY (US). <https://www.osti.gov/biblio/789786>

- Hu et al. 2021 SPT5 stabilizes RNA polymerase II, orchestrates transcription cycles, and maintains the enhancer landscape. *Mol. Cell* 81, 4413–4424.e5. doi:10.1016/j.molcel.2021.08.029
- Hu, Y., Delviks-Frankenberry, K.A., Wu, C. *et al.* Structural insights into PPP2R5A degradation by HIV-1 Vif. *Nat Struct Mol Biol* **31**, 1492–1501 (2024). <https://doi.org/10.1038/s41594-024-01314-6>
- Huo, Dawei, Hao Chen, Yiming Cheng, et al. 2020 JMJD6 Modulates DNA Damage Response through Downregulating H4K16ac Independently of Its Enzymatic Activity. *Cell Death and Differentiation* 27(3): 1052–1066.
- Isobe et al. 2009, Adenovirus E1A inhibits SCF(Fbw7) ubiquitin ligase. *J Biol Chem* 284:27766–27779. <https://doi.org/10.1074/jbc.M109.006809>
- Jang, Moon Kyoo, Kazuki Mochizuki, Meisheng Zhou, et al. 2005 The Bromodomain Protein Brd4 Is a Positive Regulatory Component of P-TEFb and Stimulates RNA Polymerase II-Dependent Transcription. *Molecular Cell* 19(4): 523–534.
- Jones, N., and T. Shenk 1979 An Adenovirus Type 5 Early Gene Function Regulates Expression of Other Early Viral Genes. *Proceedings of the National Academy of Sciences of the United States of America* 76(8): 3665–3669.
- Jonkers et al. 2014 Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *Elife* 3, e02407. (doi:10.7554/eLife.02407)
- Jubin et al. 2016 The PARP family: insights into functional aspects of poly (ADP-ribose) polymerase-1 in cell growth and survival. *Cell Prolif.* 2016 Aug;49(4):421–37. doi: 10.1111/cpr.12268. Epub 2016 Jun 22. PMID: 27329285
- Jumper, J., Evans, R., Pritzel, A. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589 (2021). <https://doi.org/10.1038/s41586-021-03819-2>
- Jun Cheng et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science* 381, eadg7492 2023. DOI:10.1126/science.adg7492
- Kajon, A. E., J. B. Weinberg, and K. R. Spindler 2019 Adenoviruses, Reference Module in Biomedical Sciences. Elsevier. doi.org/10.1016/B978-0-12-801238-3.00086-6
- Kalvakolanu, D.V., Bandyopadhyay, S.K., Harter, M.L., et al. 1991. Inhibition of interferon-inducible gene expression by adenovirus E1A proteins: block in transcriptional complex formation. *Proc. Natl. Acad. Sci. USA* 88, 7459–7463.
- Kim and Sharpless 2006, The regulation of INK4/ARF in cancer and aging. *Cell*. Oct 20;127(2):265–75. doi: 10.1016/j.cell.2006.10.003. PMID: 17055429.
- Kimelman, D 1986 A Novel General Approach to Eucaryotic Mutagenesis Functionally Identifies Conserved Regions within the Adenovirus 13S E1A Polypeptide. *Molecular and Cellular Biology* 6(5): 1487–1496.
- King, Cason R., Ali Zhang, Tanner M. Tessier, Steven F. Gameiro, and Joe S. Mymryk 2018 Hacking the Cell: Network Intrusion and Exploitation by Adenovirus E1A. Michael J. Imperiale and Danielle A. Garsin, eds. *MBio* 9(3): e00390-18, /mbio/9/3/mBio.00390-18.atom.

- Koedoot E, Smid M, Foekens JA, Martens JW, Le Dev_ edec SE, van de Water B. Co-regulated gene expression of splicing factors as drivers of cancer progression. *Sci Rep* 2019;9(1):1–13.
- Kolonin and Finley 2000, A role for cyclin J in the rapid nuclear division cycles of early *Drosophila* embryogenesis. *Dev Biol.* 2000 Nov 15;227(2):661–72. doi: 10.1006/dbio.2000.9916. PMID: 11071782.
- Kotsantis, P., Petermann, E. & Boulton, S. J. Mechanisms of oncogene-induced replication stress: jigsaw falling into place. *Cancer Discov.* 8, 537–555 (2018).
- Krebs et al. 2017 Genome-wide single-molecule footprinting reveals high RNA polymerase II turnover at paused promoters. *Mol. Cell* 67, 411–422.e4. (doi:10.1016/j.molcel.2017.06.027)
- Kremer, Eric J., and Glen R. Nemerow 2015 Adenovirus Tales: From the Cell Surface to the Nuclear Pore Complex. *PLoS Pathogens* 11(6): e1004821.
- Krueger, Brian J., Katayoun Varzavand, Jeffrey J. Cooper, and David H. Price 2010 The Mechanism of Release of P-TEFb and HEXIM1 from the 7SK SnRNP by Viral and Cellular Activators Includes a Conformational Change in 7SK. *PloS One* 5(8):e12335.
- Labuzan et al. 2020, Inhibition of protein phosphatase methylesterase 1 dysregulates MAP kinase signaling and attenuates muscle cell differentiation. *Gene.* 2020 May 20;739:144515. doi: 10.1016/j.gene.2020.144515. Epub 2020 Feb 26. PMID: 32112987.
- Lau et al. 2015. DNA tumor virus oncogenes antagonize the cGAS-STING DNA-sensing pathway. *Science* 350:568 –571. <https://doi.org/10.1126/science.aab3291>.
- Li, L., Han, W., Chen, Y. et al. MiR-3613-3p inhibits hypertrophic scar formation by down-regulating arginine and glutamate-rich 1. *Mol Cell Biochem* 476, 1025–1036 (2021). <https://doi.org/10.1007/s11010-020-03968-4>
- Liis Kolberg, Uku Raudvere, Ivan Kuzmin, Priit Adler, Jaak Vilo, Hedi Peterson: g:Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update) *Nucleic Acids Research*, May 2023; doi:10.1093/nar/gkad347
- Lillie, James W., Paul M. Loewenstein, Michael R. Green, and Maurice Green 1987 Functional Domains of Adenovirus Type 5 E1a Proteins. *Cell* 50(7): 1091–1100.
- Lion and Wold 2021, Adenoviruses in Howley and Knipe 2021, *Fields Virology* 7th Ed. Vol 2 DNA Viruses, Wolters Kluwer Health
- Liu et al. 1990, A specific member of the ATF transcription factor family can mediate transcription activation by the adenovirus E1a protein. *Cell* 61:1217–1224. [https://doi.org/10.1016/0092-8674\(90\)90686-9](https://doi.org/10.1016/0092-8674(90)90686-9).
- Liu et al. 2015 Ready, pause, go: regulation of RNA polymerase II pausing and release by cellular signaling pathways. *Trends Biochem. Sci.* 40, 516–525. (doi:10.1016/j.tibs.2015.07.003)
- Liu, Fang, and Michael R. Green 1994 Promoter Targeting by Adenovirus E1a through Interaction with Different Cellular DNA-Binding Domains. *Nature* 368(6471): 520–525.

- Liu, Wen, Qi Ma, Kaki Wong, et al. 2013 Brd4 and JMJD6-Associated Anti-Pause Enhancers in Regulation of Transcriptional Pause Release. *Cell* 155(7): 1581–1595.
- Liu, Xin, and Ronen Marmorstein 2007 Structure of the Retinoblastoma Protein Bound to Adenovirus E1A Reveals the Molecular Basis for Viral Oncoprotein Inactivation of a Tumor Suppressor. *Genes & Development* 21(21): 2711–2716.
- Love MI, Huber W, Anders S (2014). “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.” *Genome Biology*, 15, 550. doi:10.1186/s13059-014-0550-8.
- Lovén, Jakob, Heather A. Hoke, Charles Y. Lin, et al. 2013 Selective Inhibition of Tumor Oncogenes by Disruption of Super-Enhancers. *Cell* 153(2): 320–334.
- Ludt A, Ustjanzew A, Binder H, Strauch K, Marini F (2022). “Interactive and Reproducible Workflows for Exploring and Modeling RNA-seq Data with pcaExplorer, ideal, and GeneTonic.” *Current Protocols*, 2(4), e411. doi:10.1002/cpz1.411.
- Luftig, R B, and R R Weihing 1975 Adenovirus Binds to Rat Brain Microtubules in Vitro. *Journal of Virology* 16(3): 696–706.
- Magomedova, Lilia, Jens Tiefenbach, Emma Zilberman, et al. 2019 ARGLU1 Is a Transcriptional Coactivator and Splicing Regulator Important for Stress Hormone Signaling and Development. *Nucleic Acids Research* 47(6): 2856–2870.
- MASCOT Matrix Science
- Matthews et al. 2023 Cell cycle control in cancer. *Nat Rev Mol Cell Biol* 23, 74–88 (2022). <https://doi.org/10.1038/s41580-021-00404-3>
- Matthews, H.K., Bertoli, C. & de Bruin, R.A.M. Cell cycle control in cancer. *Nat Rev Mol Cell Biol* 23, 74–88 (2022). <https://doi.org/10.1038/s41580-021-00404-3>
- Mihaly Varadi, et al. AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences, *Nucleic Acids Research*, Volume 52, Issue D1, 5 January 2024, Pages D368–D375, <https://doi.org/10.1093/nar/gkad1011>
- Miles et al 2023, DichroIDP: a method for analyses of intrinsically disordered proteins using circular dichroism spectroscopy. *Commun Biol* 6, 823 (2023). <https://doi.org/10.1038/s42003-023-05178-2>
- Min, S., Ji, JH., Heo, Y. et al. Transcriptional regulation and chromatin dynamics at DNA double-strand breaks. *Exp Mol Med* 54, 1705–1712 (2022). <https://doi.org/10.1038/s12276-022-00862-5>
- Mitra et. a. 2012 Estrogen receptor-α recruits P-TEFb to overcome transcriptional pausing in intron 1 of the MYB gene. *Nucleic Acids Res.* 40, 5988–6000. (doi:10.1093/nar/gks286)
- Murray, Vincent, Jon K. Chen, and Long H. Chung 2018 The Interaction of the Metallo-Glycopeptide Anti-Tumour Drug Bleomycin with DNA. *International Journal of Molecular Sciences* 19(5): 1372.
- Nakato R., Sakata T., Methods for ChIP-seq analysis: A practical workflow and advanced applications, *Methods*, 2020.

- Nguyen M, Branton P, Roy S, Nicholson D, Alnemri E, Yeh W, Mak T, Shore G 1998 E1A-induced processing of procaspase-8 can occur independently of F ADD and is inhibited by Bcl-2. *J Biol Chem* 273:33099-33102
- Nguyen, V. T., T. Kiss, A. A. Michels, and O. Bensaude 2001 7SK Small Nuclear RNA Binds to and Inhibits the Activity of CDK9/Cyclin T Complexes. *Nature* 414(6861): 322–325.
- Olanubi O, Frost JR, Radko S, Pelka P. Suppression of Type I interferon response by E1A via RuvBL1/Pontin, *J Virol*, 91(8): e02484-16,2017.
- Park E, Gong E-Y, Romanelli MG, Lee K. 2012 Suppression of estrogen receptor- α transactivation by thyroid transcription factor-2 in breast cancer cells. *Biochem Biophys Res Commun*; 421(3):532–7. [16]
- Pelka P, Ablack JNG, Torchia J, Turnell AS, Grand RJA, Mymryk JS. 2009, Transcriptional control by adenovirus E1A conserved region 3 via p300/CBP, *Nucleic Acids Res* 37: 1095-1106
- Pelka, P., J. N. G. Ablack, G. J. Fonseca, A. F. Yousef, and J. S. Mymryk 2008 Intrinsic Structural Disorder in Adenovirus E1A: A Viral Molecular Hub Linking Multiple Diverse Processes. *Journal of Virology* 82(15): 7252–7263.
- Pied, Noémie, and Harald Wodrich 2019 Imaging the Adenovirus Infection Cycle. *FEBS Letters* 593(24): 3419–3448.
- Pirnie, Stephan P., Ahmad Osman, Yinzhou Zhu, and Gordon G. Carmichael 2017 An Ultraconserved Element (UCE) Controls Homeostatic Splicing of ARGLU1 mRNA. *Nucleic Acids Research* 45(6): 3473–3486
- Portal, Daniel, Bo Zhao, Michael A. Calderwood, et al. 2011 EBV Nuclear Antigen EBNA1 Dismisses Transcription Repressors NCoR and RBPJ from Enhancers and EBNA2 Increases NCoR-Deficient RBPJ DNA Binding. *Proceedings of the National Academy of Sciences of the United States of America* 108(19): 7808–7813.
- Prusinkiewicz, Martin A., Jessie Tu, Mackenzie J. Dodge, et al. 2020 Differential Effects of Human Adenovirus E1A Protein Isoforms on Aerobic Glycolysis in A549 Human Lung Epithelial Cells. *Viruses* 12(6): 610
- Quinlan, Kate G. R., Alexis Verger, Alister Kwok, et al. 2006 Role of the C-Terminal Binding Protein PxDLS Motif Binding Cleft in Protein Interactions and Transcriptional Repression. *Molecular and Cellular Biology* 26(21): 8202–8213.
- Radko, Sandi, Maria Koleva, Kris M. D. James, et al. 2014 Adenovirus E1A Targets the DREF Nuclear Factor to Regulate Virus Gene Expression, DNA Replication, and Growth. *Journal of Virology* 88(22): 13469–13481.
- Radko, Sandi, Richard Jung, Oladunni Olanubi, and Peter Pelka 2015 Effects of Adenovirus Type 5 E1A Isoforms on Viral Replication in Arrested Human Cells. *PLoS ONE* 10(10)
- Rahl et al. 2010 c-Myc regulates transcriptional pause release. *Cell* 141,432–445. (doi:10.1016/j.cell.2010.03.030)
- Rhyasen, Garrett W., Yi Yao, Jingwen Zhang, et al. 2018 BRD4 Amplification Facilitates an Oncogenic Gene Expression Program in High- Grade Serous Ovarian

Cancer and Confers Sensitivity to BET Inhibitors. Eric Asselin, ed. PLOS ONE 13(7): e0200826.

- Rothkamm, Kai, Stephen Barnard, Jayne Moquet, et al. 2015 DNA Damage Foci: Meaning and Significance. *Environmental and Molecular Mutagenesis* 56(6): 491–504.
- Rowe, W. P., R. J. Huebner, L. K. Gilmore, R. H. Parrott, and T. G. Ward 1953 Isolation of a Cytopathogenic Agent from Human Adenoids Undergoing Spontaneous Degeneration in Tissue Culture. *Experimental Biology and Medicine* 84(3): 570–573.
- Sacta MA et al. 2018 Gene-specific mechanisms direct glucocorticoid-receptor-driven repression of inflammatory response genes in macrophages. *Elife* 7, e34864. (doi:10.7554/eLife.34864)
- Schaeper, U., J. M. Boyd, S. Verma, et al. 1995a Molecular Cloning and Characterization of a Cellular Phosphoprotein That Interacts with a Conserved C-Terminal Domain of Adenovirus E1A Involved in Negative Modulation of Oncogenic Transformation. *Proceedings of the National Academy of Sciences* 92(23): 10467–10471.
- Schaeper, U., J. M. Boyd, S. Verma, et al. 1995b Molecular Cloning and Characterization of a Cellular Phosphoprotein That Interacts with a Conserved C-Terminal Domain of Adenovirus E1A Involved in Negative Modulation of Oncogenic Transformation. *Proceedings of the National Academy of Sciences* 92(23): 10467–10471.
- Schmitz-Winnenthal F.H., Galindo-Escobedo L.V., Rimoldi D., Geng W., Romero P., Koch M., et al. Potential target antigens for immunotherapy in human pancreatic cancer. 2007;252(2):290_8.
- Scott Bachus, Nikolas Akkerman, Lauren Fulham, Drayson Graves, Rafe Helwer, Jordan Rempel, Peter Pelka, ARGLU1 enhances promoter-proximal pausing of RNA polymerase II and stimulates DNA damage repair, *Nucleic Acids Research*, Volume 52, Issue 10, 10 June 2024, Pages 5658–5675, <https://doi.org/10.1093/nar/gkae208>
- Shao R, Hu M, Zhou B, Lin S, Chiao P. von Lindern R. Spohn B, Hung M (1999) E1A sensitizes cells to tumor necrosis factor-induced apoptosis through inhibition of IKB kinases and nuclear factor KB activities. *J Biol Chem* 274:21495-21498
- Shi J, Yu X, Li G, Zhao X, Chen J, Fang Y, Yang Y, Wang T, Xu T, Bian L, Lyu L, He Y. DTL promotes head and neck squamous cell carcinoma progression by mediating the degradation of ARGLU1 to regulate the Notch signaling pathway. *Int J Biol Macromol*. 2024 Feb;259(Pt 2):129184. doi: 10.1016/j.ijbiomac.2023.129184
- Shisler J. Duerksen-Hughes P, Hermiston T. Wold W. Gooding L 1996 Induction of susceptibility to tumor necrosis factor by E1A is dependent on binding to either p300 or p105-Rb and induction of DNA synthesis. *J Virol* 70:68-77
- Sobell HM. Actinomycin and DNA transcription. *Proc Natl Acad Sci U S A*. 1985 Aug;82(16):5328-31. doi: 10.1073/pnas.82.16.5328
- Ström et al. 1998, AR1 is an integral part of the adenovirus type 2 E1A-CR3 transactivation domain. *J Virol* 72: 5978 –5983.

- Sun et al. 2016, Regulation of p27Kip1 phosphorylation and G1 cell cycle progression by protein phosphatase PPM1G. *Am J Cancer Res.* 2016 Oct 1;6(10):2207-2220. PMID: 27822412; PMCID: PMC5088286.
- Svensson, C., M. Bondesson, E. Nyberg, et al. 1991 Independent Transformation Activity by Adenovirus-5 E1A-Conserved Regions 1 or 2 Mutants. *Virology* 182(2): 553–561.
- Täuber, Birgitt, and Thomas Dobner 2001 Adenovirus Early E4 Genes in Viral Oncogenesis. *Oncogene* 20(54). Nature Publishing Group: 7847–7854.
- Teodoro JG. Shore GC. Branton PE 1995 Adenovirus E1A proteins induce apoptosis by both p53-dependent and p53-independent mechanisms. *Oncogene* 11:467 474
- Tessier, Tanner M., Mackenzie J. Dodge, Katelyn M. MacNeil, et al. 2021 Almost Famous: Human Adenoviruses (and What They Have Taught Us about Cancer). *Tumour Virus Research* 12: 200225.
- Thomas A. White E 1998 Suppression of the p300-dependent mdm2 negative-feedback loop induces the p53 apoptotic function. *Genes Dev* 12: 197519X5
- Trapnell, C., Williams, B., Pertea, G. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28, 511–515 (2010). <https://doi.org/10.1038/nbt.1621>
- Vachon, Virginia K., and Graeme L. Conn 2016 Adenovirus VA RNA: An Essential pro-Viral Non-Coding RNA. *Virus Research* 212: 39–52.
- Vaid et al. 2020 Release of promoter-proximal paused Pol II in response to histone deacetylase inhibition. *Nucleic Acids Res.* 48, 4877–4890. (doi:10.1093/nar/gkaa234)
- van den Heuvel D, van der Weegen Y, Boer DEC, Ogi T, Luijsterburg MS. Transcription-Coupled DNA Repair: From Mechanism to Human Disorder. *Trends Cell Biol.* 2021 May;31(5):359-371. doi: 10.1016/j.tcb.2021.02.007
- Vijayalingam et al. 2013 Adenovirus L-E1A activates transcription through mediator complex-dependent recruitment of the super elongation complex. *J Virol*
- Voß, H., Schlumbohm, S., Barwikowski, P. et al. HarmonizR enables data harmonization across independent proteomic datasets with appropriate handling of missing values. *Nat Commun* 13, 3523 (2022). <https://doi.org/10.1038/s41467-022-31007-x>
- Vos et al. 2018a Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. *Nature* 560, 607–612. (doi:10.1038/s41586-018-0440-4)
- Vos et al. 2018b Structure of paused transcription complex Pol II-DSIF-NELF. *Nature* 560, 601–606. (doi:10.1038/s41586-018-0442-2)
- Wada, T, T Takagi, Y Yamaguchi, D Watanabe, and H Handa 1998 Evidence That P-TEFb Alleviates the Negative Effect of DSIF on RNA Polymerase II-Dependent Transcription in Vitro. *The EMBO Journal* 17(24): 7395–7403.
- Walker, Peter J., Stuart G. Siddell, Elliot J. Lefkowitz, et al. 2020 Changes to Virus Taxonomy and the Statutes Ratified by the International Committee on Taxonomy of Viruses (2020). *Archives of Virology* 165(11): 2737–2748.

- Wang et al. 2023 Nuclear DJ-1 Regulates DNA Damage Repair via the Regulation of PARP1 Activity. *Int J Mol Sci.* 2023 May 12;24(10):8651. doi: 10.3390/ijms24108651. PMID: 37239999; PMCID: PMC10218208.
- Wang, et al. 2023 Bromodomain and extraterminal (BET) proteins: biological functions, diseases and targeted therapy. *Sig Transduct Target Ther* **8**, 420 (2023). <https://doi.org/10.1038/s41392-023-01647-6>
- Wang, Ranran, June F. Yang, Flora Ho, Erle S. Robertson, and Jianxin You 2020 Bromodomain-Containing Protein BRD4 Is Hyperphosphorylated in Mitosis. *Cancers* 12(6): 1637.
- Welch and Tsai 2022, Mechanisms of DNA damage-mediated neurotoxicity in neurodegenerative disease. *EMBO Rep.* 2022 Jun 7;23(6):e54217. doi: 10.15252/embr.202154217. Epub 2022 May 2. PMID: 35499251
- Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>.
- Wickham, Thomas J., Patricia Mathias, David A. Cheresh, and Glen R. Nemerow 1993 Integrins $\text{Av}\beta 3$ and $\text{Av}\beta 5$ Promote Adenovirus Internalization but Not Virus Attachment. *Cell* 73(2): 309–319.
- Wlodarchak, and Xing, 2016. PP2A as a master regulator of the cell cycle. *Critical Reviews in Biochemistry and Molecular Biology*, 51(3), 162–184. <https://doi.org/10.3109/10409238.2016.1143913>
- Wold WSM, Hermiston TW, Tollefson AE (1995) E3 transcription unit of adenovirus. *Curr Top Microbiol Immunol* 199/1:237-274
- Yang, Z., Q. Zhu, K. Luo, and Q. Zhou 2001 The 7SK Small Nuclear RNA Inhibits the CDK9/Cyclin T1 Kinase to Control Transcription. *Nature* 414(6861): 317–322.
- Yang, Zhiyuan, Jasper H. N. Yik, Ruichuan Chen, et al. 2005 Recruitment of P-TEFb for Stimulation of Transcriptional Elongation by the Bromodomain Protein Brd4. *Molecular Cell* 19(4): 535–545.
- Yao, F., Huang, S., Liu, J. et al. Deletion of ARGLU1 causes global defects in alternative splicing in vivo and mouse cortical malformations primarily via apoptosis. *Cell Death Dis* 14, 543 (2023). <https://doi.org/10.1038/s41419-023-06071-w>
- Ye F, Wu P, Zhu Y, Huang G, Tao Y, Liao Z, Guan Y. Construction of the prognostic signature of alternative splicing revealed the prognostic predictor and immune microenvironment in head and neck squamous cell carcinoma. *Front Genet.* 2022 Oct 21;13:989081. doi: 10.3389/fgene.2022.989081.
- Yousef et al. 2009 Requirements for E1A dependent transcription in the yeast *Saccharomyces cerevisiae*. *BMC Mol Biol*
- Yousef, A. F., G. J. Fonseca, P. Pelka, et al. 2010 Identification of a Molecular Recognition Feature in the E1A Oncoprotein That Binds the SUMO Conjugase UBC9 and Likely Interferes with PolySUMOylation. *Oncogene* 29(33). Nature Publishing Group: 4693–4704.

- Yu et al. 2020 Negative elongation factor complex enables macrophage inflammatory responses by controlling anti-inflammatory gene expression. *Nat. Commun.* 11, 2286. (doi:10.1038/s41467-020-16209-5)
- Yu G (2024). “Thirteen years of clusterProfiler.” *The Innovation*, 5(6), 100722. doi:10.1016/j.xinn.2024.100722, <https://doi.org/10.1016/j.xinn.2024.100722>.
- Zhang et al. 2019 Rs6265 polymorphism in brain-derived neurotrophic factor (Val/Val and Val/Met) promotes proliferation of bladder cancer cells by suppressing microRNA-205 and enhancing expression of cyclin J. *J Cell Biochem.* 2019 May;120(5):7297-7308. doi: 10.1002/jcb.28004. Epub 2018 Nov 1. PMID: 30387205.
- Zhang, Dingxiao, Pingping Jiang, Qinqin Xu, and Xiaoting Zhang 2011 Arginine and Glutamate-Rich 1 (ARGLU1) Interacts with Mediator Subunit 1 (MED1) and Is Required for Estrogen Receptor-Mediated Gene Transcription and Breast Cancer Cell Growth. *Journal of Biological Chemistry* 286(20): 17746–17754.
- Zhang, Jingwen, Austin M. Dulak, Maureen M. Hattersley, et al. 2018 BRD4 Facilitates Replication Stress-Induced DNA Damage Response. *Oncogene* 37(28): 3763–3777.
- Zhao et al. 2020, Adenovirus in the omics era - a multipronged strategy. *FEBS Lett.* 2020 Jun;594(12):1879-1890. doi: 10.1002/1873-3468.13710. Epub 2020 Jan 26. PMID: 31811727.
- Zhao, L.-J., T. Subramanian, and G. Chinnadurai 2008 Inhibition of Transcriptional Activation and Cell Proliferation Activities of Adenovirus E1A by the Unique N-Terminal Domain of CtBP2. *Oncogene* 27(39): 5214–5222.
- Zitnik et al. 2024 Current and future directions in network biology. *Bioinform Adv.* Aug 14;4(1):vbae099. doi: 10.1093/bioadv/vbae099. PMID: 39143982; PMCID: PMC11321866.
- Zuber, Johannes, Junwei Shi, Eric Wang, et al. 2011 RNAi Screen Identifies Brd4 as a Therapeutic Target in Acute Myeloid Leukaemia. *Nature* 478(7370): 524–528.

Appendix

Figures

1 Adenovirus E1A RNA-Seq

WT.UI genes	WT.UI padj	13S.UI genes	13S.UI padj	12S.UI genes	12S.UI padj	YCD.UI genes	YCD.UI padj	13S.WT genes	13S.WT padj
CDC25 A	4.87E-278	HSPA1B	1.38E-302	CCNE2	7.17E-242	NAMPT	2.37E-290	HSPA6	3.21E-201
CCNE2	4.87E-278	RN7SL4 71P	1.38E-302	CDC25 A	2.37E-207	PDE5A	2.37E-290	E2F2	3.73E-29
HSPA1B	4.87E-278	CDC25 A	1.38E-300	ZNF367	2.17E-152	CXCL5	2.37E-290	HSPA1B	2.46E-26
RN7SL4 71P	4.87E-278	CCNE2	1.33E-280	BRI3BP	8.64E-146	IL7R	2.37E-290	BLM	1.42E-25
UHRF1	4.87E-276	CCNE1	4.45E-253	CCNE1	4.86E-142	TNFSF1 5	2.37E-290	HSPA1A	4.41E-20
DTL	1.95E-271	HSPA1A	4.45E-253	WDR76	1.44E-137	ADH1B	2.37E-290	EDNRA	1.17E-19
CCNE1	3.16E-271	UHRF1	1.38E-204	DTL	4.72E-133	HSPA1B	2.37E-290	CDCA5	2.84E-19
ZNF367	5.83E-259	PAQR4	2.65E-201	UHRF1	2.14E-130	INHBA	2.37E-288	ASPM	3.67E-18
MCM7	7.59E-255	MCM7	1.66E-199	HSPA1B	5.21E-127	BMP2	4.96E-288	DDIT4	5.14E-18
WDR76	1.46E-247	SYNM	2.45E-196	SLFN11	2.08E-114	KIAA11 99	6.19E-268	ZDHHC 23	5.14E-18
PAQR4	6.92E-221	NR4A1	6.55E-192	CDCA7	5.50E-110	TXNIP	8.01E-255	MYBL2	5.33E-18
SYNM	1.36E-213	DTL	2.49E-185	PAQR4	1.90E-107	SOD2	6.66E-250	DIO2	2.83E-17
BRI3BP	1.09E-212	SPHK1	4.39E-182	FOS	9.60E-104	BNIP3	2.03E-247	NAV3	3.01E-17
PPAT	7.76E-210	ZNF367	1.05E-181	HSPA4L	1.11E-98	RN7SL4 71P	6.83E-246	KIAA11 99	4.01E-17
PPAP2B	3.17E-190	FOS	3.54E-174	CDT1	1.11E-98	DIO2	2.65E-234	KIF2C	9.90E-17
MCM4	6.70E-190	PPAT	2.08E-173	MCM2	2.44E-97	LDHA	9.07E-225	DNAJB1	2.67E-16
RNF138	4.26E-189	WDR76	8.17E-172	MCM3	5.55E-97	HSPA1A	1.53E-224	BRCA1	2.85E-16
DDX21	1.98E-188	MIR17H G	9.50E-172	CDC7	7.37E-96	RGS4	3.04E-212	TRIB1	2.85E-16
MCM6	8.71E-185	CKB	4.90E-156	MCM7	4.74E-93	FJX1	1.29E-194	ZCCHC 24	5.24E-16
MCM3	2.00E-184	DDX21	1.83E-154	SYNM	8.28E-93	FZD8	7.88E-194	B3GAL NT1	5.38E-16
CDCA7	2.00E-184	RNF138	5.09E-149	PPAT	3.12E-92	APCDD 1L	5.05E-188	C8orf4	6.10E-16
FOS	1.20E-179	MB21D 1	1.28E-148	CHAF1 B	4.45E-91	SPHK1	7.24E-186	TMEM1 19	6.18E-16

CKB	2.73E-178	BRI3BP	3.41E-147	MCM6	1.33E-86	PTGS2	1.03E-184	SLC7A1 1	7.48E-16
MIR17HG	9.35E-178	NCR3LG1	5.61E-145	MCM10	4.00E-86	ANGPTL 2	3.07E-184	ZNF530	7.71E-16
RRP1B	2.08E-177	NETO2	1.02E-142	ORC1	1.47E-85	DDX21	3.44E-180	CHAF1 A	1.17E-15

Table S1: Top 25 genes by adjusted pvalue, samples 1-5

12S.WT genes	12S.WT padj	YCD.W T genes	YCD.W T padj	13S.12S genes	13S.12 S padj	YCD.13 S genes	YCD.13 S padj	YCD.12 S genes	YCD.12 S padj
HSPA6	8.94E-82	KIAA1199	4.66E-294	HSPA6	6.85E-258	CXCL5	2.88E-264	KIAA1199	1.11E-251
RN7SL471P	3.70E-77	INHBA	4.66E-294	HSPA1B	1.45E-191	CXCL1	2.88E-264	NAMPT	1.11E-251
HSPA1B	6.55E-73	CXCL5	4.66E-294	HSPA1A	2.93E-147	INHBA	2.88E-262	INHBA	1.11E-251
HSPA1A	6.61E-56	CXCL1	4.66E-294	RN7SL471P	4.56E-131	KIAA1199	5.13E-258	CXCL5	1.11E-251
TXNIP	2.16E-53	TNFSF15	4.66E-294	HNRNPU-AS1	2.09E-85	TNFSF15	5.10E-215	IL7R	1.11E-251
HNRNPU-AS1	3.68E-51	IL7R	4.66E-292	RN7SL3	3.66E-74	CXCL6	2.67E-200	TNFSF15	1.11E-251
MIR17HG	4.99E-50	NAMPT	1.07E-239	MYC	5.55E-67	NAMPT	1.20E-179	CXCL1	1.11E-249
CHD7	4.39E-48	CCL2	8.73E-199	NR4A1	1.07E-54	CCL2	1.71E-177	BMP2	4.88E-229
RN7SL3	9.17E-48	BNIP3	2.11E-194	SPHK1	8.57E-54	IL7R	4.88E-177	BNIP3	1.20E-228
SPHK1	1.83E-45	IL1B	8.26E-174	DIO2	7.47E-50	BNIP3	1.80E-165	SLC16A6	1.04E-225
MCM7	9.59E-41	CXCL6	9.20E-172	RGS4	3.78E-49	CXCL3	1.02E-154	HSPA6	2.49E-225
MIR155HG	1.29E-40	WDR76	1.73E-161	TRIB1	1.37E-46	SOD2	3.95E-154	DIO2	9.00E-220
HPDL	1.67E-40	TNFAIP3	1.95E-159	MIR17HG	2.67E-46	TNFAIP3	1.88E-131	RGS4	5.28E-197
CDC25A	3.29E-40	SMOX	6.25E-156	FOXL1	8.20E-45	IL1B	1.05E-126	IER3	5.28E-197
WEE1	5.65E-39	CCNE2	2.67E-154	RN7SL832P	1.03E-44	IL8	2.26E-114	IL1B	4.48E-187
RPP25	1.89E-36	NAV3	2.87E-154	TXNIP	1.54E-44	CCNE2	8.04E-101	SOD2	9.19E-185
KIAA0907	3.90E-36	DIO2	3.28E-153	MIR155HG	1.34E-43	WDR76	1.20E-99	TNFAIP3	4.26E-184
RRP1B	4.82E-36	SLC16A6	2.87E-150	SOCS3	4.91E-43	NDRG1	5.88E-96	SMOX	2.35E-178
SYNM	4.80E-35	SOD2	2.30E-145	IL7R	1.87E-42	SLC16A6	1.06E-94	BDKRB1	5.13E-177
HIST1H3PS1	8.07E-35	DTL	8.83E-144	HIST1H3PS1	8.35E-42	SMOX	3.36E-89	HSPA1B	6.10E-174

RN7SL83 2P	1.46E- 34	CXCL3	1.59E- 143	RPLP0P2	2.35E- 37	BMP2	1.50E- 87	CCL2	2.46E- 173
RASEF	1.49E- 34	HSPA6	4.62E- 142	C8orf4	2.38E- 37	CXCL2	2.43E- 84	APCDD 1L	1.83E- 161
BTN1A1	2.69E- 33	BMP2	4.34E- 136	TNFSF15	1.03E- 36	NFKB2	2.81E- 81	C8orf4	2.22E- 154
RP11- 304L19.5	7.17E- 33	IER3	1.17E- 131	LANCL1	1.12E- 36	IER3	4.48E- 81	LIF	1.56E- 153
CCNE1	1.85E- 32	ATAD2	1.90E- 129	APCDD1 L	1.37E- 36	DTL	4.66E- 79	TRIB1	2.39E- 152

Table S2: Top 25 genes by adjusted pvalue, samples 6-10

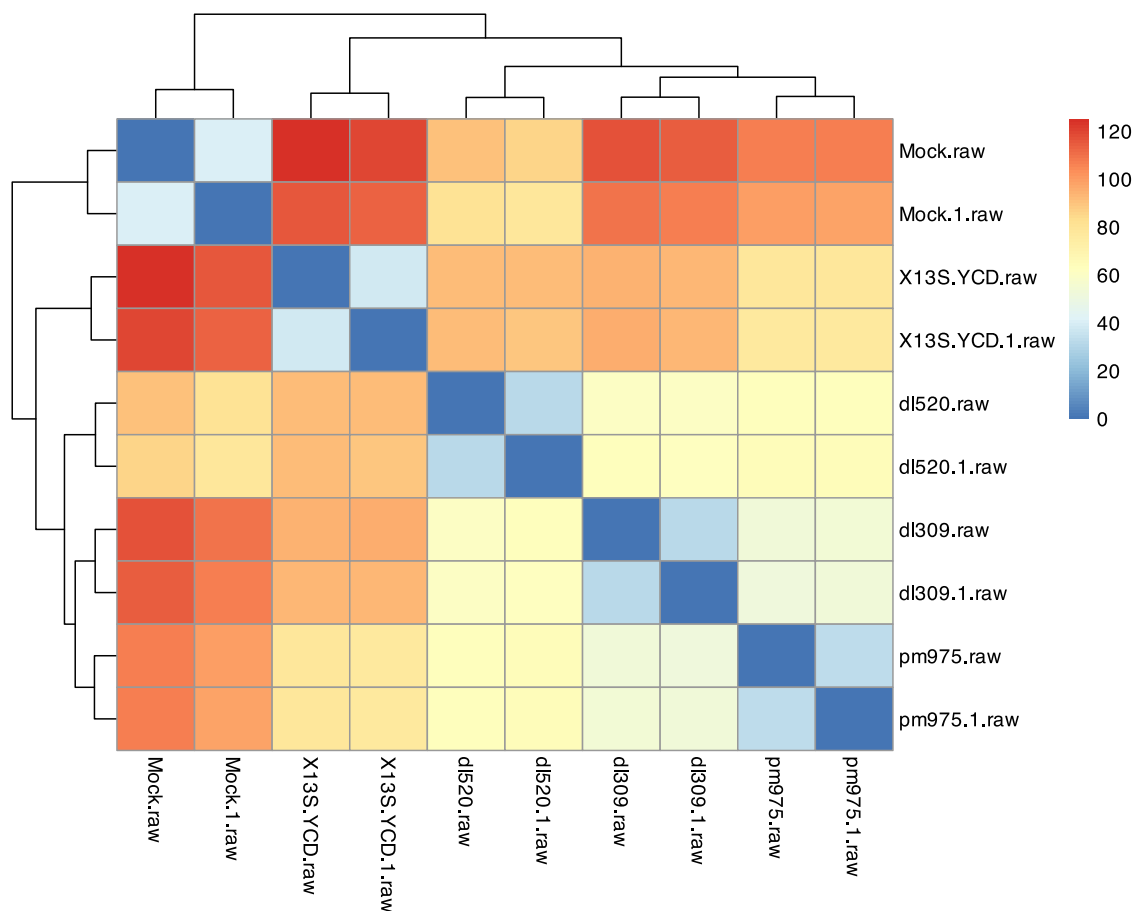


Figure S1: Sample to Sample Heat Map. Heat map showing the differences between each replicate. Blue indicates similarity (i.e. the diagonal comparing the same samples are the most blue), red indicates difference (the samples that are the most different are Mock0 and YCD0).

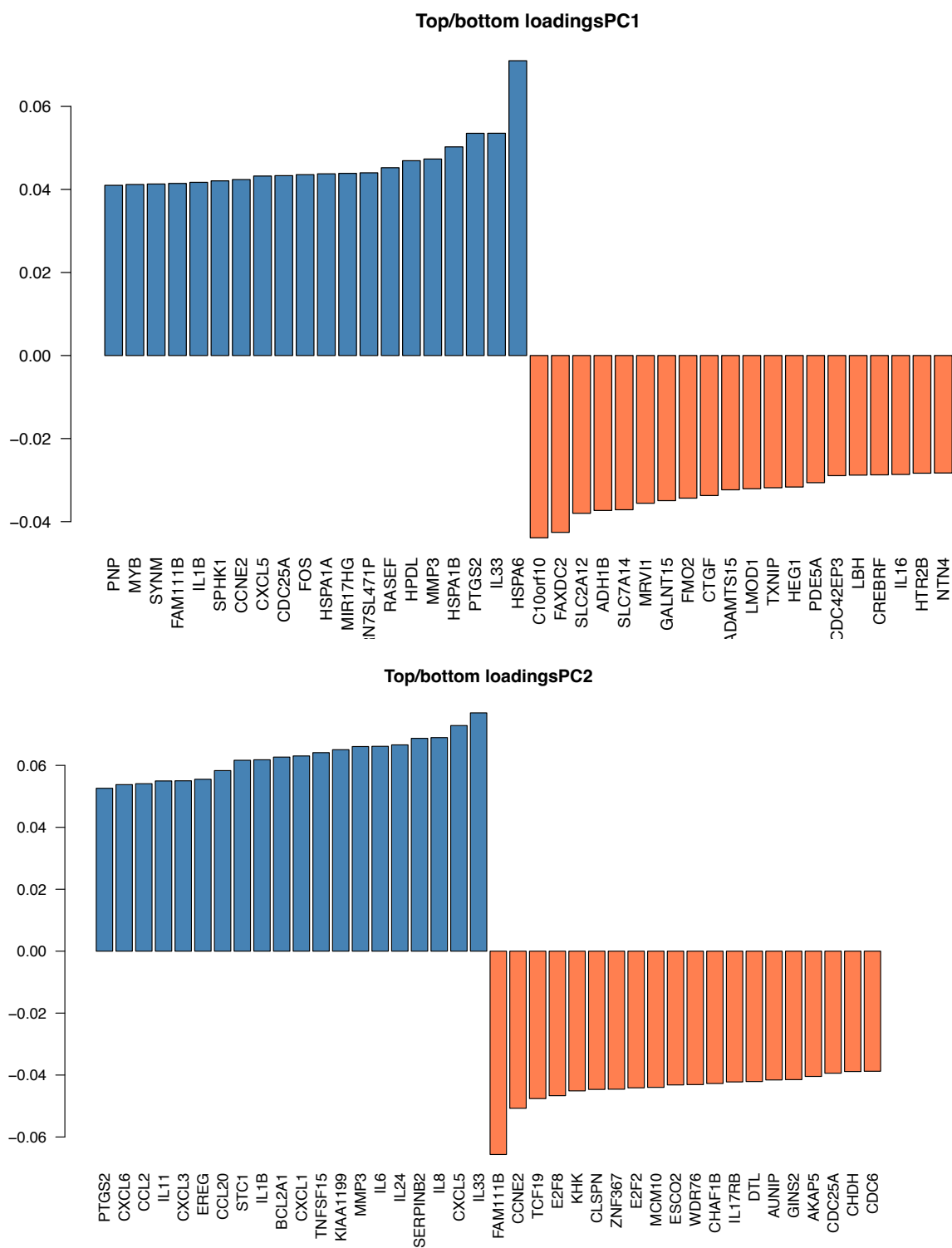


Figure S2: PCA Top/Bottom Loading Genes, PC1 and PC2. Top loadings of a PCA component are the genes that contribute most to the positive side of the component, bottom loadings are the genes that contribute most to the negative side of the component.

PC1 Negative Component (Mock, 12S)		PC1 Positive Component (WT, 13S, YCD)	
Term	log2pval	Term	log2pval
glomerulus vasculature development	18.61	ribosome biogenesis	57.79
SMAD protein signal transduction	17.17	DNA replication initiation	52.30
cell migration	15.41	regulation of DNA-templated DNA replication	51.61
positive regulation of protein phosphorylation	15.18	rRNA processing	50.22
cell-substrate adhesion	14.07	cell cycle DNA replication	44.78
sphingolipid biosynthetic process	13.80	regulation of DNA-templated DNA replication initiation	38.73
positive regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	13.66	nuclear DNA replication	38.66
cell motility	13.57	double-strand break repair via break-induced replication	37.40
muscle cell proliferation	13.32	DNA unwinding involved in DNA replication	32.15
chromosome condensation	12.91	mitotic cell cycle	31.37
regulation of heart rate by cardiac conduction	12.70	response to stress	29.52
negative regulation of epithelial cell proliferation	12.61	ribosomal large subunit biogenesis	21.37
regulation of action potential	12.52	double-strand break repair via homologous recombination	21.37
positive regulation of cell migration	12.44	negative regulation of DNA replication	18.25
bleb assembly	12.36	mitotic DNA replication	17.45
cardiac atrium development	12.15	positive regulation of cell population proliferation	17.21
calcium ion import across plasma membrane	11.91	tissue development	16.65
glomerular mesangium development	11.80	DNA replication checkpoint signaling	16.35
respiratory tube development	11.40	response to other organism	16.35
epithelial to mesenchymal transition involved in endocardial cushion formation	11.32	positive regulation of DNA replication	16.02
ventricular trabecula myocardium morphogenesis	11.32	cell-cell signaling	15.12
system process	11.25	response to glucocorticoid	14.93
male gonad development	10.85	G1/S transition of mitotic cell cycle	14.00
positive regulation of lipid biosynthetic process	10.48	positive regulation of miRNA transcription	13.84
lens development in camera-type eye	10.06	DNA strand elongation involved in DNA replication	13.84

Table S3: PCA PC1 Positive and Negative Loadings GO terms

PC2 Negative Component (13S, 12S, WT)		PC2 Positive Component (Mock/UI, YCD)	
Term	log2pval	Term	log2pval
double-strand break repair	99.39	inflammatory response	84.60
regulation of cell cycle	88.21	cell surface receptor signaling pathway	81.01
cell cycle DNA replication	87.80	response to cytokine	72.32
double-strand break repair via homologous recombination	77.84	response to lipopolysaccharide	66.50
nuclear DNA replication	76.27	angiogenesis	56.21
chromosome segregation	76.03	negative regulation of multicellular organismal process	55.21

negative regulation of cell cycle phase transition	73.14	cellular response to lipopolysaccharide	51.19
regulation of DNA-templated DNA replication	68.76	negative regulation of signal transduction	49.06
DNA replication initiation	65.67	positive regulation of cell migration	48.83
regulation of mitotic cell cycle	63.49	cytokine-mediated signaling pathway	46.51
mitotic cell cycle phase transition	62.27	positive regulation of cell differentiation	41.86
DNA-templated DNA replication maintenance of fidelity	60.00	positive regulation of inflammatory response	39.38
cell division	59.66	response to hypoxia	38.94
DNA duplex unwinding	59.31	positive regulation of cell communication	38.66
meiotic cell cycle	52.08	positive regulation of angiogenesis	38.06
meiosis I cell cycle process	50.89	positive chemotaxis	37.00
DNA strand elongation involved in DNA replication	50.43	positive regulation of MAPK cascade	35.69
mitotic sister chromatid segregation	49.69	response to organic cyclic compound	32.96
mitotic DNA replication	48.45	cell-cell signaling	32.63
positive regulation of cell cycle process	48.11	positive regulation of epithelial cell proliferation	31.59
DNA unwinding involved in DNA replication	46.79	negative regulation of apoptotic process	30.37
meiotic nuclear division	45.07	positive regulation of gene expression	29.52
negative regulation of mitotic cell cycle phase transition	43.53	skeletal system development	26.84
homologous recombination	43.47	cellular response to interleukin-1	26.58
replication fork processing	43.32	regulation of ERK1 and ERK2 cascade	25.99

Table S4: PCA PC2 Positive and Negative Loadings GO terms

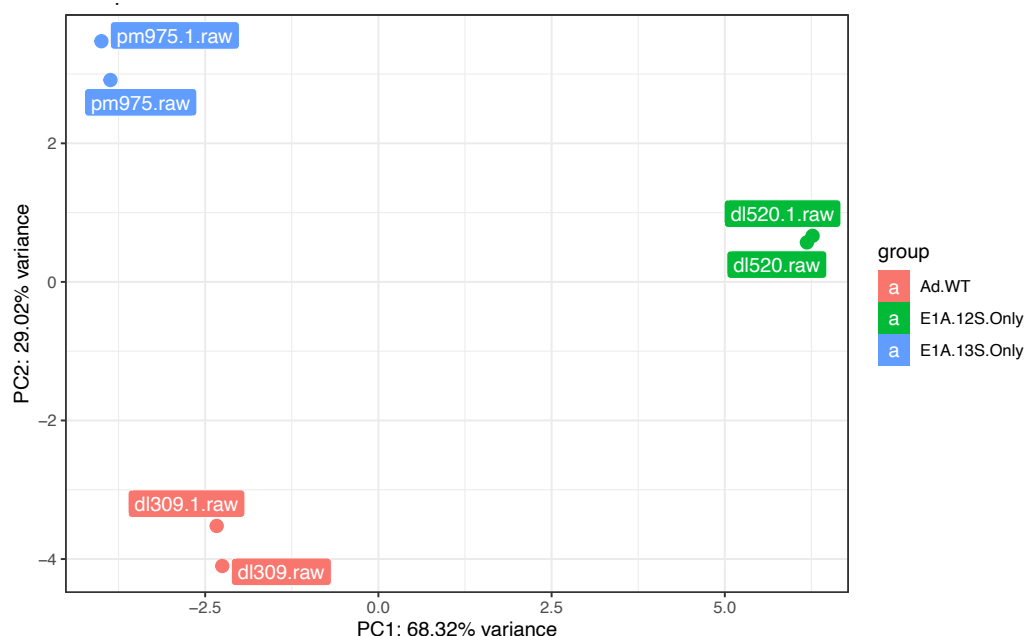


Figure S4: PCA 3 sample chart. RNA-Seq PCA. Principal Component Analysis plot generated using pcaExplorer with scaling via DESeq2 rlog transformation and displaying the first two principal coordinates. Individual replicates are colored according to group (WT,

13S, 12S. The first component (PC1) clusters WT and 13S together, separate from 12S and explains 68.32% of the variance. In second component (PC2) which explains 29.02% of the variance, WT and 12S are closer to each other than 13S.

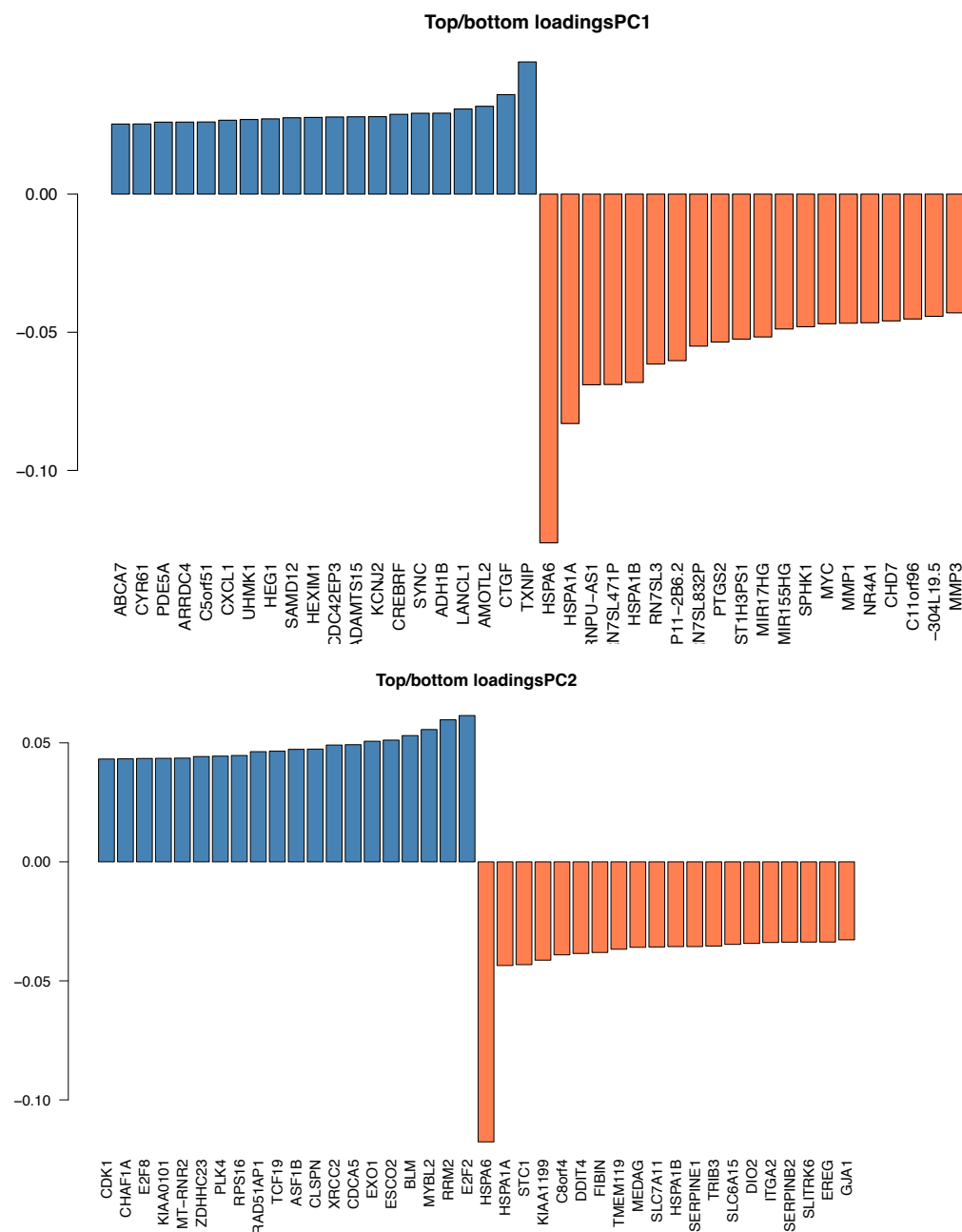


Figure S5: PCA Top/Bottom Loading Genes, PC1 and PC2. 3 samples (WT, 13S, 12S). Top loadings of a PCA component are the genes that contribute most to the positive side of the component, bottom loadings are the genes that contribute most to the negative side of the component.

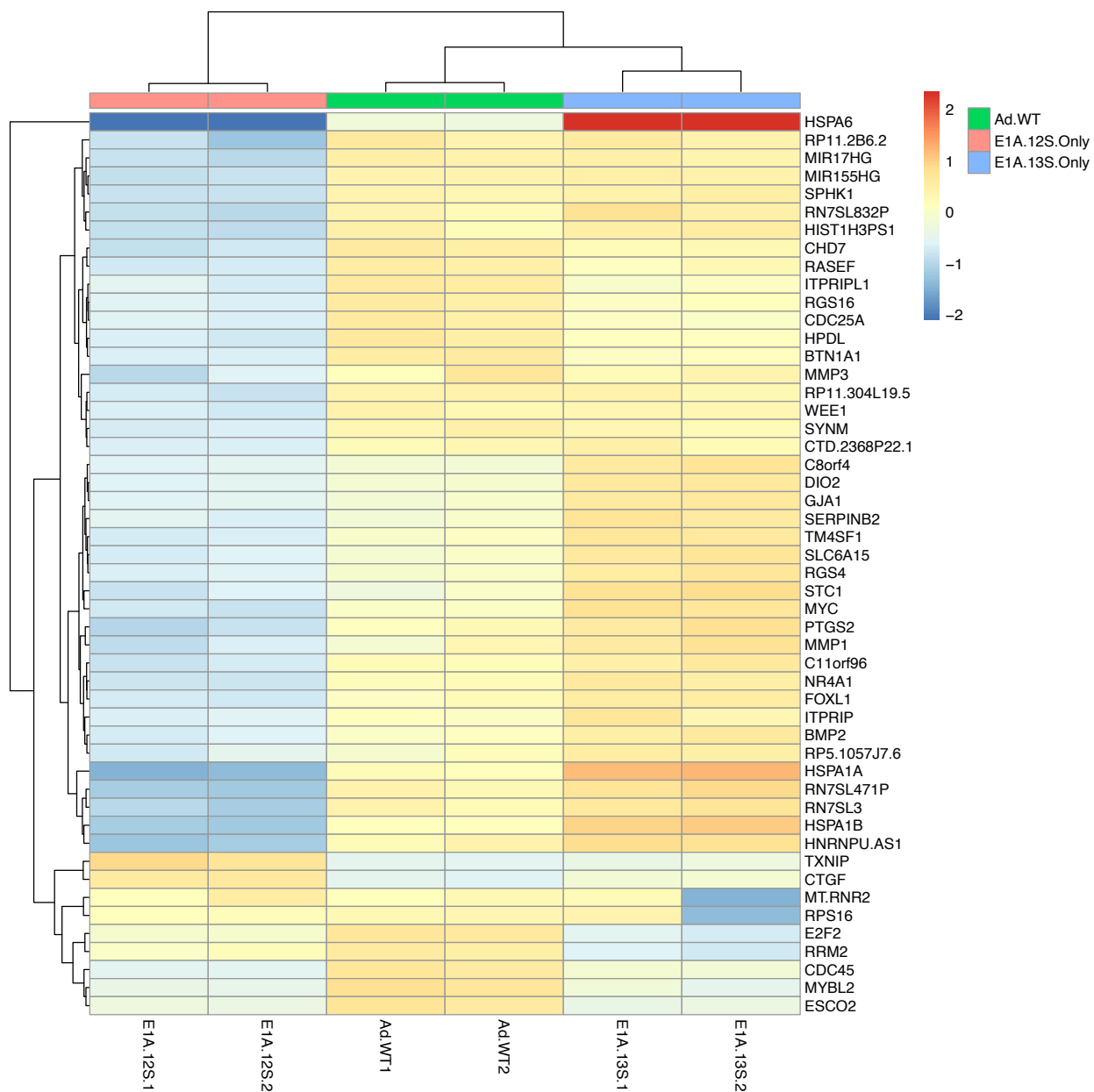
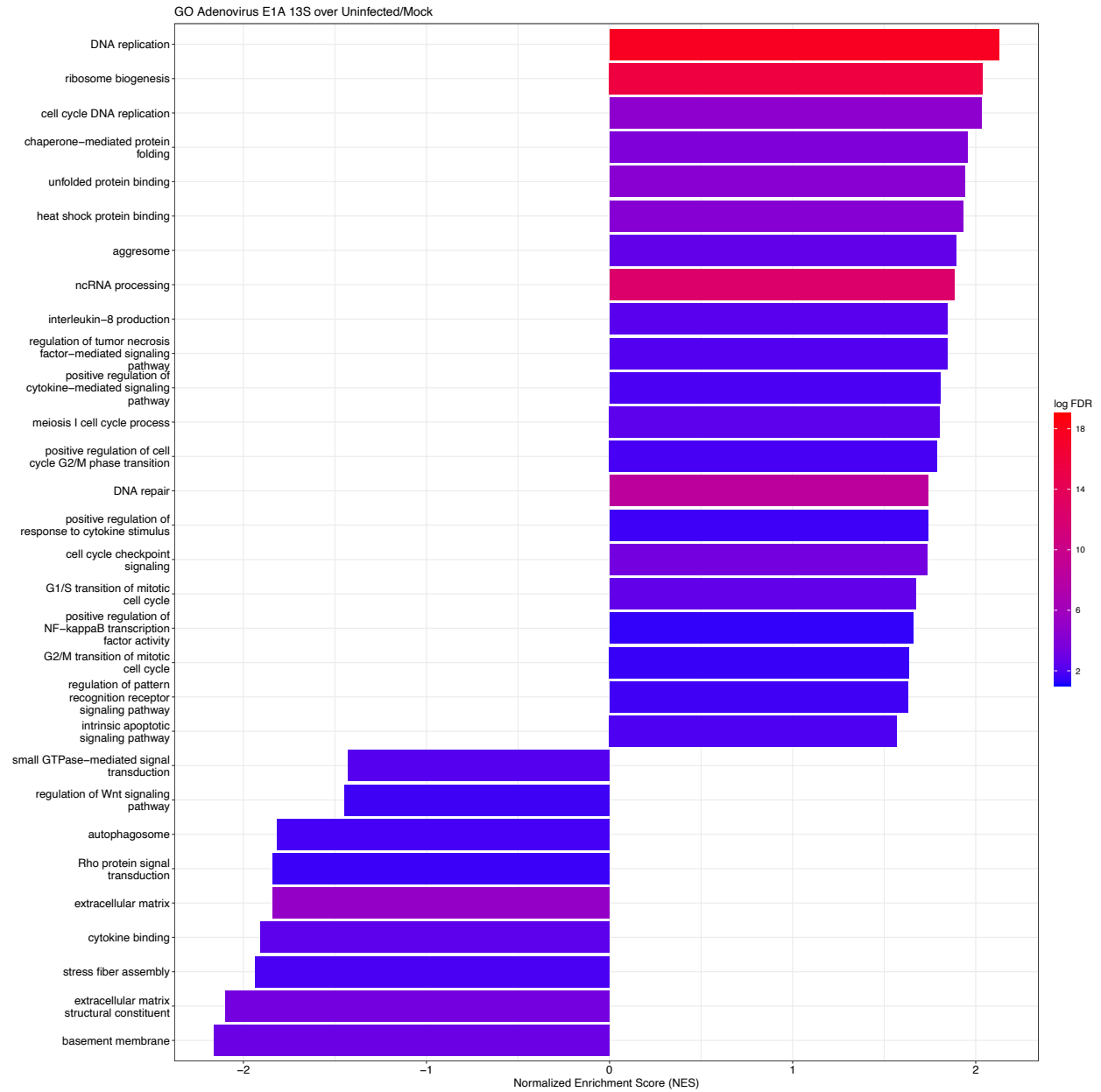
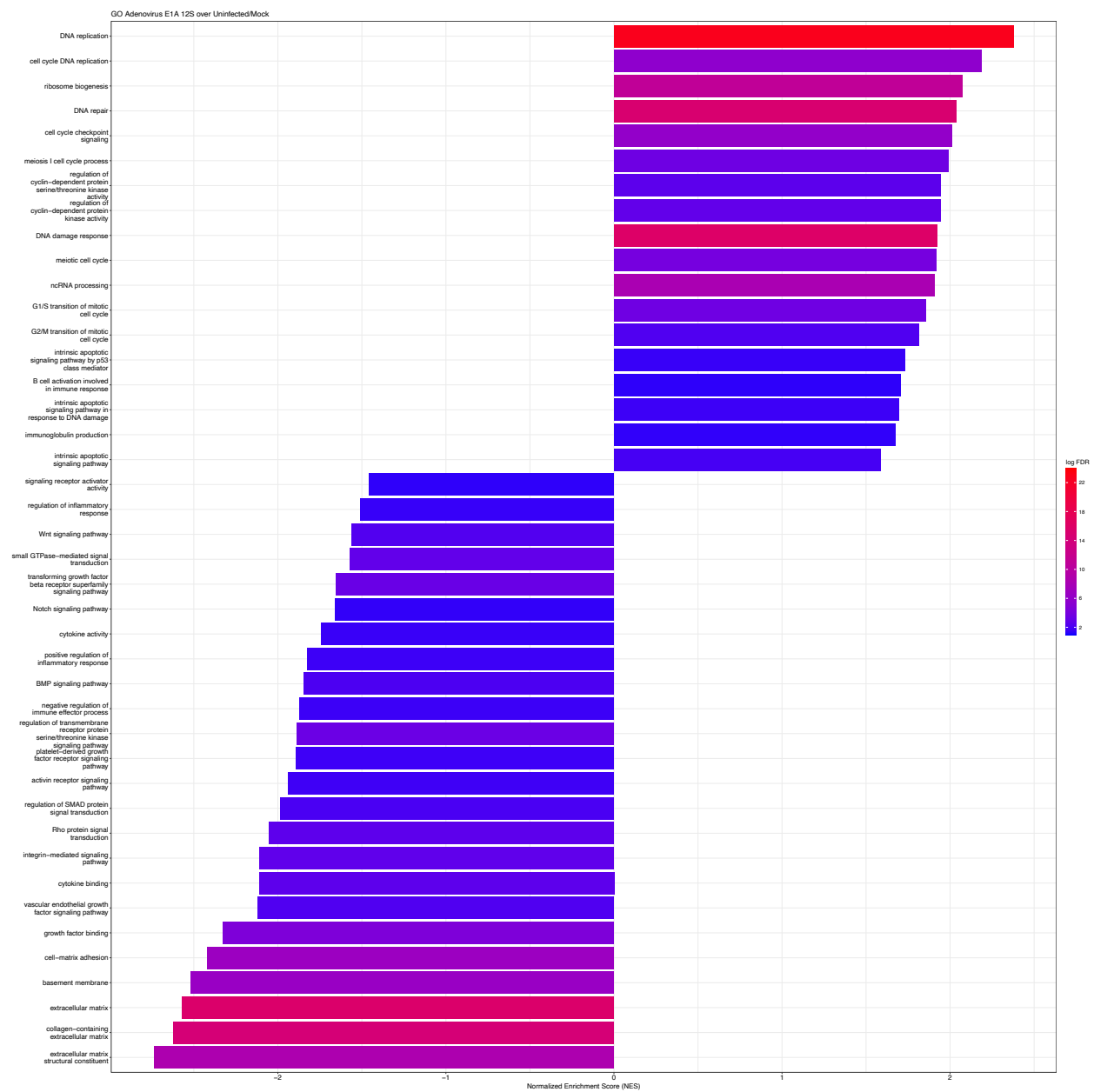


Figure S6: Heat Map 3 Samples (WT, 13S, 12S), 50 most variable genes. Heat map generated using pheatmap with hierarchical clustering (samples and transcripts) after using DESeq2 rlog transformation for scaling. Boxes in red are upregulated, blue are downregulated.





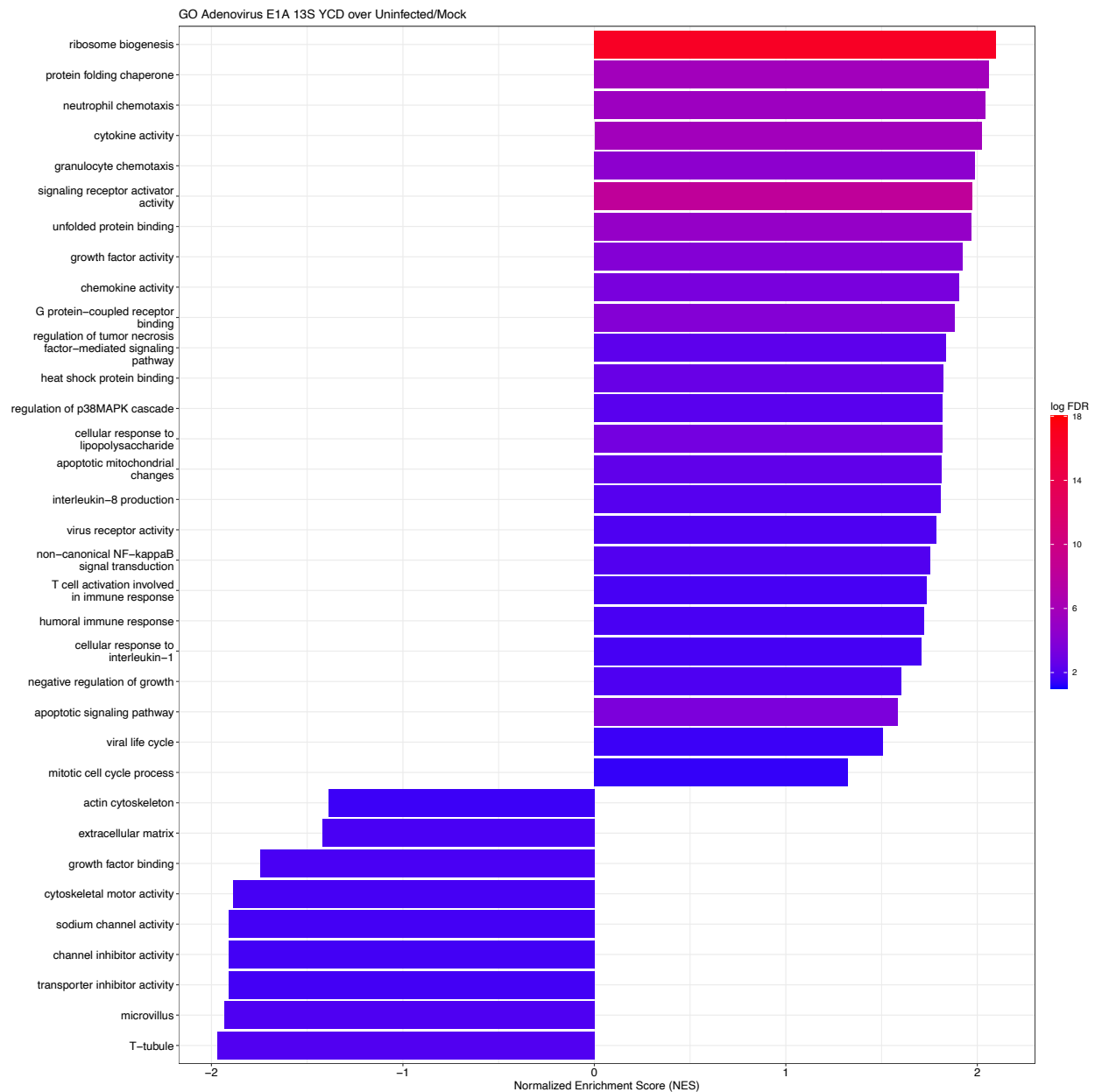


Figure S7: RNA-Seq GO Bar plots (13S/UI, 12S/UI, YCD/UI). Gene set enrichment analysis (GSEA) of differential expression comparisons. The x axis is the normalized expression score (NES), indicating the extent to which the GO terms (y axis) are upregulated (NES > 0) or downregulated (NES < 0). Bars are colored based on logFDR (false discovery rate) (scale bar on side).

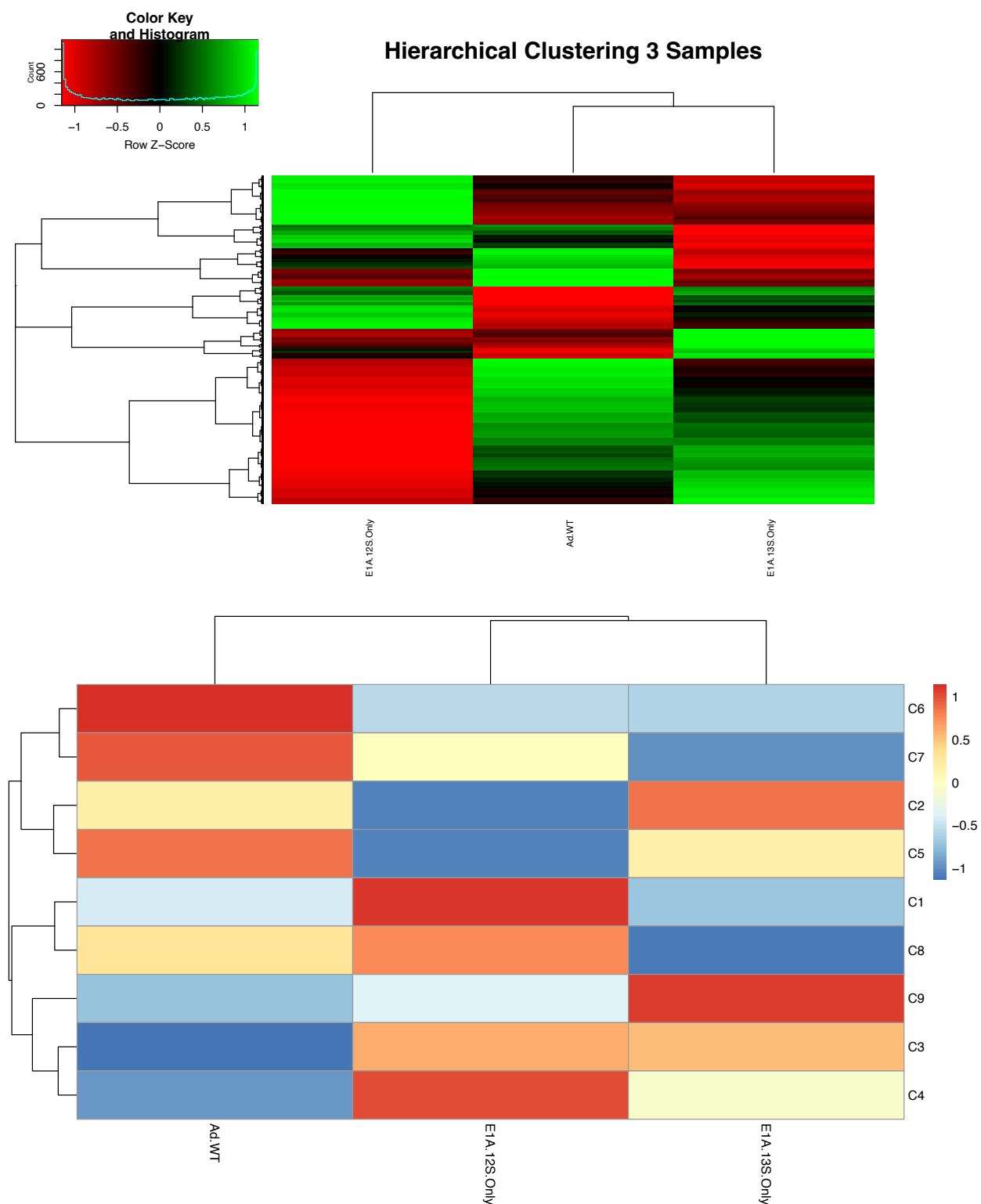


Figure S8: Heat Map Clusters A. Three samples (WT, 12S, and 13S) were clustered according to similarity producing 9 clusters. The bottom chart was produced from the first by averaging the gene expression within each cluster per sample.

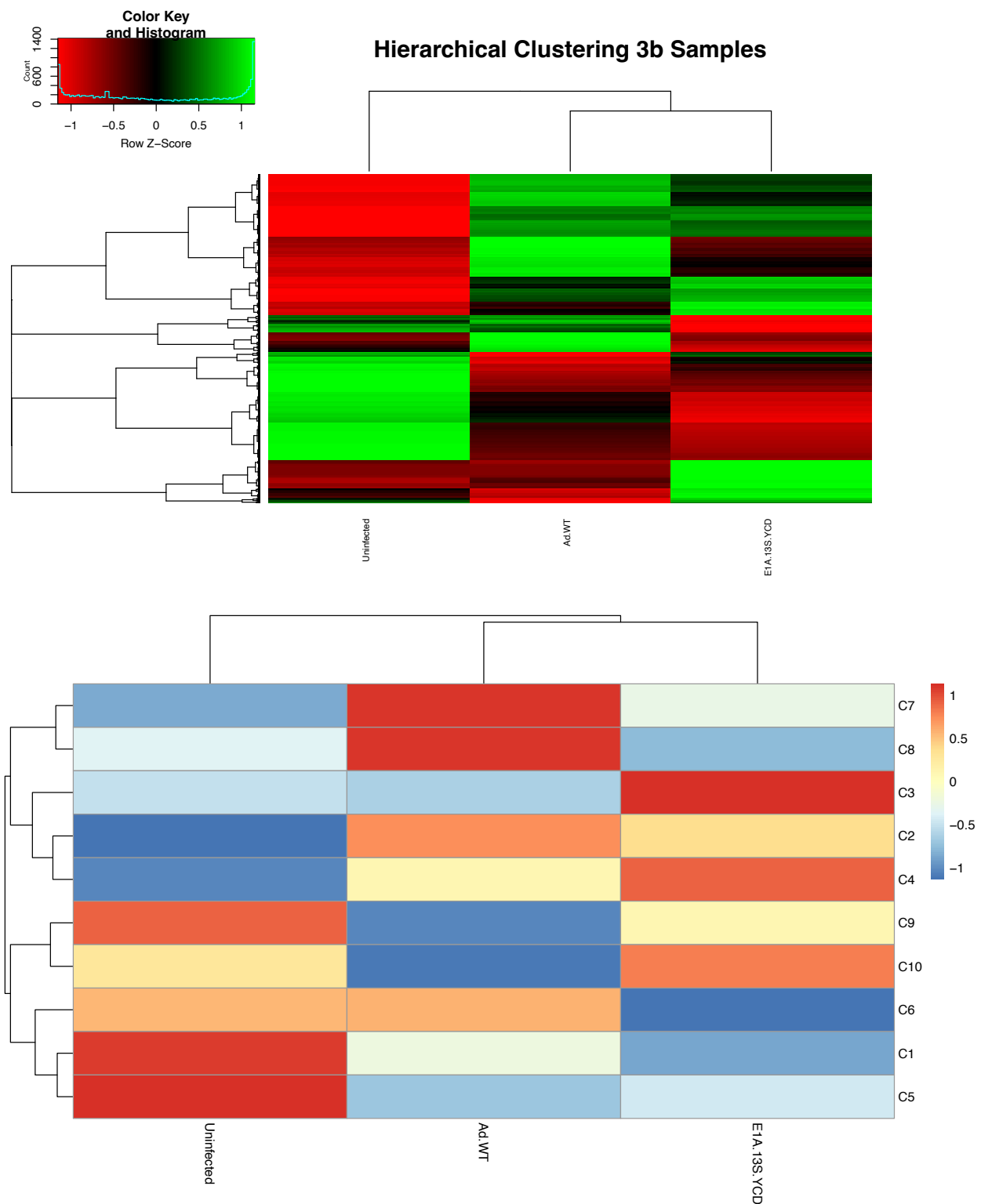


Figure S9: Heat Map Clusters B. Three samples (WT,UI, and YCD) were clustered according to similarity producing 9 clusters. The bottom chart was produced from the first by averaging the gene expression within each cluster per sample.

Cluster	#Genes	Expression			GO Categories	miRNA	TFs
		WT	13S	12S			
C1	781				actin, cytoskeleton binding, protein phosphorylation	miR-335-5p	
C2	953				folding chaperone, cell junction, IL-4 and IL-13 signaling, Cori cycle (lactic acid cycle), anion transmembrane transport		MAZ, Zbtb37, ZNF183, E2F, WT1,
C3	310				cell adhesion, migration, stress response, inflammatory response		IRF2, IRF2, STAT2, IRF3, POU2F1
C4	381				cell proliferation, MAPK cascade	miR-355-5p, miR-320e, miR-21-5p	
C5	1401				ribosome biogenesis, DNA replication	let7b5p, miR985p, miR2155p, miR193b3p, miR6153p, miR165p	E2F, cMyc, DP1, TCF1, CHD2
C6	283				cell cycle, MutSbeta and MutLbeta complexes, DNA repair, MCM8-MCM9 complex, retinoblastoma gene	miR193b3p, miR1925p, miR2155p	E2Fs, TBR2, IRX1, DPs Rb, TCF1, WT1, MYC
C7	332				Microtubule binding, DNA translocase, cell cycle	miR1925p, miR2155p, miR193b3p	E2Fs, FOXI1, RNF96, NFYA, HMGXB3
C8	383				cytoskeletal protein binding, cell cycle, organelle fission, centromere complex assembly	miR2155p	E2Fs SP1, HTF4, MRF4,
C9	478				signaling, inflammatory response	miR3355p, miR1523p	C/EBPalpha, C/EBP, foxm1, TCF3, CIZ

Table S6: Cluster Table for Wild Type Adenovirus and Adenoviruses with E1A 13S and E1A 12S Isoforms, using hierarchical clustering. Red indicates underexpression, green

overexpression, and white average expression. Clusters were analyzed in g:profiler for overexpressed GO terms, miRNAs, and transcription factors (TFs).

Cluster	#Genes	Expression			GO Categories	miRNA	TFs
		UI	WT	YCD			
C1	1100				Ion binding, gated channel activity	hsa-miR-335-5p	
C2	1100				Ribosome Biogenesis, RNA metabolism	miR-16-5p, let-7b-5p, miR-98-5p	E2F 1-3, ZF5, HES-7, Hey2, cMyc
C3	610				signaling, cytokines in immunity	miR-335-5p, miR-98-5p	AP-1 TFs, Oct-3, p65-RelA
C4	624				Unfolded protein binding, Chaperone, ubiquitin, apoptosis	miR-6743-5p, hsa-miR-4688,	ZF5, Hey1, E2F2, Egr-1 TCF-1
C5	448				cell motility, migration, cytoskeleton	miR-335-5-, miR-21-5-p, miR-4762-3p	ZNF148, MSGN1, MAZ, TIEG1. PATZ
C6	276				Cell cycle	miR-215-5p, miR-192-5p	HMGXB3, Slug, NFYA
C7	640				DNA damage response, retinoblastoma gene	miR-215-5p, miR-192-5p, miR-193b-3p	E2Fs, ZF5, DP-2, Egr-1, TCF-1
C8	318				DNA helicase activity, ATP binding, interstrand cross-link repair	miR-192-5p, miR-193b-3p, miR-215-5p	E2Fs, ZF5, DPs, p300, Rb
C9	197				metabolic regulation		
C10	89				Growth factor response	mir-7-1	FOXO3A

Table S7: Cluster Table for Uninfected, Wild Type Adenovirus, and Adenovirus with E1A 13S YCD Mutation, using hierarchical clustering. Red indicates underexpression, green overexpression, and white average expression. Clusters were analyzed in g:profiler for overexpressed GO terms, miRNAs, and transcription factors (TFs).

Transcript	log2FC	pvalue	padj
<i>IL1B</i>	5.600	3.45E-190	4.48E-187

<i>EREG</i>	4.445	4.06E-112	1.72E-109
<i>IL1A</i>	3.304	9.29E-26	3.49E-24
<i>CYP1A1</i>	3.144	5.66E-10	5.74E-09
<i>EDN1</i>	2.110	2.79E-07	1.94E-06
<i>IGF2</i>	1.976	1.78E-10	1.91E-09
<i>SPHK1</i>	1.703	1.05E-59	1.41E-57
<i>TBX3</i>	1.346	4.79E-75	1.05E-72
<i>WNT5A</i>	1.323	4.34E-52	4.40E-50
<i>FOSL1</i>	1.231	4.28E-13	5.99E-12
<i>HES1</i>	1.219	1.37E-15	2.44E-14
<i>CITED2</i>	1.200	1.28E-48	1.14E-46
<i>CCND1</i>	0.986	2.72E-38	1.79E-36
<i>TBX2</i>	0.782	4.84E-25	1.72E-23
<i>TGFB1</i>	0.743	2.16E-14	3.46E-13
<i>CCDC57</i>	0.731	9.65E-09	8.27E-08
<i>SHB</i>	0.719	1.12E-09	1.10E-08
<i>EGFR</i>	0.702	2.40E-14	3.81E-13
<i>DDX3X</i>	0.635	2.50E-15	4.39E-14
<i>PLCB1</i>	0.618	3.18E-13	4.50E-12
<i>DYRK3</i>	0.613	2.82E-09	2.61E-08
<i>MSX2</i>	0.611	1.96E-06	1.20E-05
<i>RGCC</i>	0.606	2.01E-05	1.03E-04
<i>ANXA1</i>	0.603	5.27E-09	4.69E-08
<i>CCNF</i>	-0.586	9.47E-06	5.14E-05
<i>FHL1</i>	-0.604	4.99E-16	9.36E-15
<i>GMNN</i>	-0.613	4.98E-10	5.07E-09
<i>DONSON</i>	-0.613	8.93E-10	8.83E-09
<i>CCP110</i>	-0.614	7.45E-13	1.02E-11
<i>FOXN3</i>	-0.617	2.63E-12	3.40E-11
<i>RBBP8</i>	-0.622	3.74E-11	4.33E-10
<i>HEXIM2</i>	-0.624	1.90E-03	6.51E-03
<i>MAD2L2</i>	-0.650	3.07E-08	2.46E-07
<i>INTS7</i>	-0.671	1.57E-14	2.55E-13
<i>KLHL22</i>	-0.673	1.53E-11	1.84E-10
<i>RNASEH2B</i>	-0.682	1.29E-09	1.25E-08
<i>NPR2</i>	-0.685	8.56E-07	5.55E-06
<i>BTG2</i>	-0.713	3.06E-12	3.93E-11
<i>RAD9A</i>	-0.719	3.20E-06	1.90E-05

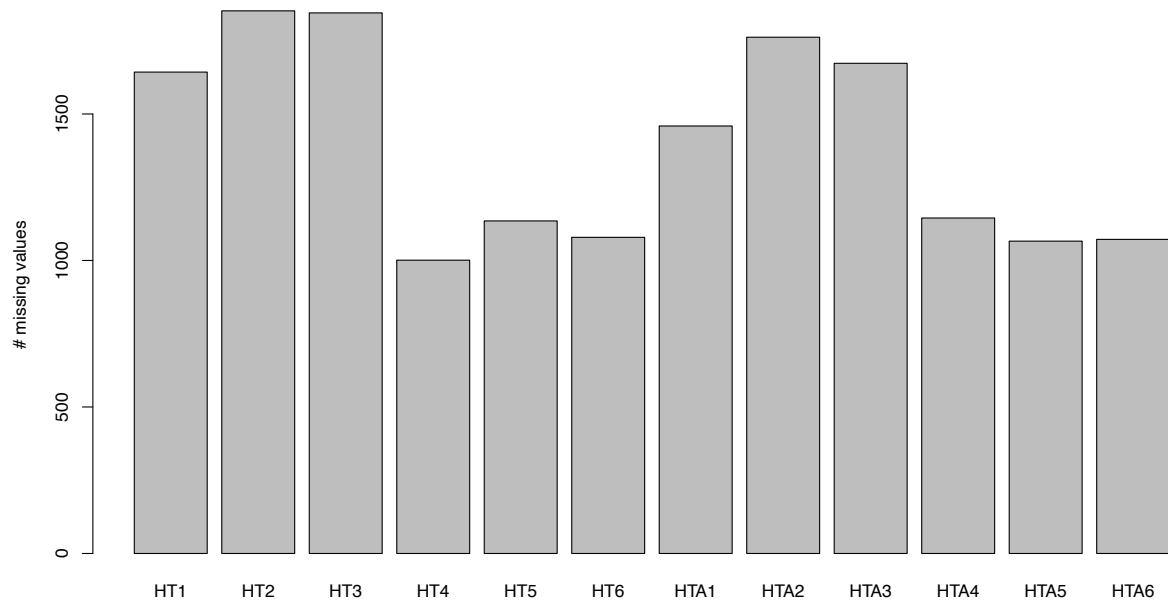
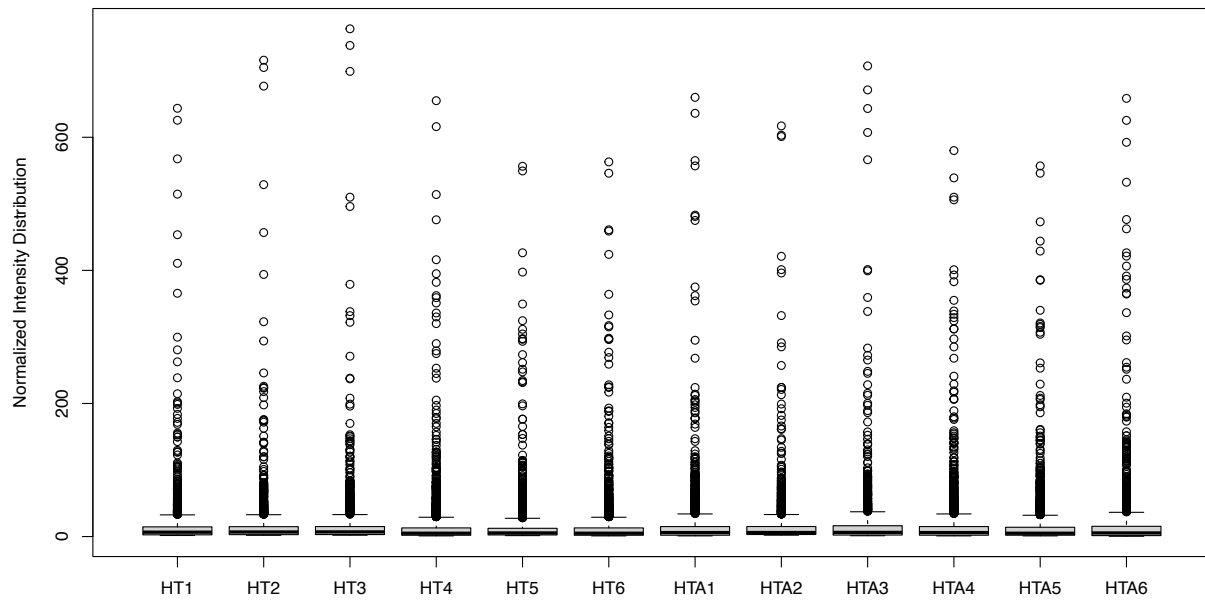
<i>TOM1L2</i>	-0.754	9.42E-12	1.16E-10
<i>FBXO43</i>	-0.758	1.81E-02	4.67E-02
<i>TIMELESS</i>	-0.778	6.87E-10	6.92E-09
<i>RBL1</i>	-0.799	2.29E-14	3.65E-13
<i>GEN1</i>	-0.799	7.98E-13	1.09E-11
<i>CDT1</i>	-0.829	1.39E-13	2.06E-12
<i>AURKB</i>	-0.831	4.87E-04	1.91E-03
<i>CHEK1</i>	-0.854	8.58E-15	1.45E-13
<i>TRIP13</i>	-0.858	1.63E-07	1.17E-06
<i>BRCA2</i>	-0.867	1.48E-14	2.41E-13
<i>SLFN11</i>	-0.890	4.98E-28	2.10E-26
<i>MRE11A</i>	-0.894	3.14E-14	4.92E-13
<i>MDM1</i>	-0.915	9.21E-12	1.13E-10
<i>NUF2</i>	-0.926	1.62E-04	7.01E-04
<i>ORC1</i>	-0.995	2.52E-14	4.00E-13
<i>CCNB1</i>	-1.000	1.47E-13	2.17E-12
<i>SUV39H1</i>	-1.025	7.10E-10	7.13E-09
<i>BIRC5</i>	-1.058	2.00E-09	1.89E-08
<i>RHNO1</i>	-1.060	1.62E-18	3.78E-17
<i>MTBP</i>	-1.064	2.51E-10	2.65E-09
<i>BUB1</i>	-1.072	8.22E-08	6.21E-07
<i>MSH2</i>	-1.115	8.24E-36	5.03E-34
<i>ATRIP</i>	-1.176	2.08E-18	4.79E-17
<i>CENPF</i>	-1.178	2.74E-18	6.18E-17
<i>CDC45</i>	-1.179	4.82E-12	6.07E-11
<i>CDC6</i>	-1.200	3.56E-06	2.09E-05
<i>DTX3L</i>	-1.205	4.49E-25	1.60E-23
<i>FBXO6</i>	-1.218	1.33E-12	1.78E-11
<i>TIPIN</i>	-1.271	1.01E-27	4.18E-26
<i>BARD1</i>	-1.273	2.16E-27	8.82E-26
<i>FBXO5</i>	-1.278	1.23E-22	3.77E-21
<i>MYOCD</i>	-1.293	9.33E-09	8.02E-08
<i>FZD3</i>	-1.296	9.86E-08	7.35E-07
<i>TICRR</i>	-1.298	2.43E-13	3.49E-12
<i>KAT2B</i>	-1.310	3.02E-24	1.03E-22
<i>ERCC6</i>	-1.326	1.04E-16	2.09E-15
<i>FANCD2</i>	-1.338	1.04E-14	1.73E-13
<i>RAD51</i>	-1.378	1.30E-10	1.42E-09
<i>NUPR1</i>	-1.384	1.55E-48	1.37E-46

<i>FBXO4</i>	-1.385	1.92E-21	5.42E-20
<i>CDC20</i>	-1.412	1.04E-10	1.15E-09
<i>ZWINT</i>	-1.444	7.81E-21	2.11E-19
<i>DNA2</i>	-1.458	6.66E-27	2.67E-25
<i>SPC24</i>	-1.466	8.40E-10	8.35E-09
<i>PKMYT1</i>	-1.472	2.01E-13	2.91E-12
<i>BCL2</i>	-1.575	1.71E-27	6.98E-26
<i>DTL</i>	-1.576	5.33E-47	4.42E-45
<i>TTK</i>	-1.590	2.38E-16	4.60E-15
<i>PARP9</i>	-1.622	2.31E-36	1.43E-34
<i>E2F1</i>	-1.741	9.25E-38	5.99E-36
<i>PPP1R13B</i>	-1.765	6.78E-09	5.94E-08
<i>CDK1</i>	-1.767	4.73E-24	1.59E-22
<i>GTSE1</i>	-1.789	5.39E-09	4.79E-08
<i>STK33</i>	-1.816	8.14E-13	1.11E-11
<i>CDCA8</i>	-1.872	3.93E-16	7.44E-15
<i>NEK2</i>	-1.906	2.01E-07	1.43E-06
<i>BUB1B</i>	-1.923	1.78E-25	6.57E-24
<i>BRCA1</i>	-1.925	7.58E-47	6.25E-45
<i>WDR76</i>	-1.931	2.32E-76	5.37E-74
<i>EME1</i>	-2.167	9.49E-12	1.17E-10
<i>BLM</i>	-2.168	3.25E-36	1.99E-34
<i>NDC80</i>	-2.169	2.27E-17	4.76E-16
<i>BRIP1</i>	-2.404	5.07E-48	4.36E-46
<i>SPC25</i>	-2.443	2.21E-07	1.56E-06
<i>CLSPN</i>	-2.552	8.57E-59	1.10E-56
<i>E2F8</i>	-3.103	2.00E-42	1.47E-40

Table S8: Full table of positively and negatively regulated cell cycle process genes

	Name	Expressed	Sequenced	Stable Cell Lines
1	dl 0-10			
2	dl 0-20			
3	dl 21-30			
4	dl 31-40			
5	dl 41-50			
6	dl 51-60			
7	dl 61-70			
8	dl 71-80			
9	dl 81-90			
10	dl 91-100			
11	dl 101-110			
12	dl 111-120			
13	dl 121-130			
14	dl 131-140			
15	dl 141-150			
16	dl 151-160			
17	dl 161-170			
18	dl 171-180			
19	dl 181-190			
20	dl 190-200			
21	dl 201-210			
22	dl 211-220			
23	dl 221-230			
24	dl 231-240			
25	dl 241-250			
26	dl 251-273			
27	dl 261-273			
28	dl 1-40			
29	dl 41-80			
30	dl 81-120			
31	dl 121-160			
32	dl 161-200			
33	dl 191-240			
34	dl 241-273			
	Total	28	30	21

Table S9: Arglu1 Deletion Mutants and Cell Lines. Green or yellow indicates the plasmids or cell lines are made and verified, yellow indicates plasmids were made by Peter Pelka, and white were not made.



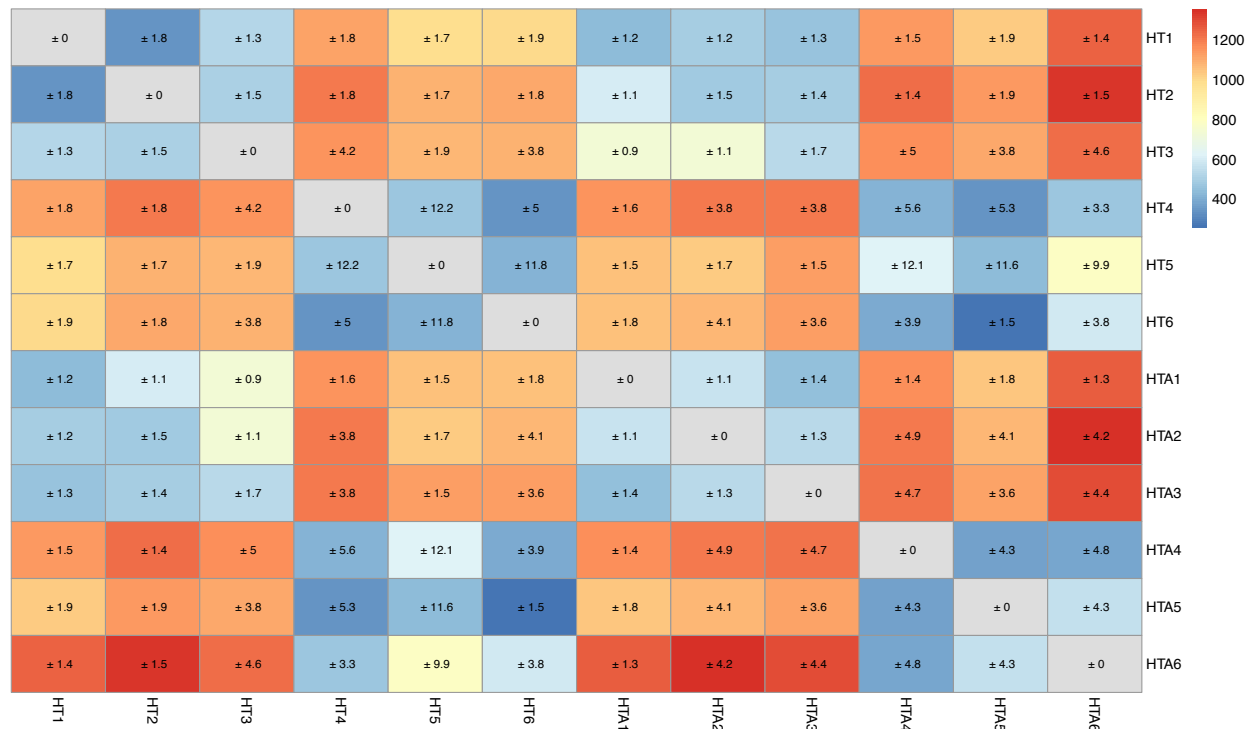
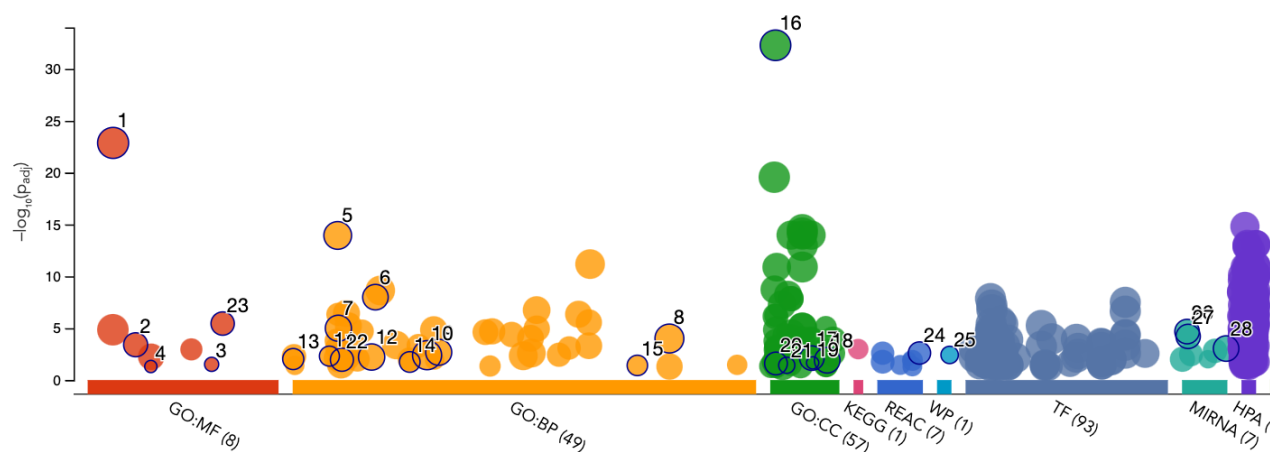


Figure S10: Proteomics Normalized Intensity Distribution (Y axis is normalized average protein abundance or intensity), Missing Values per sample, and Sample to Sample Heat Map showing the differences between each replicate. Grey indicates the samples are identical (i.e. the diagonal), blue indicates similarity and red indicates difference (the samples that are the most different are HTA2-HTA6 and HT2-HTA6).



ID	Source	Term ID	Term Name	Padj (query_... ↑)
16	GO:CC	GO:0005737	cytoplasm	5.277×10 ⁻³³
1	GO:MF	GO:0005515	protein binding	1.368×10 ⁻²³
5	GO:BP	GO:0006996	organelle organization	1.144×10 ⁻¹⁴
6	GO:BP	GO:0015031	protein transport	1.025×10 ⁻⁸
23	GO:MF	GO:0050839	cell adhesion molecule binding	3.592×10 ⁻⁶
7	GO:BP	GO:0007049	cell cycle	1.013×10 ⁻⁵
26	MIRNA	MIRNA:hsa-mi...	hsa-miR-155-5p	2.125×10 ⁻⁵
27	MIRNA	MIRNA:hsa-mi...	hsa-miR-193b-3p	6.215×10 ⁻⁵
8	GO:BP	GO:1901564	organonitrogen compound metabolic process	9.934×10 ⁻⁵
2	GO:MF	GO:0016491	oxidoreductase activity	3.982×10 ⁻⁴
28	MIRNA	MIRNA:hsa-mi...	hsa-miR-92a-3p	9.193×10 ⁻⁴
9	GO:BP	GO:0033554	cellular response to stress	2.105×10 ⁻³
24	REAC	REAC:R-HSA-3...	Transcriptional Regulation by TP53	2.478×10 ⁻³
25	WP	WP:WP5300	TROP2 regulatory signaling	3.806×10 ⁻³
10	GO:BP	GO:0032502	developmental process	4.019×10 ⁻³
11	GO:BP	GO:0006354	DNA-templated transcription elongation	4.953×10 ⁻³
12	GO:BP	GO:0012501	programmed cell death	6.032×10 ⁻³
17	GO:CC	GO:0070161	anchoring junction	7.735×10 ⁻³
13	GO:BP	GO:0000082	G1/S transition of mitotic cell cycle	9.484×10 ⁻³
22	GO:BP	GO:0007346	regulation of mitotic cell cycle	1.105×10 ⁻²
18	GO:CC	GO:0099080	supramolecular complex	1.243×10 ⁻²
14	GO:BP	GO:0022900	electron transport chain	1.814×10 ⁻²
19	GO:CC	GO:0070449	elongin complex	2.206×10 ⁻²
20	GO:CC	GO:0005769	early endosome	2.533×10 ⁻²
3	GO:MF	GO:0048407	platelet-derived growth factor binding	3.069×10 ⁻²
21	GO:CC	GO:0030686	90S preribosome	3.769×10 ⁻²
15	GO:BP	GO:0140053	mitochondrial gene expression	3.902×10 ⁻²
4	GO:MF	GO:0019826	oxygen sensor activity	4.961×10 ⁻²

Figure S11: Arglu1 Proteomics g:Profiler Figure showing the gene ontology terms from Arglu1 over expression analysis (OEA).

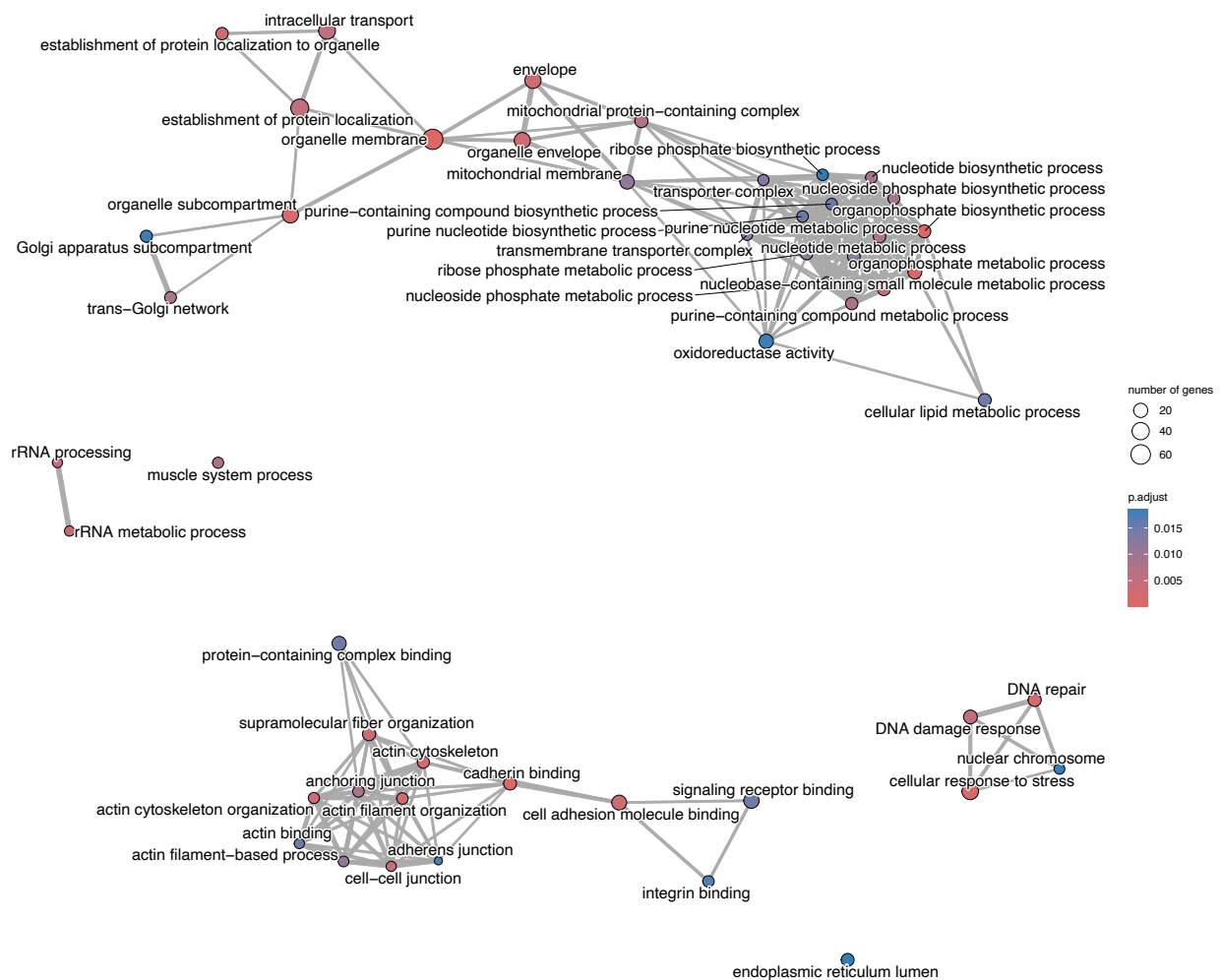


Figure S12: Arglu1 Proteomics GSEA Enrichment Map. Map showing connections between the 50 GO terms with the lowest adjusted pvalue (indicated by color, blue is lower or more significant, red is higher). The size of each dot indicates the amount of genes in each category.

Protein	pval	LogFC	t_stat	Std. Err.	avg intensity	n_approx	n_obs
<i>PCNA</i>	0.02713	24.62	2.59	9.52	55.90	12.00	12
<i>CALR</i>	0.09556	23.76	1.84	12.91	103.33	12.00	12
<i>HSP90B1</i>	0.07069	22.22	2.02	10.99	81.69	12.00	12
<i>NASP</i>	0.00780	20.81	3.32	6.28	42.92	12.00	12
<i>FLNC</i>	0.05063	19.83	2.22	8.93	47.56	12.00	12
<i>KHSRP</i>	0.02188	18.96	2.71	6.99	65.39	12.00	12
<i>HSPE1</i>	0.14703	14.25	1.57	9.06	25.16	12.00	12
<i>TPM4</i>	0.01304	13.28	3.01	4.41	79.76	12.00	12
<i>FUBP1</i>	0.25441	13.08	1.21	10.82	29.22	12.00	12
<i>P4HB</i>	0.21477	12.90	1.32	9.74	84.11	12.00	12
<i>H2BC11</i>	0.14073	12.70	1.60	7.94	9.57	9.89	10
<i>PSMD2</i>	0.05160	12.52	2.21	5.67	46.56	12.00	12
<i>CDV3</i>	0.17996	12.50	1.44	8.67	32.95	12.00	12
<i>HSP90AB4P</i>	0.17846	12.47	1.45	8.62	-1.67	4.98	4
<i>HDGF</i>	0.04003	12.39	2.36	5.25	47.59	12.00	12
<i>PDLIM1</i>	0.00265	12.02	3.97	3.03	22.18	12.00	12
<i>PPA1</i>	0.18011	12.01	1.44	8.33	40.59	12.00	12
<i>PRDX1</i>	0.25558	11.76	1.21	9.75	55.66	12.00	12
<i>EEA1</i>	0.36545	11.76	0.95	12.40	24.05	12.00	12
<i>EFTUD2</i>	0.13010	11.69	1.65	7.09	43.88	12.00	12
<i>PCBP1</i>	0.05008	11.28	2.23	5.06	34.72	12.00	12
<i>RCN3</i>	0.04921	10.98	2.24	4.91	14.79	12.00	12
<i>DDX3Y</i>	0.34798	10.88	0.98	11.05	0.39	4.41	4
<i>SSB</i>	0.06732	10.51	2.05	5.12	30.95	12.00	12
<i>PPP1CA</i>	0.07790	10.37	1.96	5.28	18.18	9.07	10

Protein	pval	LogFC	t_stat	Std. Err.	avg intensity	n_approx	n_obs
<i>ACTG1</i>	0.34199	-38.29	-1.00	38.38	580.35	12.00	12
<i>ANPEP</i>	0.01638	-23.73	-2.88	8.24	27.40	12.00	12
<i>ACTN1</i>	0.40018	-12.63	-0.88	14.37	131.76	12.00	12
<i>PGAM1</i>	0.24688	-10.19	-1.23	8.29	95.76	12.00	12
<i>PYGB</i>	0.05114	-8.85	-2.21	4.00	15.97	12.00	12
<i>PGK2</i>	0.21868	-8.54	-1.31	6.51	-6.64	4.00	1
<i>GAPDH</i>	0.80888	-8.44	-0.25	33.99	605.32	12.00	12
<i>MAP1B</i>	0.13532	-8.36	-1.62	5.15	33.17	12.00	12

<i>WDR1</i>	0.23707	-8.07	-1.26	6.42	39.38	12.00	12
<i>SLC25A6</i>	0.03131	-7.71	-2.50	3.08	-3.08	4.19	2
<i>RRBP1</i>	0.32498	-7.54	-1.04	7.28	20.89	12.00	12
<i>GSN</i>	0.06595	-7.30	-2.06	3.54	8.67	12.00	12
<i>DPYSL2</i>	0.05227	-7.27	-2.20	3.30	37.42	12.00	12
<i>PSAP</i>	0.04250	-7.18	-2.32	3.09	24.42	12.00	12
<i>GSTO1</i>	0.06341	-6.91	-2.09	3.31	33.79	12.00	12
<i>PLEC</i>	0.66432	-6.89	-0.45	15.41	12.38	12.00	12
<i>SERPINE2</i>	0.00644	-6.62	-3.43	1.93	-1.61	3.94	3
<i>DSG2</i>	0.00454	-6.36	-3.64	1.75	-1.37	3.76	3
<i>NEFL</i>	0.25742	-6.28	-1.20	5.23	-0.06	5.01	4
<i>NNMT</i>	0.00180	-6.13	-4.21	1.46	12.97	12.00	12
<i>ALDH1A3</i>	0.00743	-6.05	-3.34	1.81	-1.65	4.01	3
<i>TF</i>	0.10733	-6.03	-1.77	3.41	-4.35	3.95	1
<i>THBS1</i>	0.10733	-6.03	-1.77	3.41	-4.35	3.95	1
<i>SERPINB5</i>	0.01194	-6.02	-3.06	1.96	4.73	9.99	10
<i>PRDX6</i>	0.16480	-6.00	-1.50	4.00	29.01	12.00	12

Table S10: Arglu1 Proteomics top and bottom proteins by Fold Change

ONTOLOGY	ID	Description	setSize	NES	pvalue	qvalue
CC	GO:0031090	organelle membrane	125	-2.4661943	1.60E-05	0.01457332
BP	GO:0090407	organophosphate biosynthetic process	18	-2.2681162	0.00040511	0.12976844
MF	GO:0045296	cadherin binding	22	2.38235137	0.00042707	0.12976844
BP	GO:0019637	organophosphate metabolic process	26	-2.1277084	0.00105689	0.22341234
BP	GO:0033554	cellular response to stress	68	2.23086827	0.00182735	0.22341234
CC	GO:0031984	organelle subcompartment	55	-2.0379365	0.00188143	0.22341234
BP	GO:0006281	DNA repair	24	2.12890291	0.00208218	0.22341234
BP	GO:0007015	actin filament organization	14	2.14039643	0.00244777	0.22341234
CC	GO:0015629	actin cytoskeleton	20	2.16036915	0.00253327	0.22341234
CC	GO:0031967	organelle envelope	60	-1.9859957	0.00285063	0.22341234
CC	GO:0031975	envelope	60	-1.9859957	0.00285063	0.22341234
BP	GO:0030036	actin cytoskeleton organization	22	2.04913542	0.00294099	0.22341234

BP	GO:0072594	<i>establishment of protein localization to organelle</i>	15	-2.1177775	0.00340767	0.22600532
MF	GO:0050839	<i>cell adhesion molecule binding</i>	35	2.11582335	0.00352162	0.22600532
BP	GO:0097435	<i>supramolecular fiber organization</i>	27	2.00463801	0.00394858	0.22600532
BP	GO:0006364	<i>rRNA processing</i>	11	-1.9297517	0.0042366	0.22600532
BP	GO:0016072	<i>rRNA metabolic process</i>	11	-1.9297517	0.0042366	0.22600532
BP	GO:0055086	<i>nucleobase-containing small molecule metabolic process</i>	23	-1.9732876	0.004834	0.22600532
BP	GO:0045184	<i>establishment of protein localization</i>	74	-1.8950743	0.00500675	0.22600532
CC	GO:0005911	<i>cell-cell junction</i>	23	1.90995398	0.0052032	0.22600532
BP	GO:0006974	<i>DNA damage response</i>	30	2.0375331	0.00539228	0.22600532
BP	GO:0046907	<i>intracellular transport</i>	58	-1.8706289	0.0054544	0.22600532
CC	GO:0098798	<i>mitochondrial protein-containing complex</i>	23	-1.932679	0.00673423	0.26690358
BP	GO:0003012	<i>muscle system process</i>	11	-1.8554701	0.00727447	0.26804413
CC	GO:0005802	<i>trans-Golgi network</i>	11	-1.8489424	0.00750003	0.26804413
BP	GO:0006163	<i>purine nucleotide metabolic process</i>	20	-1.963233	0.00793918	0.26804413
BP	GO:0072521	<i>purine-containing compound metabolic process</i>	20	-1.963233	0.00793918	0.26804413
CC	GO:0070161	<i>anchoring junction</i>	36	1.87488152	0.00861478	0.27579097
BP	GO:0009165	<i>nucleotide biosynthetic process</i>	14	-1.9126678	0.00907626	0.27579097
BP	GO:1901293	<i>nucleoside phosphate biosynthetic process</i>	14	-1.9126678	0.00907626	0.27579097
BP	GO:0030029	<i>actin filament-based process</i>	25	1.86487966	0.01132761	0.31528218
CC	GO:0031966	<i>mitochondrial membrane</i>	35	-1.8726278	0.01138112	0.31528218
CC	GO:1902495	<i>transmembrane transporter complex</i>	10	-1.9016897	0.01161617	0.31528218
BP	GO:0006753	<i>nucleoside phosphate metabolic process</i>	22	-1.8617099	0.01210523	0.31528218
BP	GO:0009117	<i>nucleotide metabolic process</i>	22	-1.8617099	0.01210523	0.31528218
BP	GO:0019693	<i>ribose phosphate metabolic process</i>	19	-1.8222078	0.01250179	0.31553902

CC	GO:1990351	<i>transporter complex</i>	11	-1.7710599	0.01280739	0.31553902
MF	GO:0044877	<i>protein-containing complex binding</i>	58	1.76705221	0.01471662	0.32990191
MF	GO:0005102	<i>signaling receptor binding</i>	35	1.82995183	0.01486173	0.32990191
MF	GO:0003779	<i>actin binding</i>	15	1.97648608	0.01496149	0.32990191
BP	GO:0044255	<i>cellular lipid metabolic process</i>	22	-1.8277007	0.01500215	0.32990191
BP	GO:0006164	<i>purine nucleotide biosynthetic process</i>	13	-1.811131	0.01556177	0.32990191
BP	GO:0072522	<i>purine-containing compound biosynthetic process</i>	13	-1.811131	0.01556177	0.32990191
MF	GO:0005178	<i>integrin binding</i>	10	1.84252488	0.01694565	0.33765496
CC	GO:0005912	<i>adherens junction</i>	10	1.83862452	0.01736149	0.33765496
MF	GO:0016491	<i>oxidoreductase activity</i>	38	-1.7678575	0.01827479	0.33765496
CC	GO:0098791	<i>Golgi apparatus subcompartment</i>	13	-1.7799959	0.01835212	0.33765496
BP	GO:0046390	<i>ribose phosphate biosynthetic process</i>	13	-1.7782674	0.01856676	0.33765496
CC	GO:0000228	<i>nuclear chromosome</i>	10	1.82860666	0.01860901	0.33765496
CC	GO:0005788	<i>endoplasmic reticulum lumen</i>	19	1.69453188	0.01860907	0.33765496
CC	GO:0005789	<i>endoplasmic reticulum membrane</i>	42	-1.7981505	0.01889074	0.33765496
BP	GO:0006511	<i>ubiquitin-dependent protein catabolic process</i>	23	1.65415196	0.02075109	0.35395721
BP	GO:0019941	<i>modification-dependent protein catabolic process</i>	23	1.65415196	0.02075109	0.35395721
BP	GO:0033108	<i>mitochondrial respiratory chain complex assembly</i>	11	-1.6967088	0.02096767	0.35395721
BP	GO:0018193	<i>peptidyl-amino acid modification</i>	22	1.65162257	0.02281106	0.36827506
CC	GO:0042175	<i>nuclear outer membrane-endoplasmic reticulum membrane network</i>	43	-1.7276121	0.02302782	0.36827506
CC	GO:0098827	<i>endoplasmic reticulum subcompartment</i>	43	-1.7276121	0.02302782	0.36827506
BP	GO:0006793	<i>phosphorus metabolic process</i>	62	-1.7196601	0.02393125	0.3761245
BP	GO:0042254	<i>ribosome biogenesis</i>	17	-1.7466088	0.02438635	0.37678101
BP	GO:0009896	<i>positive regulation of catabolic process</i>	21	1.67851744	0.02822949	0.42889009
CC	GO:0016020	<i>membrane</i>	219	-1.5862895	0.02890792	0.43199761
BP	GO:0009057	<i>macromolecule catabolic process</i>	54	1.72432638	0.02941432	0.43247533

BP	GO:0009259	<i>ribonucleotide metabolic process</i>	18	- 1.6729317	0.03006929	0.43508782
CC	GO:0005840	<i>ribosome</i>	12	- 1.7190252	0.03117479	0.44403562
BP	GO:0043632	<i>modification-dependent macromolecule catabolic process</i>	24	1.57321234	0.03239753	0.45435239
CC	GO:0031965	<i>nuclear membrane</i>	16	- 1.6766938	0.03536068	0.45816514
CC	GO:0098796	<i>membrane protein complex</i>	41	- 1.7100358	0.03557814	0.45816514
BP	GO:0003008	<i>system process</i>	42	- 1.6922752	0.03680203	0.45816514
BP	GO:0006796	<i>phosphate-containing compound metabolic process</i>	61	- 1.6422701	0.03694581	0.45816514
CC	GO:0005654	<i>nucleoplasm</i>	138	1.49031548	0.03721841	0.45816514
BP	GO:0140053	<i>mitochondrial gene expression</i>	13	- 1.6417042	0.03730273	0.45816514
CC	GO:0005743	<i>mitochondrial inner membrane</i>	27	- 1.6490795	0.03815364	0.45816514
BP	GO:0072657	<i>protein localization to membrane</i>	24	- 1.6813937	0.03831938	0.45816514
BP	GO:0051234	<i>establishment of localization</i>	125	- 1.6177874	0.0385051	0.45816514
BP	GO:0022900	<i>electron transport chain</i>	14	- 1.6037215	0.03896104	0.45816514
BP	GO:0007600	<i>sensory perception</i>	11	- 1.6151579	0.03939394	0.45816514
MF	GO:0005215	<i>transporter activity</i>	26	- 1.6352987	0.03956344	0.45816514
BP	GO:0034470	<i>ncRNA processing</i>	16	- 1.652479	0.03960396	0.45816514
BP	GO:0009260	<i>ribonucleotide biosynthetic process</i>	12	- 1.6011749	0.03970588	0.45816514
BP	GO:0002252	<i>immune effector process</i>	12	- 1.593071	0.04117647	0.46460608
BP	GO:0030163	<i>protein catabolic process</i>	38	1.67654249	0.04128342	0.46460608
CC	GO:0005740	<i>mitochondrial envelope</i>	37	- 1.670694	0.04230769	0.46471984
BP	GO:0051960	<i>regulation of nervous system development</i>	17	- 1.6116685	0.04231312	0.46471984
BP	GO:0044281	<i>small molecule metabolic process</i>	52	- 1.617044	0.04320988	0.46723249
MF	GO:0140677	<i>molecular function activator activity</i>	27	1.60171954	0.043567	0.46723249
CC	GO:0044391	<i>ribosomal subunit</i>	10	- 1.5633614	0.04658385	0.48860762

BP	GO:0006886	<i>intracellular protein transport</i>	36	-1.6571647	0.04663212	0.48860762
BP	GO:0006950	<i>response to stress</i>	110	1.42814421	0.04727709	0.48973641
MF	GO:0003713	<i>transcription coactivator activity</i>	13	1.56930332	0.04831573	0.49487198

Table S11: Arglu1 Proteomics GSEA table

source	term_name	term_id	padj	-LOG10 padj	term_size	query_size	intersection_size
GO:CC	<i>cytoplasm</i>	GO:0005737	5.28E-33	32.278	12345	417	351
GO:MF	<i>protein binding</i>	GO:0005515	1.37E-23	22.864	14838	415	388
GO:CC	<i>intracellular anatomical structure</i>	GO:0005622	2.78E-20	19.556	17834	417	403
HPA	<i>cervix</i>	HPA:0630000	1.55E-15	14.810	6930	325	274
GO:CC	<i>intracellular organelle</i>	GO:0043229	3.16E-15	14.500	16018	417	373
GO:CC	<i>organelle</i>	GO:0043226	6.75E-15	14.171	16808	417	383
GO:CC	<i>membrane-enclosed lumen</i>	GO:0031974	1.06E-14	13.974	6657	417	207
GO:CC	<i>organelle lumen</i>	GO:0043233	1.06E-14	13.974	6657	417	207
GO:CC	<i>intracellular organelle lumen</i>	GO:0070013	1.06E-14	13.974	6657	417	207
GO:BP	<i>organelle organization</i>	GO:0006996	1.14E-14	13.942	3613	410	142
HPA	<i>vagina</i>	HPA:0620000	8.48E-14	13.072	6000	325	247
HPA	<i>vagina; squamous epithelial cells[,â•Low]</i>	HPA:0620221	8.48E-14	13.072	6000	325	247
HPA	<i>fallopian tube</i>	HPA:0210000	8.53E-14	13.069	7466	325	283
GO:CC	<i>membrane-bounded organelle</i>	GO:0043227	1.04E-13	12.984	15671	417	365
HPA	<i>esophagus</i>	HPA:0190000	1.27E-13	12.897	6679	325	264
HPA	<i>esophagus; squamous epithelial cells[,â•Low]</i>	HPA:0190221	1.27E-13	12.897	6679	325	264
HPA	<i>placenta</i>	HPA:0380000	3.33E-13	12.478	7517	325	283
HPA	<i>pancreas</i>	HPA:0350000	5.73E-12	11.242	7409	325	278
GO:BP	<i>cellular component organization or biogenesis</i>	GO:0071840	6.66E-12	11.177	8361	410	241
HPA	<i>esophagus; squamous epithelial cells[,â•Medium]</i>	HPA:0190222	8.56E-12	11.068	5058	325	216

HPA	<i>urinary bladder</i>	HPA:0610000	1.02E-11	10.993	7304	325	275
HPA	<i>urinary bladder; urothelial cells[,â•Low]</i>	HPA:0610831	1.02E-11	10.993	7304	325	275
HPA	<i>cervix; squamous epithelial cells[,â•Low]</i>	HPA:0630221	1.26E-11	10.900	5460	325	227
GO:CC	<i>intracellular membrane-bounded organelle</i>	GO:0043231	1.27E-11	10.895	14605	417	344
GO:CC	<i>cytosol</i>	GO:0005829	1.31E-11	10.883	5487	417	173

Table S12: Arglu1 Proteomics g:Profiler table

Gene Name	Peak Score	Log2 FC	pval	padj	Chr	Annotation	Distance to TSS	Nearest PromoterID	Gene Description	Gene Type
<i>RNU1-3</i>	550.3	6.257	5.068E-19	1.260E-14	chr1_K1270713v1_random	TTS (NR_004400).9	243	NR_004408	RNA, U1 small nuclear 3	snRNA
<i>RNU2-1</i>	1026.4	5.339	3.662E-14	4.554E-10	chr17	TTS (NR_002716).5	307	NR_002716	RNA, U2 small nuclear 1	snRNA
<i>SNORD3D</i>	278.2	5.429	8.872E-14	7.356E-10	chr17	TTS (NR_006882)	650	NR_006882	small nucleolar RNA, C/D box 3D	snoRNA
<i>RNU1-2</i>	333.5	6.065	3.318E-13	2.021E-09	chr1	TTS (NR_004400).4	270	NR_004427	RNA, U1 small nuclear 2	snRNA
<i>WDR74</i>	249.4	6.477	4.062E-13	2.021E-09	chr11	5' UTR (NM_001369451, exon 1 of 12)	213	NM_001369451	WD repeat domain 74	protein-coding
<i>RNU1-3</i>	249.4	5.768	8.772E-13	3.637E-09	chr1	TTS (NR_004400)	463	NR_004408	RNA, U1 small nuclear 3	snRNA
<i>RNU1-3</i>	215.4	5.734	4.413E-12	1.489E-08	chr1_K1270713v1_random	TTS (NR_004400).9	898	NR_004408	RNA, U1 small nuclear 3	snRNA
<i>TFCP2L1</i>	209.2	5.456	4.790E-12	1.489E-08	chr2	Intergenic	210009	NM_014553	transcription factor CP2 like 1	protein-coding
<i>SNHG7</i>	186.7	4.990	3.092E-11	8.545E-08	chr9	non-coding (NR_024542, exon 1 of 4)	105	NR_003672	small nucleolar RNA host gene 7	ncRNA
<i>RNU5E-1</i>	186.3	6.834	5.070E-11	1.261E-07	chr1	TTS (NR_002754)	144	NR_002754	RNA, U5E small nuclear 1	snRNA
<i>RNU12</i>	184.7	4.690	7.180E-11	1.624E-07	chr22	promoter-TSS (NM_032311)	253	NR_029422	RNA, U12 small nuclear	snRNA
<i>RNVU1-18</i>	187.1	5.684	9.809E-11	2.033E-07	chr1	TTS (NR_004400).6	337	NR_004400	RNA, variant U1 small nuclear 18	snRNA
<i>SNORD118</i>	149	5.134	2.061E-10	3.944E-07	chr17	TTS (NR_033294)	133	NR_033294	small nucleolar RNA, C/D box 118	snoRNA
<i>SNORA50C</i>	153.1	6.557	3.224E-10	5.012E-07	chr17	TTS (NM_001288733)	337	NR_002995	small nucleolar RNA, H/ACA box 50C	snoRNA

<i>LOC101928453</i>	165.9	6.690	2.964E-10	5.012E-07	chr10	intron (NR_134578, intron 2 of 3)	111477	NR_120638	uncharacterized LOC101928453	ncRNA
<i>MIR21</i>	195.1	7.499	3.195E-10	5.012E-07	chr17	Intergenic	2580	NR_029493	microRNA 21	ncRNA
<i>RNU1-2</i>	136	6.008	5.271E-10	7.712E-07	chr1	Intergenic	9203	NR_004427	RNA, U1 small nuclear 2	snRNA
<i>RNU5F-1</i>	174.6	4.447	1.048E-09	1.449E-06	chr1	TTS (NR_002753)	168	NR_002753	RNA, U5F small nuclear 1	snRNA
<i>RNU5E-1</i>	147.4	9.588	1.274E-09	1.668E-06	chr1	TTS (NR_002754)	695	NR_002754	RNA, U5E small nuclear 1	snRNA
<i>CCDC85B</i>	134.8	5.669	2.169E-09	2.569E-06	chr11	exon (NM_006848, exon 1 of 1)	194	NM_006848	coiled-coil domain containing 85B	protein-coding
<i>RNU1-1</i>	142.6	9.484	2.098E-09	2.569E-06	chr1	TTS (NR_004400).5	580	NR_004430	RNA, U1 small nuclear 1	snRNA
<i>RNVU1-18</i>	140.2	9.464	2.317E-09	2.620E-06	chr1	Intergenic	33200	NR_004400	RNA, variant U1 small nuclear 18	snRNA
<i>PHLDA1</i>	134.8	9.411	2.880E-09	3.114E-06	chr12	exon (NM_007350, exon 1 of 2)	395	NM_007350	pleckstrin homology like domain family A member 1	protein-coding
<i>RNU1-3</i>	131.3	9.338	3.778E-09	3.257E-06	chr1	promoter-TSS (NR_004400).4	-147	NR_004408	RNA, U1 small nuclear 3	snRNA
<i>GNA12</i>	144.4	7.086	3.797E-09	3.257E-06	chr7	Intergenic	-30036	NM_001293092	G protein subunit alpha 12	protein-coding

Table S13: ChIP-Seq RNAP2 HT/IgG Peaks by pvalue

Gene Name	Log2 FC	pvalue	padj	Chr	Peak Score	Annotation	Distance to TSS	Nearest PromoterID	Gene Description	Gene Type
<i>RNU1-3</i>	6.358	2.443E-26	7.275E-22	chr1_K1270713v1_random	599.6	TTS (NR_004400).9	231	NR_004408	RNA, U1 small nuclear 3	snRNA
<i>RNU1-3</i>	6.028	4.746E-22	7.067E-18	chr1	421.6	TTS (NR_004400).4	261	NR_004408	RNA, U1 small nuclear 3	snRNA
<i>WDR74</i>	6.655	2.035E-18	2.021E-14	chr11	304.5	5' UTR (NM_001369451, exon 1 of 12)	192	NM_001369451	WD repeat domain 74	protein-coding
<i>RNU1-4</i>	5.840	4.714E-18	3.510E-14	chr1	281.8	TTS (NR_004400)	458	NR_004421	RNA, U1 small nuclear 4	snRNA
<i>SNORD3D</i>	5.324	1.088E-17	6.480E-14	chr17	283.8	TTS (NR_006882)	651	NR_006882	small nucleolar RNA, C/D box 3D	snoRNA
<i>SNORD118</i>	5.712	1.471E-17	7.303E-14	chr17	223.3	TTS (NR_033294)	144	NR_033294	small nucleolar RNA, C/D box 118	snoRNA
<i>TFCP2L1</i>	5.789	2.333E-16	9.926E-13	chr2	287.4	Intergenic	209983	NM_014553	transcription factor CP2 like 1	protein-

										codin g
RNU5F-1	4.6 60	2.3 27E -15	8.6 65E -12	chr1	211. 1	TTS (NR_002753)	167	NR_002 753	RNA, U5F small nuclear 1	snRN A
RNU12	6.8 22	4.7 66E -15	1.5 77E -11	chr22	235. 1	promoter-TSS (NM_032311)	167	NR_029 422	RNA, U12 small nuclear	snRN A
RNU4-1	5.1 90	1.2 71E -14	3.7 87E -11	chr12	174. 6	TTS (NR_003925)	134	NR_003 925	RNA, U4 small nuclear 1	snRN A
RNU1-3	5.4 65	1.8 73E -14	5.0 71E -11	chr1_KI27 0713v1_ra ndom	228. 8	TTS (NR_004400).9	876	NR_004 408	RNA, U1 small nuclear 3	snRN A
RNU2-1	5.4 34	3.0 24E -14	7.5 06E -11	chr17	121 9.9	TTS (NR_002716).5	307	NR_002 716	RNA, U2 small nuclear 1	snRN A
RNU5E-1	7.1 11	4.8 19E -14	1.1 04E -10	chr1	217. 5	TTS (NR_002754)	210	NR_002 754	RNA, U5E small nuclear 1	snRN A
SNHG7	5.0 07	3.8 58E -13	8.2 07E -10	chr9	219. 7	promoter-TSS (NR_024542)	78	NR_024 542	small nucleolar RNA host gene 7	ncRN A
HIST2H 2AB	4.1 20	4.5 79E -13	9.0 91E -10	chr1	281. 6	TTS (NM_003517)	219	NM_175 065	histone cluster 2 H2A family member b	protei n- codin g
CCDC8 5B	5.8 08	7.4 51E -13	1.3 87E -09	chr11	158. 6	5' UTR (NM_006848, exon 1 of 1)	106	NM_006 848	coiled-coil domain containing 85B	protei n- codin g
THOC6	6.1 81	1.9 14E -12	3.3 53E -09	chr16	167. 4	promoter-TSS (NM_024339)	-156	NM_001 347704	THO complex 6	protei n- codin g
CCDC8 5B	4.3 46	2.1 98E -12	3.5 49E -09	chr11	159. 4	TTS (NM_001300856)	460	NM_006 848	coiled-coil domain containing 85B	protei n- codin g
PHLDA 1	6.2 67	2.2 64E -12	3.5 49E -09	chr12	178. 8	exon (NM_007350, exon 1 of 2)	305	NM_007 350	pleckstrin homology like domain family A member 1	protei n- codin g
SNORA 50C	4.9 00	3.9 92E -12	5.9 45E -09	chr17	127. 4	TTS (NM_001288733)	922	NR_002 995	small nucleolar RNA, H/ACA box 50C	snoR NA
SNORD 104	3.8 03	6.0 99E -12	8.6 50E -09	chr17	158. 9	promoter-TSS (NR_004380)	5	NR_004 380	small nucleolar RNA, C/D box 104	snoR NA
LINC01 410	4.7 90	1.0 70E -11	1.4 49E -08	chr9	123. 2	intron (NR_121647, intron 1 of 4)	1172	NR_121 647	long intergenic non- protein coding RNA 1410	ncRN A
SNORA 10	3.9 87	1.1 41E -11	1.4 78E -08	chr16	170. 5	TTS (NM_002952)	1224	NR_002 327	small nucleolar RNA, H/ACA box 10	snoR NA
KRBA2	6.5 87	1.4 73E -11	1.8 28E -08	chr17	161. 9	TTS (NM_001315530)	153	NM_001 304947	KRAB-A domain containing 2	protei n- codin g
FOSL1	5.4 29	1.5 36E -11	1.8 30E -08	chr11	141. 5	exon (NM_001300856, exon 1 of 3)	117	NM_001 300856	FOS like 1, AP-1 transcription factor subunit	protei n- codin g

Table S14: ChIP-Seq RNAP2 HTA/IgG Peaks by pvalue

Homer Known Motif Enrichment Results (Arglu1.Motifs/)

[Homer de novo Motif Results](#)
[Gene Ontology Enrichment Results](#)
[Known Motif Enrichment Results \(txt file\)](#)

Total Target Sequences = 6548, Total Background Sequences = 42934

Rank	Motif	Name	P-value	log P-value	q-value (Benjamini)	# Target Sequences with Motif	% of Targets Sequences with Motif	# Background Sequences with Motif	% of Background Sequences with Motif
1		Fra1(bZIP)/BT549-Fra1-ChIP-Seq(GSE46166)/Homer	1e-56	-1.310e+02	0.0000	330.0	5.04%	785.7	1.83%
2		Fra2(bZIP)/Striatum-Fra2-ChIP-Seq(GSE43429)/Homer	1e-54	-1.262e+02	0.0000	314.0	4.79%	741.9	1.73%
3		Fosl2(bZIP)/3T3L1-Fosl2-ChIP-Seq(GSE56872)/Homer	1e-54	-1.251e+02	0.0000	263.0	4.01%	556.4	1.30%
4		Fos(bZIP)/TSC-Fos-ChIP-Seq(GSE110950)/Homer	1e-51	-1.190e+02	0.0000	328.0	5.01%	824.0	1.92%
5		Jun-AP1(bZIP)/K562-cJun-ChIP-Seq(GSE31477)/Homer	1e-51	-1.179e+02	0.0000	220.0	3.36%	427.4	1.00%
6		Atf3(bZIP)/GBM-ATF3-ChIP-Seq(GSE3912)/Homer	1e-50	-1.159e+02	0.0000	358.0	5.46%	957.5	2.23%
7		JunB(bZIP)/DendriticCells-Junb-ChIP-Seq(GSE36099)/Homer	1e-50	-1.153e+02	0.0000	323.0	4.93%	818.5	1.91%
8		AP-1(bZIP)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	1e-47	-1.096e+02	0.0000	391.0	5.97%	1125.2	2.62%
9		BATF(bZIP)/Th17-BATF-ChIP-Seq(GSE39756)/Homer	1e-46	-1.064e+02	0.0000	341.0	5.21%	929.5	2.17%
10		CTCF(Zf)/CD4+-CTCF-ChIP-Seq(Barski_et_al.)/Homer	1e-43	-9.970e+01	0.0000	291.0	4.44%	754.7	1.76%
11		Bach2(bZIP)/OCILy7-Bach2-ChIP-Seq(GSE44420)/Homer	1e-31	-7.255e+01	0.0000	173.0	2.64%	399.5	0.93%
12		BORIS(Zf)/K562-CTCF-ChIP-Seq(GSE32465)/Homer	1e-20	-4.768e+01	0.0000	539.0	8.23%	2316.2	5.40%
13		NF-E2(bZIP)/K562-NFE2-ChIP-Seq(GSE31477)/Homer	1e-19	-4.494e+01	0.0000	64.0	0.98%	102.6	0.24%
14		Rfx2(HTH)/LoVo-RFX2-ChIP-Seq(GSE49402)/Homer	1e-17	-4.133e+01	0.0000	163.0	2.49%	495.5	1.15%

file:///Users/rafehelwer/Docs/Homer/Motif_files/Motifs.Arglu1.IgG/knownResults.html

15		RFX(HTH)/K562-RFX3-ChIP-Seq(SRA012198)/Homer	1e-17	-4.040e+01	0.0000	160.0	2.44%	487.8	1.14%
16		Nrf2(bZIP)/Lymphoblast-Nrf2-ChIP-Seq(GSE37589)/Homer	1e-17	-3.974e+01	0.0000	54.0	0.82%	83.7	0.20%
17		Bach1(bZIP)/K562-Bach1-ChIP-Seq(GSE31477)/Homer	1e-12	-2.876e+01	0.0000	38.0	0.58%	58.3	0.14%
18		Rfx1(HTH)/NPC-H3K4me1-ChIP-Seq(GSE16256)/Homer	1e-11	-2.561e+01	0.0000	207.0	3.16%	819.4	1.91%
19		Rfx5(HTH)/GM12878-Rfx5-ChIP-Seq(GSE31477)/Homer	1e-10	-2.506e+01	0.0000	253.0	3.86%	1062.6	2.48%
20		NFE2L2(bZIP)/HepG2-NFE2L2-ChIP-Seq(Encode)/Homer	1e-10	-2.387e+01	0.0000	27.0	0.41%	36.6	0.09%
21		X-box(HTH)/NPC-H3K4me1-ChIP-Seq(GSE16256)/Homer	1e-10	-2.370e+01	0.0000	107.0	1.63%	349.9	0.82%
22		Fli1(ETS)/CD8-FLI-ChIP-Seq(GSE20898)/Homer	1e-10	-2.367e+01	0.0000	1214.0	18.53%	6679.4	15.56%
23		ERG(ETS)/VCaP-ERG-ChIP-Seq(GSE14097)/Homer	1e-7	-1.782e+01	0.0000	1220.0	18.62%	6896.3	16.07%
24		MafK(bZIP)/C2C12-MafK-ChIP-Seq(GSE36030)/Homer	1e-7	-1.617e+01	0.0000	156.0	2.38%	654.2	1.52%
25		RUNX(Runt)/HPC7-Runx1-ChIP-Seq(GSE22178)/Homer	1e-6	-1.568e+01	0.0000	351.0	5.36%	1737.6	4.05%
26		ETV4(ETS)/HepG2-ETV4-ChIP-Seq(ENCODE)/Homer	1e-6	-1.529e+01	0.0000	1339.0	20.44%	7725.9	18.00%
27		RUNX2(Runt)/PCa-RUNX2-ChIP-Seq(GSE38889)/Homer	1e-6	-1.505e+01	0.0000	333.0	5.08%	1646.6	3.84%
28		NF1-halfsite(CTF)/LNCaP-NF1-ChIP-Seq(Unpublished)/Homer	1e-6	-1.423e+01	0.0000	1296.0	19.78%	7498.2	17.47%
29		PU.1-IRF(ETS:IRF)/Bcell-PU.1-ChIP-Seq(GSE21512)/Homer	1e-5	-1.361e+01	0.0000	674.0	10.29%	3693.3	8.61%
30		ETS1(ETS)/Jurkat-ETS1-ChIP-Seq(GSE17954)/Homer	1e-5	-1.360e+01	0.0000	937.0	14.30%	5297.6	12.34%
31		ETV1(ETS)/GIST48-ETV1-ChIP-Seq(GSE22441)/Homer	1e-5	-1.359e+01	0.0000	1240.0	18.93%	7174.9	16.72%
32		EWS:ERG-fusion(ETS)/CADO_ES1-	1e-5	-1.331e+01	0.0000	309.0	4.72%	1542.7	3.59%

file:///Users/rafehelwer/Docs/Homer/Motif_files/Motifs.Arglu1.IgG/knownResults.html

		EWS:ERG-ChIP-Seq(SRA014231)/Homer							
33		RUNX1(Runt)/Jurkat-RUNX1-ChIP-Seq(GSE29180)/Homer	1e-5	-1.236e+01	0.0001	430.0	6.56%	2268.6	5.29%
34		Elk1(ETS)/Hela-Elk1-ChIP-Seq(GSE31477)/Homer	1e-5	-1.202e+01	0.0001	858.0	13.10%	4867.8	11.34%
35		Eiv2(ETS)/ES-ER71-ChIP-Seq(GSE59402)/Homer	1e-5	-1.192e+01	0.0001	731.0	11.16%	4092.6	9.54%
36		GABPA(ETS)/Jurkat-GABPA-ChIP-Seq(GSE17954)/Homer	1e-5	-1.171e+01	0.0001	900.0	13.74%	5137.0	11.97%
37		Elk4(ETS)/Hela-Elk4-ChIP-Seq(GSE31477)/Homer	1e-4	-1.122e+01	0.0002	851.0	12.99%	4853.7	11.31%
38		EWS:FLI1-fusion(ETS)/SK_N_MC-EWS:FLI1-ChIP-Seq(SRA014231)/Homer	1e-4	-1.045e+01	0.0003	481.0	7.34%	2623.8	6.11%
39		MafA(bZIP)/Islet-MafA-ChIP-Seq(GSE30298)/Homer	1e-4	-1.038e+01	0.0004	470.0	7.17%	2559.8	5.96%
40		TEAD1(TEAD)/HepG2-TEAD1-ChIP-Seq(Encode)/Homer	1e-3	-8.909e+00	0.0015	358.0	5.46%	1930.7	4.50%
41		RUNX-AML(Runt)/CD4+-PolII-ChIP-Seq(Barski_et_al.)/Homer	1e-3	-8.860e+00	0.0015	311.0	4.75%	1652.6	3.85%
42		TEAD(TEA)/Fibroblast-PU.1-ChIP-Seq(Unpublished)/Homer	1e-3	-8.715e+00	0.0017	195.0	2.98%	978.5	2.28%
43		ZNF528(Zf)/HEK293-ZNF528.GFP-ChIP-Seq(GSE58341)/Homer	1e-3	-8.188e+00	0.0028	7.0	0.11%	8.9	0.02%
44		ELF1(ETS)/Jurkat-ELF1-ChIP-Seq(SRA014231)/Homer	1e-3	-8.011e+00	0.0033	782.0	11.94%	4556.0	10.62%
45		PU.1(ETS)/ThioMac-PU.1-ChIP-Seq(GSE21512)/Homer	1e-3	-7.611e+00	0.0048	291.0	4.44%	1566.4	3.65%
46		FOXP1(Forkhead)/H9-FOXP1-ChIP-Seq(GSE31006)/Homer	1e-3	-7.553e+00	0.0050	97.0	1.48%	446.9	1.04%
47		Fox:Ebox(Forkhead.bHLH)/Panc1-Foxa2-ChIP-Seq(GSE47459)/Homer	1e-3	-7.089e+00	0.0078	272.0	4.15%	1467.1	3.42%
48		BMYB(HTH)/Hela-BMYB-ChIP-Seq(GSE27030)/Homer	1e-2	-6.851e+00	0.0097	680.0	10.38%	3972.9	9.26%
49		STAT4(Stat)/CD4-Stat4-ChIP-Seq(GSE22104)/Homer	1e-2	-6.822e+00	0.0098	336.0	5.13%	1859.5	4.33%

file:///Users/rafehelwer/Docs/Homer/Motif_files/Motifs.Arglu1.IgG/knownResults.html

50		TEAD4(TEA)/Tropoblast-Tea4-ChIP-Seq(GSE37350)/Homer	1e-2	-6.618e+00	0.0118	374.0	5.71%	2096.3	4.88%
51		Rbpj1(?)/Panc1-Rbpj1-ChIP-Seq(GSE47459)/Homer	1e-2	-6.558e+00	0.0122	776.0	11.85%	4585.7	10.68%
52		MYB(HTH)/ERMYB-Myb-ChIP-Seq(GSE22095)/Homer	1e-2	-6.536e+00	0.0123	969.0	14.79%	5799.2	13.51%
53		CEBP:AP1(bZIP)/ThioMac-CEBPb-ChIP-Seq(GSE21512)/Homer	1e-2	-6.337e+00	0.0147	186.0	2.84%	977.0	2.28%
54		Smad3(MAD)/NPC-Smad3-ChIP-Seq(GSE36673)/Homer	1e-2	-6.278e+00	0.0153	1615.0	24.65%	9925.3	23.13%
55		Nanog(Homeobox)/mES-Nanog-ChIP-Seq(GSE11724)/Homer	1e-2	-6.103e+00	0.0179	1543.0	23.55%	9476.5	22.08%
56		p53(p53)/Saos-p53-ChIP-Seq(GSE15780)/Homer	1e-2	-6.085e+00	0.0179	47.0	0.72%	197.0	0.46%
57		p53(p53)/Saos-p53-ChIP-Seq/Homer	1e-2	-6.085e+00	0.0179	47.0	0.72%	197.0	0.46%
58		AMYB(HTH)/Testes-AMYB-ChIP-Seq(GSE44588)/Homer	1e-2	-6.077e+00	0.0179	764.0	11.66%	4534.8	10.57%
59		Stat3(Stat)/mES-Stat3-ChIP-Seq(GSE11431)/Homer	1e-2	-6.014e+00	0.0182	343.0	5.24%	1928.3	4.49%
60		FOXK1(Forkhead)/HEK293-FOXK1-ChIP-Seq(GSE51673)/Homer	1e-2	-5.971e+00	0.0187	211.0	3.22%	1134.2	2.64%
61		SPDEF(ETS)/VCaP-SPDEF-ChIP-Seq(SRA014231)/Homer	1e-2	-5.876e+00	0.0202	664.0	10.14%	3918.9	9.13%
62		ELF3(ETS)/PDAC-ELF3-ChIP-Seq(GSE64557)/Homer	1e-2	-5.835e+00	0.0208	385.0	5.88%	2191.3	5.11%
63		Hoxa10(Homeobox)/ChickenMSG-Hoxa10.Flag-ChIP-Seq(GSE86088)/Homer	1e-2	-5.789e+00	0.0214	112.0	1.71%	559.1	1.30%
64		ETS(ETS)/Promoter/Homer	1e-2	-5.776e+00	0.0214	462.0	7.05%	2667.4	6.22%
65		YY1(Zf)/Promoter/Homer	1e-2	-5.577e+00	0.0256	135.0	2.06%	696.9	1.62%
66		HNF6(Homeobox)/Liver-Hnf6-ChIP-Seq(ERP000394)/Homer	1e-2	-5.563e+00	0.0256	99.0	1.51%	489.7	1.14%
67		Foxo3(Forkhead)/U2OS-Foxo3-ChIP-Seq(E-MTAB-2701)/Homer	1e-2	-5.335e+00	0.0317	137.0	2.09%	713.8	1.66%

file:///Users/rafehelwer/Docs/Homer/Motif_files/Motifs.Arglu1.IgG/knownResults.html

68		Rfx6(HTH)/Min6b1-Rfx6.HA-ChIP-Seq(GSE62844)/Homer	1e-2	-5.324e+00	0.0317	870.0	13.28%	5243.4	12.22%
69		STAT5(Stat)/mCD4+-Stat5-ChIP-Seq(GSE12346)/Homer	1e-2	-4.979e+00	0.0439	103.0	1.57%	523.2	1.22%
70		TEAD3(TEA)/HepG2-TEAD3-ChIP-Seq(Encode)/Homer	1e-2	-4.886e+00	0.0475	377.0	5.75%	2180.0	5.08%
71		FoxL2(Forkhead)/Ovary-FoxL2-ChIP-Seq(GSE60858)/Homer	1e-2	-4.789e+00	0.0516	126.0	1.92%	661.2	1.54%
72		Stat3+il21(Stat)/CD4-Stat3-ChIP-Seq(GSE19198)/Homer	1e-2	-4.752e+00	0.0528	368.0	5.62%	2129.4	4.96%
73		ZNF416(Zf)/HEK293-ZNF416.GFP-ChIP-Seq(GSE58341)/Homer	1e-2	-4.746e+00	0.0528	1197.0	18.27%	7360.0	17.15%
74		ELF5(ETS)/T47D-ELF5-ChIP-Seq(GSE30407)/Homer	1e-2	-4.736e+00	0.0528	459.0	7.01%	2694.5	6.28%
75		Foxo1(Forkhead)/RAW-Foxo1-ChIP-Seq(Fan_et_al.)/Homer	1e-2	-4.626e+00	0.0575	701.0	10.70%	4217.8	9.83%

Table S15: Arglu1 ChIP-Seq Known Motif Enrichment using Homer

Name	Arglu1 logPval	HTA logPval	HT logPval	Arglu1 %	HTA %	HT %
cpglisland	-13118.23	-30366.61	-32650.78	53.10	43.39	45.06
promoters	-5377.85	-17085	-17884	21.77	24.41	24.68
utr5	-2908.89	-10786.02	-10758.91	11.77	15.41	14.85
exons	-1860.06	-6707.63	-6494.78	7.53	9.58	8.96
protein-coding	-1721.29	-6316.49	-6039.68	6.97	9.02	8.33
coding	-810.46	-2512.02	-2477.18	3.28	3.59	3.42
ncRNA	-129.81	-451.22	-477.56	0.53	0.64	0.66
rRNA	-33.32	0.73	0.73	0.13	0.00	0.00
pseudo	-18.27	-41.07	-49	0.07	0.06	0.07
miRNA	-4.98	-24.44	-30.31	0.02	0.03	0.04
tts	-1.08	-385.02	-243.76	0.00	0.55	0.34
intergenic	0	0	0	0.00	0.00	0.00
scRNA	0	0.01	0.01	0.00	0.00	0.00
snRNA	0.04	-19.97	-12.09	0.00	0.03	0.02
snoRNA	0.2	-2.13	-2.13	0.00	0.00	0.00
utr3	33.66	4.46	32.47	-0.14	-0.01	-0.04

gaps	551.17	1644.05	1641.52	-2.23	-2.35	-2.27
introns	692.86	3058.5	2977.19	-2.80	-4.37	-4.11
Total	-24706.31	-69989.87	-72468.26			

Table S16: Genome Ontology Comparison

Gene	Prot_pval	logFC	logPS	logPI
<i>AGPS</i>	0.033	-4.045	2.570	5.979
<i>AK4</i>	0.026	-2.683	3.536	7.646
<i>ALDH1A3</i>	0.007	-6.053	2.007	4.343
<i>ARL6IP1</i>	0.005	-3.146	2.738	4.195
<i>ARMC8</i>	0.005	-3.165	3.288	7.664
<i>BACH1</i>	0.028	-3.029	2.070	5.074
<i>BAD</i>	0.028	-3.111	2.696	2.873
<i>BASP1</i>	0.026	5.760	4.483	6.898
<i>C6orf120</i>	0.026	-2.866	2.655	3.804
<i>CARS2</i>	0.026	-2.683	2.381	8.393
<i>CCNA2</i>	0.003	-3.972	2.776	3.334
<i>CDC37</i>	0.017	-5.798	3.501	2.346
<i>CDC42EP4</i>	0.006	2.216	2.454	12.324
<i>CDH2</i>	0.038	-1.766	3.489	8.324
<i>CDKN2D</i>	0.028	-3.111	2.406	2.783
<i>CEP41</i>	0.026	-2.866	2.220	6.594
<i>CHD3</i>	0.024	-3.842	2.454	1.249
<i>CNN3</i>	0.007	9.156	2.738	6.228
<i>DYNC2H1</i>	0.027	-2.914	2.850	5.694
<i>EIF4G3</i>	0.024	-3.842	2.795	5.407
<i>ELOA</i>	0.026	-2.683	3.624	6.372
<i>EPB41L3</i>	0.001	9.340	3.645	6.263
<i>EPS15</i>	0.031	2.584	3.070	4.275
<i>ERRFI1</i>	0.043	2.431	4.268	4.019
<i>ETV6</i>	0.027	-2.914	2.275	11.183
<i>FAF2</i>	0.005	-3.648	3.039	8.894
<i>FAM136A</i>	0.015	3.556	2.454	3.219
<i>FUT11</i>	0.026	-2.683	3.007	4.106
<i>GALNT2</i>	0.026	-2.866	4.170	5.254
<i>GEMIN4</i>	0.004	-3.558	2.406	4.001
<i>GTPBP1</i>	0.026	-2.866	2.131	10.737

<i>HM13</i>	0.007	-5.085	2.160	3.202
<i>IVD</i>	0.002	-3.251	2.302	5.019
<i>JAM3</i>	0.027	-2.914	3.233	6.734
<i>LACTB</i>	0.005	-3.146	2.353	3.810
<i>LEMD2</i>	0.026	-2.866	2.922	3.425
<i>LRBA</i>	0.027	-2.914	2.070	7.950
<i>MAP2K1</i>	0.016	-5.121	2.635	10.620
<i>MKI67</i>	0.001	-4.974	2.070	5.096
<i>MRPL39</i>	0.005	-3.146	2.454	5.618
<i>MRPL40</i>	0.026	-2.866	2.501	5.654
<i>MTRR</i>	0.026	-2.683	2.302	6.927
<i>MYO18A</i>	0.005	-3.146	2.070	7.683
<i>MYOF</i>	0.007	-4.234	2.275	4.752
<i>NAGLU</i>	0.028	-3.029	2.973	7.801
<i>NDUFS3</i>	0.048	-2.585	2.406	1.147
<i>NOP14</i>	0.026	-3.453	2.381	5.017
<i>NPTX1</i>	0.026	-2.683	3.248	5.759
<i>NR3C1</i>	0.026	-2.683	2.546	6.239
<i>PAK4</i>	0.005	-3.146	2.716	7.264
<i>PCNT</i>	0.020	-3.252	2.220	6.889
<i>PDIA6</i>	0.019	6.123	2.275	5.593
<i>PDLIM1</i>	0.003	12.018	2.478	6.073
<i>PEX14</i>	0.026	-2.866	2.991	8.116
<i>PKIG</i>	0.008	-3.901	3.442	3.333
<i>PPM1G</i>	0.003	7.886	2.546	3.980
<i>PPP2R5C</i>	0.028	-3.111	2.328	8.075
<i>PTPN1</i>	0.026	-2.683	3.635	9.749
<i>RAB18</i>	0.026	-2.683	2.353	7.346
<i>RAD23B</i>	0.015	5.277	3.569	6.611
<i>RHEB</i>	0.009	-4.453	2.635	5.588
<i>SH3GL1</i>	0.019	3.626	2.275	3.564
<i>SLC25A13</i>	0.033	-4.045	2.302	12.098
<i>SLC39A14</i>	0.026	-3.453	2.160	6.776
<i>SLFN5</i>	0.026	-2.683	2.101	5.940
<i>SMAD1</i>	0.020	-3.252	2.220	7.495
<i>TAGLN3</i>	0.037	-2.098	2.635	5.045
<i>TET2</i>	0.026	-2.683	2.501	6.025
<i>TFAM</i>	0.001	-3.859	3.406	6.827
<i>TIMM50</i>	0.026	-3.453	2.614	5.751

<i>TM9SF3</i>	0.026	-2.866	3.055	7.235
<i>TMSB10</i>	0.002	-4.565	2.007	1.816
<i>TRIM4</i>	0.028	-3.111	2.430	2.979
<i>TSHZ3</i>	0.035	-3.694	3.417	7.188
<i>TUBB2A</i>	0.028	-3.111	2.070	6.327
<i>TXNDC5</i>	0.005	8.079	2.328	6.192
<i>UACA</i>	0.026	-2.866	2.039	7.003
<i>UAP1L1</i>	0.004	-3.453	2.716	2.288
<i>UBE4B</i>	0.026	-2.683	2.614	8.534
<i>UBR5</i>	0.026	-2.683	2.635	9.473
<i>VIRMA</i>	0.026	-2.683	2.192	3.750
<i>ZNF219</i>	0.028	-3.029	2.160	2.521

Table S17: Proteomics and Arglu1 ChIP-Seq Overlaps