Global Analysis of Mouse Polyomavirus Infection Reveals Dynamic Regulation of Viral and Host Gene Expression and Promiscuous Viral RNA Editing

Seth B. Garren
University of Connecticut - Storrs, sethbgarren@gmail.com

Follow this and additional works at: https://opencommons.uconn.edu/dissertations

Recommended Citation
https://opencommons.uconn.edu/dissertations/935
Global Analysis of Mouse Polyomavirus Infection Reveals Dynamic Regulation of Viral and Host Gene Expression and Promiscuous Viral RNA Editing

Seth B Garren, PhD
University of Connecticut, 2015

Abstract
Murine polyomavirus lytically infects mouse cells, transforms rat cells in culture, and is highly oncogenic in rodents. We have used deep sequencing to follow polyomavirus infection of mouse NIH3T6 cells at various times after infection and analyzed both the viral and cellular transcriptomes. Alignment of sequencing reads to the viral genome illustrated the transcriptional profile of the early-to-late switch with both early-strand and late-strand RNAs being transcribed at all time points. A number of novel insights into viral gene expression emerged from these studies, including the demonstration of widespread RNA editing of viral transcripts at late times in infection. By late times in infection, 359 host genes were seen to be upregulated and 857 were downregulated. Gene ontology analysis indicated transcripts involved in translation, metabolism, RNA processing, DNA methylation, and protein turnover were upregulated while transcripts involved in extracellular adhesion, cytoskeleton, zinc finger binding, SH3 domain, and GTPase activation were downregulated. The levels of a number of long noncoding RNAs were also altered. The long noncoding RNA MALAT1, which is involved in splicing speckles and used as a marker in many late-stage cancers, was noticeably downregulated, while several other abundant noncoding RNAs were strongly upregulated. We discuss these results in light of what
is currently known about the polyoma life cycle and its effects on host cell growth and metabolism.
Global Analysis of Mouse Polyomavirus Infection Reveals Dynamic Regulation of Viral and Host Gene Expression and Promiscuous Viral RNA Editing

Seth B Garren

B.S., University of Connecticut, 2007

A Dissertation
Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy at the University of Connecticut 2015
Global Analysis of Mouse Polyomavirus Infection Reveals Dynamic Regulation of Viral and Host Gene Expression and Promiscuous Viral RNA Editing

Presented by

Seth B Garren, B.S.

Major Advisor

Dr. Gordon G. Carmichael

Associate Advisor

Dr. Sandra K. Weller

Associate Advisor

Dr. Brenton R. Graveley

Associate Advisor

Dr. David I. Dorsky

University of Connecticut

2015
Acknowledgements

My experience through this process has shaped who I am today and I could not have gotten through it without the help and support of so many people. This lab has been like a second home to me and all of the people who have come and gone during my unusually long time here have each influenced me in their own way.

My advisor Dr. Gordon Carmichael has been a reliable source of patience and optimism even when I couldn’t find any evidence that I was ever going to make progress. He showed me a level of excitement for science that I haven’t seen before and our sessions planning and troubleshooting experiments have proven to be the core defining experience of my scientific career. Even when my initial projects failed to bear fruit, he confidently transitioned me to a project that not only gave me experience with new sequencing technology but also allowed me to take a more holistic look at scientific questions. Ironically, it was through my failures that he helped me succeed.

Our lab manager Kim Morris has been both a reliable resource and a good friend. She not only fixes all of the problems in the lab but also seems to be the one everyone goes to when something goes wrong with the building. She has always made the lab feel like a fun and interesting place to be.

I would like to thank Dr. Michael Duff for his help with all of the data analysis on the
polyoma project. His expertise and clever problem solving has proven invaluable in getting our paper submitted.

I would like to thank my committee members Dr. Sandra Weller, Dr. Brenton Graveley, and Dr. David Dorsky for their guidance and suggestions on my project. They have been able to think of questions, and make observations that we sometimes take for granted and the story is all the better for it.

Finally I would like to thank my friends and family who have been supportive and have long since learned to never ask a graduate student when they are going to graduate.
Table of Contents

APPROVAL PAGE i
ACKNOWLEDGEMENTS ii
TABLE OF CONTENTS iv
LIST OF TABLES vi
LIST OF FIGURES vii

I. INTRODUCTION

A. Mouse Polyomavirus Biology

1. Discovery 1

2. Characterization

a. Transformation 4

b. Structure 5

c. Genome 6

d. Mutant Analysis 8

i. Temperature-sensitive mutant 8

ii. Host range nontransforming mutant 11

e. Transcription 15

f. Noncoding control region 18

g. Tumor antigens 28

i. Large T 31

ii. Middle T 34

iii. Small t 38

B. Polyomavirus Life Cycle
1. Virus Attachment and Entry 39
2. Early Phase 40
3. Early-Late Switch 42
4. Late Phase 47
5. Host Cell Lysis 48
C. Thesis Objective and Specific Aims 49

II. MATERIALS AND METHODS
A. Cell Culture and Virus Infection
B. Preparation of TruSeq™ Stranded Library 52
C. Analysis 53

III. RESULTS
A. RNA expression in NIH 3T6 cells infected with the Py59RA strain of mouse polyomavirus 56
B. The polyomavirus transcriptome 61
C. Splicing analysis 71
D. Leader-to-leader splicing 74
E. A-to-I editing of viral RNAs 77
F. Effects of viral infection on host gene expression 85

IV. DISCUSSION 107

V. FUTURE DIRECTIONS 112

VI. REFERENCES 114
# List of Tables

1. Viral RNA splicing events during infection 60
2. Viral splicing events during infection 73
3. Alignment of reads from late leader repeats. 76
4. Single base mismatches across the polyoma genome 82
5. Number of host genes differentially expressed between mock and Py59RA infected samples 87
6. Gene ontology analysis 90
7. Significant changes in noncoding RNAs 99
8. Transfind transcription factor prediction for 36 hour upregulated 104
9. Transfind transcription factor prediction for 36 hour downregulated 106
List of Figures

1. Map of transcription and splicing in the mouse polyoma virus 3
2. Shift in the transcription start site of the early strand late in infection 20
3. Enhancers and host transcription factor binding sites in the polyoma noncoding control region 22
4. Functional regions of polyoma tumor antigen proteins 30
5. Leader-to-leader splicing 46
6. Library sizes 58
7. Py59RA reference sequence 63
8. Alignment of the time course reads to the Py59RA genome (unscaled) 65
9. Alignment of the time course reads to the Py59RA genome (scaled) 68
10. Alignment of 48 hour aphidicolin + and - reads to Py59RA genome 70
11. Reads with up to 3 A-G mismatches as potential ADAR editing sites 80
12. Accumulation of hyper-editing clusters during the time course 86
13. Validation of upregulated genes 94
14. Validation of downregulated genes 96
15. Example of host protein coding transcripts significantly upregulated or downregulated in infected samples compared to mock infection 98
16. Example of host noncoding transcripts significantly upregulated or downregulated in infected samples compared to mock infection 103
I. INTRODUCTION

For over 60 years mouse polyomavirus has proven instrumental in the understanding of both viruses from the Polyomaviridae family as well as the endogenous host factors the virus utilizes to create the tumorigenic phenotype for which it is named. Mouse polyomavirus is a small circular double-stranded DNA virus with a genome that is only 5,327 base pairs. It consists of two divergent transcription units whose products are alternatively spliced to produce mRNAs for the large, middle, and small tumor antigens at early times and the VP1, VP2, and VP3 viral coat proteins at late times (Fig. 1). It is through the interaction of the tumor antigens that the virus can alter host-signaling pathways to drive the cell into S-phase to initiate the DNA replication the virus requires in order to copy itself and undergo an early-late switch to begin producing viral coat proteins. This thesis will focus on the biology of the mouse polyomavirus infection with respect to both virus and host gene expression.

A. Mouse Polyomavirus Biology

1. Discovery

Mouse polyomavirus was discovered in 1953 when a cell free extract from mice infected with a leukemia virus also caused the development of salivary gland tumors when inoculated in newborn mice (Gross 1953). This tumor agent was cultured using mouse embryo cells and was called “polyoma” for its ability to produce a variety of tumors (Stewart 1958). The development of a polyoma plaque assay allowed for pure strains of virus to be quantified by overlaying an infected monolayer of cells with an agar
Figure 1. Map of transcription and splicing in the mouse polyomavirus.

Map of the circular doublestranded DNA genome of the mouse polyomavirus. Early strand transcripts are shown in blue and the late strand transcripts are shown in red. Alternative splicing of the early strand produces the Large, middle, and small t antigens while the late strand splices into VP1, VP2, and VP3. Early and late transcripts are on opposite strands flanking the noncoding control region, which contains the origin of replication (marked as “ori”). Also shown are the polyadenylation sites which are roughly opposite from the noncoding control region on the genome.
nutrient mixture and counting the number of plaques that form when the cells lyse (Winocour 1959). Media from polyoma cultures was shown to induce agglutination in red blood cells allowing for the hemagglutination inhibition assay to be used to detect antibodies against the virus (Eddy 1958) as well as measure the rate of adsorption (Crawford 1962A). Antibodies from a variety of mouse colonies showed a positive reaction for polyoma, indicating that the virus was already present in many mouse stocks independent of strain (Rowe 1959). By culturing primary cells from the kidneys of infected newborn mice, high titers of pure virus could be obtained (Winocour 1963). The development of these techniques made it possible to study the virus in further detail.

2. Characterization

a. Transformation

Depending on the type of cells polyomavirus is cultured on, it either undergoes a permissive “lytic” infection, where the cells produce large quantities of virus and are lysed in the process, or a non-permissive “transforming” infection where the cells take on the same tumor phenotype but do not produce virus. This transformation infection was first observed when polyomavirus was cultured on hamster cells (Vogt 1960). Hamster cells transformed by polyoma could be passaged and caused tumors when injected into live animals but showed no evidence of virus production (Habel 1960, Vogt 1962). Despite this lack of apparent virus production, antibodies to a tumor antigen of viral origin could still be detected in the sera from animals injected with transformed virus-free tumors when it was exposed to cell extracts from both mouse and hamster cells infected with polyomavirus (Habel 1965). In addition, pulse labeled RNA from transformed cells
showed hybridization to viral DNA (Benjamin 1966). Taken together, these experiments suggested that viral gene products were still present in the transformed cells. Experiments with polyoma and a similar primate virus SV40 showed transformed cells to contain viral DNA integrated into the host genome (Sambrook 1968). Exposure of polyoma to DNA damage via radiation before infection showed that the lytic infection of mouse cells is roughly twice as sensitive to radiation compared to the transformation of hamster cells, indicating roughly half of the viral genome is required for transformation (Basilico 1965, Benjamin 1965, Laterjet 1967). Viral mutants were isolated that lost the ability to grow in permissive mouse cell lines but could grow when cultured in mouse cells that had already been transformed by a previous abortive polyoma infection (Benjamin 1970). This indicated that the mutated viral gene products required for transformation were provided by the genes integrated in the transformed cell line from the previous infection, while the gene products necessary for a complete lytic infection were still intact in the mutated virus. At this point transformation appeared to be necessary but not sufficient for a complete lytic infection. We now know these mutants were unable to express middle and small t antigens, which is why they could only infect previously transformed cells. Since the mutants still expressed Large T, VP1, VP2, and VP3, they could still replicate, undergo an early late switch, and complete a lytic infection.

b. Structure

Electron microscopy of polyoma virus particles revealed two types of particles of approximately 450 angstrom in diameter, full virus particles and empty capsids (Wildy
1960). While both particles shared the same structure and ability to induce agglutination in red blood cells, the two types of particles could be separated based on buoyant density by centrifugation in rubidium chloride. Only the particles in the denser gradient were able to infect cells, suggesting that the other gradient contained empty capsids (Crawford 1962B). The viral capsids were observed to be composed of 72 subunits or capsomeres arranged in an icosohedral structure (Klug 1965, Finch 1974). When virus particles were run on SDS-polyacrylamide gels, 7 distinct polypeptides could be identified, VP1, VP2, VP3, and four mouse histone proteins H2a, H2b, H3, and H4 (Frearson 1972, Friedmann 1972, McMillen 1974, Roblin 1971, Seehafer 1974). VP1 is the largest and most abundant capsid protein at ~48kDa, while VP2 and VP3 are 35 and 23 kDa respectively (Hewick 1975). VP2 and 3 share polypeptide sequence with each other but not with VP1 (Friedmann 1976). This is because VP2 and VP3 are contained within the intron of VP1 (Kamen 1980).

c. Genome

Phenol extracted nucleic acid from polyoma infected mouse embryo cultures was shown to lose its infectivity upon treatment with DNase but not RNase indicating a DNA based genome (DiMayorca 1959). The polyoma virus genome showed a high resistance to boiling and demonstrated spontaneous renaturation (Weil 1963A). When extracted polyoma DNA was separated by velocity sedimentation, it produced components consistent with both linear and circular genomes, both of which retained transforming ability (Crawford 1964). Analytical ultracentrifugation of polyoma DNA showed three types of sedimentation corresponding to super-coiled closed circular double helices,
genomes with a single stranded break, and double stranded linear molecules (Radloff 1967, Vinograd 1965, Vinograd 1968). These findings along with electron microscopy demonstrated that the polyoma genome is circular and double-stranded (Weil 1963B). When polyoma genomes were digested with the restriction enzyme Eco RI that only cuts in one place on the genome and treated with the single stranded DNA binding protein 32 from bacteriophage T4, a single denaturation loop per molecule could be visualized by electron microscopy at one of three unique sites corresponding to A/T rich regions (Yaniv 1974). One of these sites would later be identified as part of the noncoding control region. Electron microscopy of intact nucleoprotein complexes revealed chromatin like structures with viral DNA bound to histones (Cremisi 1976). The first attempts to map the polyoma genome used digestion by different combinations of restriction enzymes to determine the order the fragments of viral DNA. This subsequently made it possible to determine the relative location of restriction sites based on the order of the fragments (Chen 1975). Restriction digestion of partially replicated intermediates (under 50% replicated) allowed for the identification of the fragment containing the origin of replication (Griffin 1974). Cleavage of partially replicated intermediates with the restriction enzyme EcoRI made it possible to visualize the bidirectional replication of polyoma using electron microscopy to determine the relative location of the origin (Crawford 1973) and terminus (Crawford 1974) with respect to known restriction digestion sites. Partial denaturation of polyoma fragments from different restriction enzymes allowed for mapping of A/T rich regions using electron microscopy (Lescure 1975). Using a sequencing method (Maxam 1977) that involves chemically attacking specific bases and running terminally radio-labeled DNA fragments on a polyacrylamide
urea gel, the regions encompassing the origin and 5’ end of the early region (Friedman 1978), the early region (Friedman 1979, Soeda 1979), and the late region (Deininger 1979, Arrand 1980) were sequenced and found to have extensive homology with SV40. The sequencing data and restriction maps were combined to make a complete map of the polyoma virus genome (Deininger 1980, Soeda 1980).

d. Mutant analysis

Attempts to ascertain the functions of the different regions of the polyoma genome were aided by the use of two major classes of polyoma mutants, temperature-sensitive and host range nontransforming mutants. Temperature-sensitive mutants function like wild type at low temperatures (32°C) but lose transforming ability and viral DNA synthesis at nonpermissive temperatures (39°C). Host range nontransforming mutants are defective in their ability to transform wild type cells but function normally when infecting cells that have already been transformed. Work with both kinds of mutants has revealed a number of major insights into the function of the early and late regions of the viral genome as well as the process of transformation by the early genes.

i. Temperature-sensitive mutants

In order to isolate temperature sensitive mutants of polyoma, large plaque strains of polyoma were treated with nitric acid long enough to reduce viability of the virus to between 1 in 10,000 to 1 in 100,000 surviving at 32°C. Of the surviving plaques 40 were replated at both 32°C and 39°C and two temperature sensitive lines ts-a and ts-b showed at least a 100 fold difference in plaque formation between 32°C and 39°C (Fried 1965).
One of the temperature-sensitive mutants ts-a was found to induce host DNA replication but failed to transform the host cell or synthesize viral DNA when cultured at nonpermissive temperatures (Fried 1970).

When ts-a was shifted to a nonpermissive temperature after viral DNA production initiated it was found that viral gene products were necessary for the initiation but not completion of viral DNA synthesis (Francke 1973). Assessment of six different temperature sensitive mutants that were isolated using similar methods showed three kinds of mutations. At nonpermissive temperatures class I mutants failed at making infectious viral particles, class II mutants were unable to transform cells or synthesize viral DNA, and class III mutants could transform cells but could not synthesize viral DNA. Mutants unable to synthesize viral DNA were called “early” and mutants that make viral DNA but fail to produce virus particles were called “late” (Eckhart 1969). Both “early” and “late” mutants were able to induce cellular DNA synthesis (Eckhart 1971A). Experiments to determine the cause of the inhibition of cellular transformation and viral DNA synthesis found that at nonpermissive temperatures, “early” mutants were defective in T antigen expression but “late” mutants were not. In addition, treatment of infected cells with DNA replication inhibitors did not affect T antigen expression in either wild type or mutants at permissive temperatures (Oxman 1972). Late mutants such as ts-59 were shown to express altered versions of all three viral coat proteins at nonpermissive temperatures that failed to form stable virus particles but could be complemented by the functioning coat proteins from early mutants such as ts-a (Gibson 1977). Cells infected with temperature-sensitive mutants defective in their ability to transform cells at nonpermissive temperatures did not revert to an untransformed
phenotype when switched to a nonpermissive temperature (Dimayorca 1969). This indicated that the disrupted viral genes are necessary for the initiation, but not maintenance, of the transformed phenotype.

Another temperature sensitive mutant, ts-3, was used to show that certain hallmarks of a transformed phenotype could be rendered temperature dependent. At nonpermissive temperatures, cells infected with ts-3 did not demonstrate a loss of topoinhibition, the halting of DNA replication when cells become too dense (Dulbecco 1970). In addition, ts-3 renders the reduction in extracellular proteins that normally occurs during transformation, temperature dependent (Eckhart 1971B). The ts-3 mutant was also shown to have different sensitivities in different host cell types, failed to complement with other mutants, and decreased temperature sensitivity when fresh medium and serum was added at a nonpermissive temperature (Eckhart 1974). When RNA from cells infected with ts-3 was annealed to radiolabeled viral early DNA, it failed to accumulate at nonpermissive temperatures indicating the temperature sensitivity affected early transcription (Cogen 1977). These observations were likely the result of the temperature sensitivity affecting the ability of the ts-3 virus to uncoat and release viral DNA at nonpermissive temperatures. This explains the lack of transcription as well as the lack of transformation hallmarks.

Taken together these experiments with temperature sensitive mutants demonstrate the primary functions of viral genes are to transform host cells, initiate viral DNA replication, and produce viral coat proteins.
ii. Host range nontransforming mutants

Early investigation into polyoma gene function required a way to culture mutant virus in order to observe loss of function. Unfortunately, most viral mutants could not be cultured directly since so much of the viral genome is critical for the viral life cycle. One strategy to overcome the problem was the use of temperature sensitive mutants. Another approach was to use what are called host range nontransforming mutants or “hr-t mutants”. Hr-t mutants are unable to grow in normal mouse cells but they can grow in mouse cells that have been previously transformed. When infected with polyoma, mouse 3T3 cells can undergo both lytic infection and abortive transformation in which some of the cells do not produce virus particles but instead integrate viral genes in the host genome (Sambrook 1968). Transformed 3T3 cells that do not produce virus particles can be cultured and used as hosts for virus mutants defective in their ability to transform cells.

In order to isolate transformed virus-free cells, mouse 3T3 cells were infected with polyoma virus and plated in soft agar suspension to select for transformed cells. Antiviral serum and receptor destroying enzyme was added to prevent reinfection of the transformed virus-free cells. This resulted in 9 clones that could be recultured in antiviral conditions. Of these only one clone, 3T3 Py3, was able to be both infected and maintained indefinitely. Polyoma virus was mutagenized with either nitrosoguanidine (NG) or hydroxylamine (HA) and cultured on both 3T3 and 3T3Py3 cells. Out of the 900 NG and 360 HA plaques tested only 3 NG (NG-18, NG-23, NG-59) and 1 HA (HA-33) isolates showed strong plaque formation on 3T3 Py3 with no plaque formation on 3T3.
These mutants were called host range nontransforming mutants because of their altered host range caused by an inability to transform cells.

Host range nontransforming mutants have proven invaluable for discovering viral gene functions as well as corroborating with temperature-sensitive mutant experiments. The hr-t mutants, like the temperature sensitive mutant ts-3, have also shown an inability to demonstrate reduced extracellular protein expression normally associated with transformation despite successfully entering the host cell and expressing viral genes (Benjamin and Burger 1970). An interesting observation from experiments with the hr-t mutant NG-18 was its ability to grow in a variety of mouse cell types including cells transformed by murine leukemia virus as well as cells in a less differentiated state such as mouse embryonic fibroblasts. This meant that host cells did not need an integrated polyoma gene or even all of the traits associated with a transformed phenotype to be able to support hr-t growth. The transformed cells that were not good hosts for hr-t mutants included cells transformed by SV40 or spontaneously through mutation (Goldman 1975). Because of hr-t mutants’ inability to transform cells, they also could not transform rat or hamster fibroblasts (Siegler 1975). These observations are consistent with the current understanding that the middle and small t antigens were inactivated in hr-t mutants (Hattori 1979). Middle and small t antigens are responsible for the transformed phenotype and are necessary for the expression of the host factors required for viral DNA synthesis. Both embryonic fibroblasts and cells transformed by mouse leukemia virus would likely have had such factors expressed at a sufficient level to compensate for loss of middle and small t. Interestingly, despite being unable to transform host cells, hr-t mutants were still
able to express Large T antigen and synthesize viral DNA in nonpermissive 3T3 cells (Staneloni 1977).

At the time it was not understood how mutated viral genes prevented transformation or where in the viral genome they were located. Using a technique that replaces regions of hr-t virus with wild type sequence via restriction digestion, the disrupted region of hr-t mutant NG-18 was rescued when replaced with wildtype sequence near the promoter of the transcript expressed at early times. This same approach also revealed the general locations of other temperature sensitive mutants such as ts25D (a type of ts-a mutant), ts-3, and ts-1260 as mapping to the distal early region, proximal late region, and distal late regions respectively (Feunteun 1976). It would later be understood that the proximal early region mutated in hr-t mutants like NG-18 affected the splicing of middle and small t but not Large T antigens (Hattori 1979).

Another process affected by hr-t mutation is the status of the viral chromatin. Previous observations of wild type polyoma have shown that viral chromosomes are highly acetylated on histones H3 and H4 (Frearson 1972, Schaffhausen 1976). This acetylation was not observed in hr-t mutants indicating the necessity of host transformation for viral chromosome packaging (Schaffhausen 1976). Results from these experiments have been replicated in a number of other hr-t mutants leading to a model of hr-t genes inducing host permissivity factors required for transformation, virus growth, and acetylation of host histones incorporated into viral chromosomes. It was later determined that small t antigen specifically is responsible for the hyperacetylation of viral chromatin in a manner dependent on its association with protein phosphatase 2A (Dahl 2007).
While the mechanism of culturing hr-t mutants and temperature sensitive mutants is different, both kinds of mutants have the ability to complement each other when infecting the same cell. Studies of both hr-t and ts-a mutants showed complementation for transformation and growth at high ts-a to hr-t ratios. Complementation between both types of mutants is reciprocal in both untransformed mouse 3T3, hamster BHK, and rat Y1 cells and at nonpermissive temperatures (Eckhart 1977). Unlike hr-t mutants, ts-a mutants were unable to grow in cells transformed by murine leukemia virus. These experiments indicated that the mutants were defective in different viral early gene functions, suggesting multiple functions for viral early genes (Fluck 1977). Both wild type and temperature sensitive mutants at nonpermissive temperatures were shown to induce DNA replication in the host cell for multiple rounds while hr-t mutants only induced a single round (Schlegel 1978). Comparisons between polyoma and SV40 versions of temperature sensitive and hr-t deletion mutants showed that the temperature sensitive mutants lose the ability to maintain transformation but not the ability to initiate it while polyoma hr-t and SV40 hr-t deletion (dl) mutants are largely unable to initiate transformation (Fluck 1979). Sequence analysis of the hr-t region of a variety of hr-t mutants revealed the C terminal region of smaller early gene products, but not the large T antigen, were not expressed (Hattori 1979). Both kinds of mutants can now be understood with regards to the viral genes being disrupted. Loss of Large T antigen precludes viral DNA synthesis but not cellular transformation. Loss of middle and small t antigens precludes cellular transformation. Loss of viral coat proteins either precludes completion of a lytic infection, or in the case of ts-3, results in a loss of the ability of virus particles to properly uncoat and transcribe viral DNA.
e. Transcription

Study of the transcription of polyoma RNA at different stages of viral infection was made difficult by the variability in the onset of virus induced DNA synthesis at the individual cell level. For this reason the DNA replication inhibitor 5-fluorodeoxyuridine (FdU) was used to synchronize polyoma-infected cells. This allowed detection of polyoma RNA via hybridization to viral DNA of both low abundance “early” RNA and extremely high abundance “late” RNA expressed before and after DNA replication respectively (Hudson 1970). In the presence of DNA replication inhibitor FdU, transcription of early RNA was followed by the expression of tumor antigen leading to the cell expressing mitogenic factors and increasing the expression of ribosomal RNA (Salomon 1977). This was consistent with a role of early gene products in driving DNA replication.

Analysis of cytoplasmic polyoma RNA from late times in infection resolved 3 species of RNA, two of which hybridized to the late region and one that covered most of the early region (Kamen 1976, Turler 1976). In vitro translation of the cytoplasmic long and short late region RNAs revealed they code for the capsid proteins VP2 and VP1 respectively (Smith 1976, Kamen 1978). At early times in infection both nuclear and cytoplasmic RNA was shown to be complementary to 40% of the early strand of the viral genome with a small amount complementary to the late strand in the nucleus (Acheson and Mieville 1978). At late times nuclear RNA was shown to be complementary to all of the late strand and 60% of the early strand while cytoplasmic RNA was only complementary to 45% of the early strand and 55% of the late strand. Taken together
this indicated that the smaller cytoplasmic gene products were likely derivatives of the larger nuclear RNAs (Beard 1976). This is remarkable because it was a good example of alternative splicing before splicing had been elucidated. The small size of the viral genome coupled with the difference in the percentage of the viral genome hybridized to nuclear and cytoplasmic RNA, indicated that some process was changing the length of cytoplasmic RNA.

Experiments looking at the viral transcripts produced at late times by electrophoresis and sucrose gradient centrifugation revealed surprisingly large RNA multiple times longer than the polyoma genome that are processed into smaller species (Acheson 1971). Radiolabeled RNA from late times in infection was found to be highly resistant to RNase treatment due to self-annealing. This indicated symmetrical transcription of both strands was occurring across the whole genome at late times (Aloni 1973). Isolation of polyoma RNA from late times in infection using oligo(dT) cellulose chromatography showed approximately 15% of giant transcripts were polyadenylated. When treated with the DNA replication inhibitor FdU, late polyoma transcripts no longer exceeded the length of the polyoma genome indicating a dependence of giant transcript production on DNA replication (Rosenthal 1976). The presence of giant nuclear viral late transcripts with inefficient polyadenylation at late times of infection was confirmed using Southern mini-blot hybridization (Birg 1977) and in vitro transcription (Condit 1977). The polyadenylation of giant transcripts was found to be consistently on the 3’ end of the late region indicating that it was not terminating at various locations (Lev 1979). Electron microscopy of giant RNAs hybridized to restriction-digested polyoma
DNA containing 91% of the genome showed that the giant transcripts were contiguous tandem repeats of the whole genome and not the result of splicing (Acheson 1978).

Analysis of the 5’ ends of late polyoma mRNAs via Ribonuclease T1 fingerprinting revealed a variable number of copies of a “late leader sequence” resulting from splicing out whole genome sized introns from the giant transcripts (Legon 1979B, Treisman 1981A). The observed difference in the rate of RNA export from the nucleus to the cytoplasm between RNA at early times and RNA at late times is consistent with giant late RNA precursors undergoing more processing. At early times radiolabeled uridine was seen in the cytoplasmic fractions <6 minutes after it was seen in nuclear fractions, while at late times the difference was between 12 to 20 minutes (Acheson 1981). This makes sense if there are longer RNAs to process at late times than at early. The tandem repeats of late leader sequences were visualized by electron microscopy of the loops cause by hybridizing late mRNA to genomic sequence (Zuckermann 1980).

Complete digestion of viral mRNA with RNase T1, T2, and A, followed by 2D electrophoresis revealed a variety of 5’ cap structures (Flavell 1979). When the cap structures were removed and the 5’ ends of late mRNAs were sequenced, they were found to have variable sequences upstream of the late leader (Flavell 1980, Cowie 1981). The location of the 5’ cap with respect to the viral sequence was found to be dependent on the ratio of available nucleotides when transcribed in vitro indicating variability in transcription start sites (Cowie 1982). Comparison between early and late transcription in vitro in HeLa whole cell extracts showed early transcripts to be consistently expressed but late transcripts showed lower expression than in vivo and the variability in late 5’ capping was not observed. Late transcription was also affected by the concentration of
template *in vitro* (Jat 1982A). Mapping of the 5’ ends of early strand mRNAs via hybridization to a HinfI restriction fragment spanning the noncoding control region followed by S1 nuclease digestion of unhybridized strands and electrophoresis showed the majority of early mRNAs are principally capped between nucleotides 147 and 152 with minor species found to be capped in the early coding region. At late times species were found to have 5’ capping occurring further upstream before the origin of replication consistent with either a shift in transcription start sites or contamination from giant transcripts (Kamen 1982)(Fig. 2).

Hybridization of the three cytoplasmic late mRNAs to the viral genome protected sequences from digestion by S1 nuclease. This allowed for the precise splice sites of VP1, VP2, and VP3 to be determined by comparing hybridization with a variety of restriction fragments to the known sequence of the late region (Kamen 1980).

**f. Noncoding control region**

The area of the polyoma genome containing the origin of replication, enhancers, the early and late promoters, and large T antigen binding sites is called the noncoding control region (Fig. 3). Much of what is understood about this region has come from deletion experiments to determine which areas affect transcription efficiency or DNA replication. Comparisons between the sequence of wild type polyoma and mutants capable of replicating in embryonic carcinoma cells also revealed key regulatory sequences.

The origin of replication consists of an A-T tract followed by a GC rich inverted repeat that is necessary for DNA replication. GC to AT base substitutions in the GC rich
Figure 2. Shift in the transcription start site of the early strand late in infection.

The noncoding control region is shown in black for both early and late times in infection. Early strand transcripts are shown in blue, late are shown in red. The early strand transcribes from the TATA box to the right of the origin at early times. The late strand transcribes from the enhancer A region. At late times the early strand transcription shifts back towards the enhancer A region.
Figure 3. Enhancers and host transcripton factor binding sites in the polyoma noncoding control region

The 4 enhancers of the polyoma noncoding control region A-D are shown on the late side of the origin. The major host transcription factors that bind to the enhancers are also shown. The origin, Large T binding sites, and early and late transcription start sites are shown with the A enhancer also functioning as a late transcription initiation site.

(This image was modified from Martelli J Virol 1996)
region were found to disrupt DNA replication when they were closer to the center of the inverted repeat (Triezenberg 1984). The introduction of deletion mutations in the noncoding control region between the origin and the late coding sequence demonstrated the existence of an enhancer element required for efficient expression of early transcripts (Tyndall 1981). When the polyoma transcriptional enhancer was inserted into a plasmid containing the rabbit hemoglobin beta-1 gene, it strongly enhanced expression of rabbit transcripts over a distance of 1400 bases (de Villiers 1981). Using the same beta-globin plasmid based approach, enhancers from polyoma and SV40 were both tested in mouse and primate cells and demonstrated a cell type preference (de Villiers 1982).

The transcriptional enhancer region was also found to be necessary for DNA replication. Deletion mutants were not only shown to have a strong negative effect on DNA replication from loss of the A-T tract but also when a section containing the transcriptional enhancer starting 60 bases into the late region and stopping in the middle of the late coding region was removed. The inability of wild-type virus to rescue these mutants demonstrated that the missing sequences function in cis (Luthman 1982). Plasmids containing a 251 base segment of the polyoma non-coding control region were expressed in MOP cells that express polyoma large T antigen to make them permissive for polyoma DNA replication. Mutation analysis revealed three regions necessary for polyoma replication, the core consisting of an A/T tract followed by a GC rich palindromic sequence, a sequence upstream of the core in the enhancer region called B that has no apparent features, and a sequence further toward the late region called A that can substitute for B’s role in DNA replication (Muller 1983). When the polyoma enhancer region was removed the virus lost the ability to replicate. Addition of an SV40
or immunoglobulin enhancer both restored DNA replication in cis and altered cell type specificity (de Villiers 1984). Further analysis of the enhancer region found redundant enhancer elements also required for DNA replication. When the region was deleted and copies of a 26 base element were introduced only two copies were required to restore DNA replication ability irrespective of orientation. Two additional enhancer domains, C and D were also identified (Veldman 1985).

Embryonal carcinoma (EC) cells were used to study regulation of transcription and DNA replication in viral genes because wild-type virus was shown to grow poorly in this cell type but could be compared to mutant virus that can express early and late genes in EC cells. EC cells express lower levels of the c-Fos/c-Jun and ets family of transcription factors required for viral transcription. The duplications in the recognition sites for these proteins on the viral genome allowed the mutants to compensate for low expression levels (Nothias 1993). These mutants were first isolated from virus that grew on EC cells differentiated into fibroblasts and were selected based on their ability to grow in undifferentiated EC cells. The mutants showed the same functionality and host range as the wild type and differed only in a 10-50 bp duplication of sequence on the late side of the origin in the noncoding control region (Katinka 1980, Katinka 1981, Vasseur 1980, Dandolo 1983). Comparison between the DNase I hypersensitive regions of wild type and EC mutant noncoding control regions showed the same DNase I profile of two hypersensitive regions on the late side of the origin despite different sequence, due to a GC rich palindrome common to both wild type and mutant that provides protection between the two regions (Herbomel 1981). Each of the hypersensitive regions was found adjacent to an enhancer element in both wild-type and mutant virus (Herbomel 1984).
When wild-type and mutant virus were both grown in embryonic carcinoma cells, the mutants failed to rescue wild-type DNA replication (as indicated by differing restriction fragment size length) suggesting the differing sequence in the noncoding control region of the mutants was acting in *cis* (Fujimura 1982).

The polyoma A and B enhancers were found to function as binding sites for host factors. When the B enhancer from polyoma was incubated with proteins from the nuclei of 3T6 cells, at least two factors called polyoma enhancer B binding proteins 1 and 2 (PEB1 and PEB2) were found to cause band shifting and protect a GC rich palindromic region homologous to immunoglobulin and SV40 enhancers from Dnase I digestion (Piette 1985). Similar attempts to reveal the factors binding to the A enhancer found two proteins called polyoma enhancer A binding protein 1 and 2 (PEA1 and PEA2). PEA1 was also found to bind to SV40 and c-fos enhancers suggesting it is the same protein as the human activator protein 1 (AP-1) (Piette 1987). A third type of protein was found to bind to polyoma enhancer A called PEA3 that cooperates with PEA1 in response to upstream oncogene expression but unlike PEA1 is not dependent on c-fos (Martin 1988, Wasylyk 1989). PEA3 was found to be the same as the transcription factors produced by the *ets*-1 and *ets*-2 proto-oncogenes (Wasylyk 1990). Interestingly, in addition to functioning as an enhancer for early transcription and DNA replication, the PEA1 and PEA3 binding sites also facilitate the initiation of late transcription in lieu of a TATA binding by the TFIID complex (Yoo 1991). Another protein, called Enhancer Factor 1A (EF-1A), previously shown to bind to the E1A enhancer region in adenovirus, was identified through enhancer fragment competition to bind to 1 site in the A enhancer domain and 2 sites in the B enhancer domain. This interaction was shown to be sufficient
to stimulate some replication of polyoma in vivo (Bolwig 1991). The D enhancer domain was shown to bind to a common host transcription factor NF-D, which is a member of the NF-1 family (Caruso 1990). This factor was later shown to be the same as YY1, which can positively activate transcription of the late promoter (Martelli 1996). Interestingly, YY1 also directly binds to the late protein VP1, suggesting a possible role for VP1 in recruiting transcription factors in the upregulation of the late promoter (Palkova 2000).

The proteins PEA2, PEA3, and NF-D have different effects on the early and late promoters under nonreplicating conditions. All three factors activate early promoter transcription in a combinatorial manner while PEA2 and 3 repress late promoter transcription in a manner dependent on direct interaction between the two proteins (Shivakumar 1998).

The activity of enhancer binding proteins PEA1 and PEA3 with respect to viral transcription and replication has also been shown to be significantly influenced by middle and small t antigens. The host range nontransforming mutants (hr-t) of polyoma fail to express middle and small t antigens but compensate for the lack of downstream effects of these proteins with a duplication of the A enhancer to rescue some DNA replication ability. When hr-t mutants were treated with the phorbol ester 12-O-tetradecanoylphorbol-13-acetate (TPA) to stimulate the pathways leading to an increase in the expression of PEA1 and PEA3, more of the DNA replication defect was rescued. When middle and small t were expressed in the presence of hr-t mutants with and without the duplication of the A enhancer the differences in DNA replication between them were abolished and replication efficiency was restored (Chen 1995). In addition to being necessary for viral DNA replication, middle and small t antigens are also necessary for
increasing late transcription at late times. When untreated hr-t mutants were compared with wild type virus treated with DNA replication inhibitor cytosine β-Arabinofuranoside (AraC) after entering the late phase, the wild type still showed higher levels of late transcription even with fewer genomes indicating that the reduction in late transcripts is not an indirect result of fewer templates from inefficient replication but from a direct reduction in late transcription (Chen 2001). The major pathways activated by middle T (Shc, PI3K, and PLCγ), along with small t antigen each additively contribute to the upregulation of PEA1 and PEA3 resulting in an increase in both viral DNA replication as well as late transcription (Chen 2006).

On the early side, the deletion of bases 200-400 base pairs before the 5’ cap sites resulted in a loss of transforming ability but still produced RNA *in vitro*. The TATA box, the 5’ cap site, and 90-100 bases of nonspecific spacer sequence are required for efficient *in vitro* transcription of the early region to occur (Jat 1982B). Deletion mutations on the early side from bases 20 to 86 abolished DNA replication in COP5 cells that express large T antigen with smaller deletions resulting in proportionally decreased DNA synthesis (Katinka 1983).

These experiments allowed for a mapping of the major functional sequences in the noncoding control region. The origin of replication, the early Large T antigen binding sites, the TATA box on the early side but not the late, and the enhancers of transcription and replication make up the majority of the sequence. Understanding of the factors that bind to the noncoding control region is important for both identifying causes of host cell specificity as well as demonstrating downstream effects of early proteins such as Large middle and small t antigens and the transcription factors they activate.
g. Tumor antigens

The early gene products of polyoma virus are the large, middle, and small tumor antigens that are necessary for transformation and DNA replication. Initial attempts to identify the tumor antigens by *in vitro* translation produced three proteins of ~90kDa, 60kDa, and 22kDa (Fig. 4). Unspliced early viral cRNA was unable to translate into Large or middle T but could translate into small t, indicating the presence of an intron. The sequences downstream of the splice site were found to be on different reading frames for each T antigen (Hunter 1978). Upstream of the splice site all of the tumor antigens were found to share a common N terminal sequence (Hutchinson 1978, Smart 1978) containing a highly conserved J domain that binds to the 70kDa family of host heat shock proteins (Cheetham 1992, Kelley 1994). Large, Middle, and small T antigens were expressed individually in Py-SV40 recombinants and shown by immunofluorescence to localize to the nucleus, cell membrane, and both nucleus and cytoplasm respectively (Zhu 1984).
All three of the tumor antigens share a J domain in their common N terminal regions but differ in their C terminal regions by each splicing to a different reading frame. Large T antigen has a nuclear localization signal and pRb-binding site in the region that is spliced out of middle and small t. Middle T has a hydrophobic region near the C terminus for cell membrane localization and a binding site for PP2A at the N terminus that it shares with small t. Middle T has unique binding sites for pp60c-src and can bind Shc, PI3K, and PLC-gamma1 when tyrosines 250, 315, and 322 are phosphorylated respectively.
i. Large T antigen

When the polyoma Large T antigen was expressed alone in fibroblasts from mouse and rat embryos, the cells demonstrated an “immortalized” phenotype in which they would grow indefinitely in lower serum concentrations but maintained an untransformed appearance. This immortalization also occurred when only 40% of the Large T antigen from the N terminal region was expressed indicating this region contained the necessary domain (Asselin 1985, Holman 1994, Rassoulzadegan 1983). This process was found to be dependent on interaction between Large T antigen and retinoblasoma family of tumor suppressor proteins. When the region of Large T necessary for binding to retinoblastoma proteins was mutated, a loss of immortalization of rat embryo fibroblasts was observed (Larose 1991). Viral genomes mutated to preclude binding of Large T to retinoblastoma proteins while expressing wild type middle and small t antigens still retained the ability to produce a transformed phenotype and tumors (Freund 1992). *In vitro* coimmunoprecipitation showed polyoma Large T antigen forms a complex with both human and mouse retinoblastoma tumor suppressor proteins pRB and p107 (Dyson 1990, Pilon 1996). The pRB binding domain of Large T fused with GST was immobilized on glutathione-sepharose beads and shown to bind to *in vitro* translated C terminal domains of pRB and p107, as well as an additional p130 retinoblastoma protein (Desjardins 1997).

The downstream effects resulting from the interaction between Large T antigen and pRb were found in some cases to be dependent on the conserved DnaJ domain common to the N terminal regions of all T antigens in addition to the pRb binding region. The DnaJ domain binds to DnaK proteins such as host heat shock chaperone Hsc70 and
stimulates ATP hydrolysis (Cheetham 1992). When Large T antigen was mutated in its pRb binding region or its DnaJ domain, the two mutants were unable to complement each other for downstream effects related to E2F promoter activation, suggesting a role for direct Hsc70 binding to facilitate Large T interactions with pRb (Sheng 1997). Despite this, Large T mutated in J domain binding to Hsc70 still demonstrated the ability to induce apoptosis in differentiating myoblasts, activate the cyclin A promoter, and alter phosphorylation of pRb and p130 (Sheng 2000).

Unlike SV40 Large T antigen, which was found to bind to the p53 tumor suppressor protein (Lane 1979, Linzer 1979), polyoma Large T antigen does not bind directly (Wang 1989). Expression of polyoma Large T antigen in mouse embryo fibroblasts with a temperature sensitive p53 produced the same loss of downstream effects of p53 (such as expression of p21 and hypophosphorylation of Retinoblastoma proteins) as nonpermissive temperatures. This loss of downstream effects was not observed when Large T mutated in its ability to bind retinoblastoma was expressed, indicating a dependence on the binding of Large T to pRb to overcome the effects of p53 and prevent apoptosis or inhibition of DNA replication (Doherty 1997). In essence, the expression of functional Large T antigen in these cells had the same effect on downstream targets of p53 as temperature sensitive p53 at nonpermissive temperatures.

Immunopurified Large T antigen was found to bind to three regions of the viral genome between the origin of replication and the early coding sequence. Each of the binding regions contain repeats of the sequence G(A/G)GGC (Cowie 1984, Dilworth 1984). Large T mutants unable to bind to the polyoma genome still retained the ability to immortalize rat embryo fibroblasts (Cowie 1986). Like SV40 Large T, polyoma Large T
was also found to function as a helicase to unwind double stranded DNA. Unlike SV40 Large T, polyoma Large T is able to unwind DNA of sequence unrelated to the polyoma origin of replication in vitro if the ratio of Large T to DNA is 100:1 (Wang 1991). The C terminal region of Large T antigen consisting of residues 264-785 was found to be sufficient for DNA replication but required complementation with the N terminal region to immortalize cells and replicate under low serum conditions (Gjorup 1994).

Using coimmunoprecipitation and GST pulldown, polyoma Large T antigen was found to bind to a subspecies of p300 with histone acetyltransferase activity (Nemethova 1999), which along with CBP is known to regulate genes that respond to cyclic AMP (Arany 1994, Arany 1995, Eckner 1994, Lundblad 1995). This association was found through a region of Large T with a high level of homology to SV40 Large T and Adenovirus E1A proteins previously shown to bind to p300/CBP (Arany 1995, Eckner 1996). Polyoma Large T was also found to bind to CBP at the N terminal Domain (1-771) and one of the middle domains (1069-1892) and p300 at the N terminal (1-596) and C terminal (1572-2370) domains in a manner independent of p53, NFkB, or ability to bind to pRb (Cho 2001). This demonstrates an additional way Large T can stimulate cell cycle progression besides binding directly to retinoblastoma proteins.

Polyoma mutants able to grow in cancer cells but not in normal cells expressed Large T with a C terminal region deficient in binding to the DNA replication inhibitor p150Sal2 (Li 2001). p150Sal2 is a tumor suppressor gene that binds to the promoters of the p53-regulated, p21 and BAX tumor suppressor genes and stimulates their expression independently of p53 to inhibit DNA replication and cell division (Li 2004). The gene products of the SALL2 gene including p150Sal2 are involved in driving serum-starved
cells into a quiescent state (Liu 2007). This suggests the binding of Large T antigen to p150⁰Sal₂ in addition to binding pRb is related to the ability of Large T to drive replication even in low serum conditions.

These interactions demonstrate a variety of roles for Large T antigen in the viral lifecycle of interacting with both the viral genome and cellular factors involved in cell cycle progression. Large T can immortalize cells and inhibit downstream effects of p53 by binding to the retinoblasoma family of proteins, prevent serum starved quiescence, and bind to the viral genome to function as a helicase in viral DNA replication.

ii. Middle T antigen

The polyoma middle T antigen is the primary driver of the transformed phenotype in host cells (Ito 1980). When virus expressing only functional middle T was expressed in rat cells, a transformed phenotype was still observed, indicating that middle T is necessary and sufficient for cellular transformation (Treisman 1981B). Middle T contains a hydrophobic C terminus that anchors it to the cell membrane (Ito 1977, Ito 1979, Silver 1978). When middle T was truncated at the C terminus and expressed with wild-type Large and small t antigens, it no longer localized to the cell membrane and the transformed phenotype was no longer observed (Carmichael 1982). Middle T is associated with the cell membrane on the inner side and does not possess ATP-binding activity (Schaffhausen 1982).

Immunoprecipitates of polyoma T antigens demonstrated the ability to phosphorylate middle T on tyrosine residues (Eckhart 1979). Through a series of deletion mutations, middle T was found to be phosphorylated on a tyrosine at residue 315
(Schaffhausen 1981). This activity was the result of an association between middle T and the cellular kinase pp60^{c-src} (Courtneidge 1983, Bolen 1984, Balmer-Hofer 1985), pp62^{c-yes} (Kornbluth 1987), or pp59^{c-fyn} (Horak 1989). When the tyrosine at residue 315 was mutated to phenylalanine to preclude phosphorylation, middle T showed a dramatic reduction in ability to induce a transformed phenotype (Carmichael 1984). The extent of transformation of rat F-111 cells by inducibly expressed middle T was found to correlate more with the level of phosphorylated middle T rather than the total amount of middle T expressed (Raptis 1985). Analysis of middle T peptides digested with trypsin and phosphorylated in vitro with radiolabeled phosphates revealed a second major phosphorylation site at tyrosine 250 (Hunter 1984). Antisera against the region of middle T containing tyrosine 315 still showed phosphorylation at tyrosine 250, while antisera against middle T in general showed additional phosphorylation at tyrosine 322 in vitro (Harvey 1984). Unlike phosphorylation of 315, phosphorylation of tyrosine 322 was found to be dispensable for transformation (Magnusson 1981, Schaffhausen 1985).

Immunoprecipitates made with antisera to Middle T antigen not defective in transformation were shown to have a phosphatidylinositol kinase activity in vitro and in vivo, while middle T mutants defective in transformation did not (Whitman 1985, Kaplan 1986). This demonstrated a dependence of middle T transformation on phosphatidylinositol metabolism. When 3T3 cells were either treated with platelet derived growth factor (PDGF) or transformed with polyoma middle T, to stimulate the phosphorylation of proteins on tyrosine, anti-phosphotyrosine immunoprecipitates included an 81-85 kDa phosphoprotein that coincided with high phosphatidylinositol kinase activity (Kaplan 1987). The protein was found to complex with middle T and
pp60^c-src (Courtneidge 1987) in a manner dependent on the phosphorylation of tyrosine 315 on middle T (Talmage 1989). This protein was called phosphatidylinositol-3-kinase (PI3K) for its ability to phosphorylate phosphatidylinositol at the D-3 position (Whitman 1988). Polyoma transformed 3T3 cells show PI3K activity in the form of elevated phosphatidylinositol-3-phosphate, phosphatidylinositol-3,4-bisphosphate, and phosphatidylinositol trisphosphate that is not observed in uninfected cells (Serunian 1990).

In addition to the 85 kDa PI3K protein associated with middle T, a 61-63 kDa and a 36-37 kDa protein were found in complex with polyoma middle T, polyoma small t, or SV40 small t antigens (Pallas 1988). These proteins were identified as the A (regulatory) and C (catalytic) subunits of the serine and threonine Protein Phosphatase 2A (PP2A) respectively (Pallas 1990, Walter 1990). PP2A associates with both middle and small T antigens through their common N terminal regions, upstream of the splice site, that mimic the cellular B subunit. Biochemical analysis revealed that the association of middle T with PP2A is necessary for binding of pp60^c-src and related proteins (Campbell 1995). Inhibition of the phosphatase activity of PP2A with okadaic acid reduced PP2A association with middle T as well as the amount of pp60^c-src bound to middle T (Glover 1999). Interestingly, mutations causing inactivation of the catalytic activity of the C subunit of PP2A precluded interaction with the cellular regulatory B subunit but did not preclude association with middle T or recruitment of pp60^c-src (Ogris 1999). This suggested that the reduction in association of PP2A with middle T in the presence of okadaic acid was not purely the result of a loss of catalytic activity.
The Src homology 2 containing protein (Shc) was found through immunoprecipitation experiments to bind to middle T antigen in a manner dependent on phosphorylation of middle T at Tyrosine 250 at the NPTY motif. This association results in the phosphorylation of Shc by pp60<sup>c-src</sup>, leading to the binding of Shc by Grb2, mimicking the signaling cascade of a tyrosine kinase associated growth factor receptor (Campbell 1994, Dilworth 1994). Shc consists of three isoforms of 46, 52, and 66 kDa (Pelicci 1992). Shc is phosphorylated on tyrosines 239/240 as well as on tyrosine 313 (Salcini 1994, Gotoh 1996). Grb2 binds to Shc at either of the phosphorylated tyrosine sites and associates with Sos1 or Gab1 (Nicholson 2001). Sos1 is a guanine nucleotide exchange factor that converts inactive Ras-GDP to active Ras-GTP resulting in the activation of Raf and the MAP Kinase signaling pathway. Gab1 is involved in the localization of Sos1 to the cell membrane where it can interact with Ras-GDP (Yamasaki 2003).

Investigation into the role phosphorylation of middle T at tyrosine 322 plays revealed the amino acids immediately downstream of tyrosine 322 (Leu323-Asp324-Ile325) constituted a common binding target of the SH2 domain of Phospholipase C gamma 1 (PLC-γ1). Immunoblotting of middle T immunoprecipitates with anti-PLC-γ1 antibody showed the presence of PLC-γ1 in a manner dependent on the presence of tyrosine 322 (Su 1995). This association allows for PLC-γ1 to be phosphorylated by pp60<sup>c-src</sup>, which is necessary to activate PLC-γ1 function. PLC-γ1 catalyzes the hydrolysis of phosphatidylinositol 4,5-bisphosphate into inositol 1,4,5-trisphosphate (IP3) and diacylglycerol (DAG), each of which activates calcium release and Protein Kinase C (PKC) respectively (Majerus 1990).
iii. Small t antigen

The small tumor antigen is nearly identical to the N terminal region of middle T and differs only by 4 amino acids downstream of the splice site that terminate on a different reading frame. Like middle T, polyoma small t binds to the A and C subunits of protein phosphatase 2A (Pallas 1990, Walter 1990). Despite both containing the same PP2A binding sequence, small and middle T antigen produce different effects when binding PP2A. Small t antigen was shown to produce cell cycle progression by associating with PP2A. When middle T mutated in its ability to associate with PI3K and Shc was expressed, it was unable to promote cell cycle progression despite retaining a wild-type PP2A association (Mullane 1998). Expression of both small and Large T antigens results in the induction of S-phase and an increase in cyclin A and cyclin E-specific protein kinase activity. Small t antigen mutated in its ability to associate with PP2A was unable to prevent the protein p27Kip1 from inhibiting cyclin E-specific activity (Schuchner 1999). Small t is also involved in overcoming p53 dependent growth arrest and apoptosis. While Large T antigen does not bind p53 directly in polyomavirus-infected cells, small and middle T antigens have been shown to resist apoptosis in p53-induced cells through PP2A and PI3K respectively (Qian 2000). In mouse primary cells and rat embryonic fibroblasts, middle T expression is sufficient to activate the ARF-p53 pathway (Lomax 2001) but coexpression of small t with an intact PP2A binding site is sufficient to overcome the effects of p53 to produce a transformed phenotype (Moule 2004). Polyomavirus small t and a mutated form of middle T that does not localize to the cell membrane (and thus cannot interact with Src family kinases) both activate the MAP.
kinase cascade via PP2A in a manner independent of Shc-Grb-Sos activation of Ras (Rodriguez-Viciana 2006). The hyperacetylation of viral chromatin also depends on small t directing the host cell to S phase during viral chromosome synthesis and packaging (Dahl 2007).

B. Polyoma Virus Life Cycle

1. Virus Attachment and Entry

The life cycle of polyomavirus begins with the attachment of the VP1 viral coat protein to the sialic acid component of the ganglioside surface receptors GD1a and GT1b on the target host 3T3 cell membrane (Chen 1997, Tsai 2003). The attachment of VP1 to the cell initiates a signal transduction cascade resulting in an initial increase in the expression of genes like c-myc and c-fos to prime the cell for DNA replication (Zullo 1987). The virus particles enter the host cell through smooth monopinocytic vesicles and localize to early endosomes and the endoplasmic reticulum (Griffith 1984, Richterova 2001, Mannova 2003, Liebl 2006). Once it arrives in the ER, the hydrophobic C terminal domain of the capsid protein VP1 is exposed by the activity of ERp29, allowing the virus to cross the ER membrane into the cytoplasm or nucleus (Magnuson 2005). The chaperone protein Hsp70 is involved in the disassembly of the virus, removing all of the coat proteins except the VP1 proteins associated directly with the viral genome in a manner similar to histone H1 (Chromy 2006). The remaining VP1 proteins are necessary to direct the viral chromatin to the nucleus via their nuclear localization signals (Li P 2001). The protein PARP-1 removes the remaining VP1, revealing the viral enhancers
and promoters and allowing for the transcription of the early genes to begin (Carbone 2006).

2. Early Phase

Once transcription of the viral genome begins the virus expresses primary transcripts from both strands with efficient transcription termination at the overlapping polyadenylation sites (Hyde-DeRuyscher 1988, Gu 2009). This causes the transcripts on the late strand to destabilize because the late leader exon at the 5’ end of the late region requires at least one read-through of the entire viral genome to splice properly (Adami 1989, Hyde-DeRuyscher 1990). The early strand primary transcripts are alternatively spliced to make mRNAs for the Large T, middle T, and small t antigens, which localize to the nucleus, plasma membrane, and nucleus plus cytoplasm, respectively (Soeda 1980, Zhu 1984).

The N terminal region of the Large T antigen binds to the Retinoblastoma family of tumor suppressors including pRb, p107, and p130, causing the release of E2F transcription factors which control the expression of cell cycle genes (Dyson 1990, Pilon 1996, Desjardins 1997). This interaction is believed to be facilitated by the DnaJ domain that interacts with host DnaK chaperone proteins like Hsc70 (Cheetham 1992, Sheng 1997). This domain is common to all three T antigens but is primarily utilized by Large T. Large T also associates with the proteins p300/CBP that function as histone acetyltransferases and transcriptional coactivators that bind to cell cycle transcription factors such as the Cyclic AMP Response Element Binding protein (CREB) and facilitate the recruitment of the transcription complex to express cell cycle genes (Arany 1994,
Arany 1995, Eckner 1994, Lundblad 1995, Nemethova 1999). In addition to promoting cell cycle gene expression, Large T also binds to the tumor suppressor p150\textsuperscript{Sal2}, which is involved in inhibiting DNA replication by activating the transcription of downstream targets of p53 such as p21 and BAX (Li 2004).

Middle T antigen localizes to the cell membrane and functions as a constitutively active receptor tyrosine kinase to activate signaling cascades resulting in the host cell acquiring a transformed phenotype. The N terminal region of middle T binds to the A and C subunits of protein phosphatase 2A by mimicking the B subunit (Pallas 1990, Walter 1990). This association allows for middle T to be phosphorylated by pp60\textsuperscript{c-src} (Courtneidge 1983) at a number of locations including tyrosines 250, 315, and 322 (Harvey 1984, Hunter 1984, Schaffhausen 1981). Once these sites are phosphorylated they become functional binding sites for a number of major signaling proteins.

Phosphorylation of tyrosine 250 allows for the binding and phosphorylation of Shc leading to the activation of the MAP kinase cascade (Campbell 1994, Dilworth 1994). Phosphorylation of tyrosine 315 allows for the binding of the 85kDa subunit of phosphatidylinositol-3-kinase (Talmage 1989). This results in phosphorylation of phosphatidylinositol leading to activation of PDK1 and subsequently, Akt resulting in survival and proliferation (Meili 1998). The phosphorylation of tyrosine 322 allows for the binding and phosphorylation of PLC-gamma, activating its catalytic function causing it to cleave PIP2 into IP3 and DAG leading to the release of calcium and the activation of Protein Kinase C and downstream oncogenes (Su 1995). It is the sustained combined activation of these pathways that give the cell its transformed phenotype.
Small t antigen functions, primarily, by binding to PP2A in a manner similar to middle T (Pallas 1990, Walter 1990). Unlike middle T, small t localizes to the cytoplasm and nucleus allowing for more direct association with PP2A and any targets not bound to the cell membrane. While the precise mechanism of action is unclear, the association of small t antigen with PP2A affects a number of cell processes such as promoting cell cycle progression (Mullane 1998, Rodriguez-Viciana 2006) and interfering with tumor suppressors such as p53 (Qian 2000, Moule 2004) or p27Kip1 (Schuchner 1999).

The combined activity of the three T antigens results in the host cell entering S phase and the activation of host transcription factors results in the expression of proteins required for the virus to replicate. The C terminal region of Large T binds directly to the viral genome at the noncoding control region between the origin of replication and the early transcription start site (Cowie 1984, Dilworth 1984). Large T antigen facilitates replication of the viral genome by recruiting host replication factors and functioning as a hexameric DNA helicase (Wang 1991).

3. Early-Late Switch

Once the host enters S phase and viral DNA replication occurs the virus undergoes an early-late switch where the expression of late-strand transcripts increases more than 50 fold relative to early-strand transcripts (Kamen 1975, Piper 1979, Hyde-DeRuyscher 1988). The onset of viral DNA replication correlates with an increase in A-to-I editing of viral RNA by a class of enzymes called Adenosine Deaminase Acting on RNA (ADARs) (Gu 2009). ADAR enzymes are abundant nuclear and cytoplasmic proteins that bind double-stranded RNAs with a preference for A-rich regions and
catalyze a deaminating reaction that alters adenosine bases to inosines (Bass 2002). The polyadenylation signals (AAUAAA) of polyoma early and late transcripts overlap completely resulting in A-rich regions on both strands when they hybridize. Whether hybridization occurs between transcripts from the same or different DNA molecules is unknown but appears to be dependent on viral DNA replication (Liu 1993). It is the editing of the polyadenylation signal that is hypothesized to result in a loss of polyadenylation efficiency leading to the giant transcripts observed at late times (Acheson 1971, Gu 2009).

While the 3’- and 5’-splice sites flanking the late-leader exon are strong, the 3’ splice sites for VP3 and VP1 are weaker. This makes leader-to-message body splicing relatively inefficient, thus favoring the stronger leader-to-leader splicing in multigenomic length precursors (Batt 1994). The loss of polyadenylation efficiency results in an approximately 50% chance a given transcription complex will terminate resulting in the observation of primary transcripts as large as 12-15 times the length of the viral genome (Acheson 1984, Hyde-DeRuyscher 1990). Splicing priority is given to the 3’ splice site of the late leader resulting in all of the full genome introns between late leaders being spliced out and only the last 5’ splice site splices to form VP1 or VP3, or retains the intron to form VP2. This process yields VP1, VP2, or VP3 mRNAs with a number of tandem repeats of the late leader sequence corresponding to the number of times the transcription complex read through the polyadenylation site before terminating (Fig. 5) (Hyde-DeRuyscher 1990). Interestingly, while the late leader exon itself contains sequences complementary to the 3’ end of mouse 18S ribosomal RNA (Soeda 1980) as well as binding sites for the 40S ribosomal subunit (Legon 1979A), the actual sequence
appears dispensable for virus viability and predominantly serves as a spacer (Adami 1986). When the sequence of the late leader was replaced with a different sequence lacking complementarity to the mouse 18S ribosomal RNA, it showed no difference in translational efficiency (Rhee 1989).
Figure 5. Leader-to-leader splicing

During the late phase of infection, the efficiency of polyadenylation decreases leading to the production of giant multigenome-long transcripts. The late mRNAs for VP1, VP2, and VP3 require the late leader exon to splice across the entire genome length introns at least once to produce stable mRNAs. The number of late leaders in a given mRNA corresponds to the number of times the polymerase went around the genome.
4. Late Phase

During the late phase of infection the host cell is rapidly dividing, even at low serum concentrations. After the onset of viral DNA replication late-strand transcripts accumulate rapidly and can constitute up to 2% of all the transcripts in the cell (Kamen 1975, Piper 1979, Hyde-DeRuyscher 1988). The late genes VP1, VP2, and VP3 are expressed at 50 to 100 times the level of early genes and the giant transcripts from both strands have the potential to hybridize, resulting in hyper-editing by ADAR across the entire viral transcriptome (Gu 2009). This likely interferes with the early transcripts more than the late transcripts because of the multiple genome long introns spliced out of late pre-mRNA that are not spliced from the early strand. These giant introns potentially reduce the chances of spliced late mRNA being edited compared to early transcripts.

VP1 localizes to the nucleus and assembles into viral capsids in a manner dependent on calcium activation (Salunke 1989, Montross 1991). This calcium dependence could be part of the reason defects in VP1 encapsidation were observed in middle T mutants (Garcea 1989). If the mutants were unable to properly interact with PLCγ and trigger a downstream calcium release, it could have affected VP1 assembly.

VP2 and VP3 require the presence of VP1 to localize to the nucleus (Delos 1993). VP1 and VP2 are able to associate with each other in the cytoplasm and utilize the nuclear localization signal of either protein (Cai 1994). All three late proteins are translated in the cytoplasm and associate with Hsc70 before localizing to the nucleus (Cripe 1995). This association is likely necessary to prevent premature capsid assembly in the cytoplasm. VP1 alone is sufficient to assemble ~5kb of DNA into capsids (Gillock 1997). VP1, but not VP2 or VP3, binds directly to DNA through the same N terminal
region that contains the nuclear localization signal in a manner independent of sequence
(Chang 1993). The encapsidation of DNA is dependent on the concentration of viral
genomes and not on any particular sequence (Spanielova 2014).

5. Host Cell Lysis

By 72 hours after the initial infection, 46% of the host cells are necrotic while
10% are apoptotic (An 2000). This is unsurprising because of the continued expression
of early gene products that interfere with a number of apoptotic pathways. The small
population of apoptotic cells is potentially the result of the minor viral coat proteins VP2
and VP3, which have been shown to induce apoptosis when transfected directly
(Huerfano 2010). Once the host cell lyses, virus is released and the process starts over
with attachment to the next host.
C. Thesis Objective and Specific Aims

The goal of this thesis is to describe global changes in transcription observed by high throughput illumina sequencing of the total RNA from a polyoma infection. Historically polyomavirus has proven invaluable for understanding host tumorigenesis through interaction of viral factors with mitogenic pathways. The virus has also allowed for understanding of host transcription and replication factors due to its small size. Despite this, a global view of transcription during polyoma infection has yet to be observed. By aligning sequencing reads from either different time points during infection or from infected cells treated with or without a DNA replication inhibitor, our lab can observe RNA from early and late phases of infection from both the virus and host. In addition, the use of new read sorting pipelines allows for more direct observation of hyper-editing predicted from the production of viral giant transcripts from both strands at late times that was once difficult to observe using conventional methods. By seeing the changes in viral and host transcription, we can discover new aspects of viral gene regulation as well as downstream host effects.

Specific Aim I: Investigate changes in transcription, splicing, and editing of the viral transcriptome at a global level

Traditional methods of investigating changes in viral RNA expression over the course of infection are costly, labor intensive, and have a limited scope. By using next generation sequencing to simultaneously observe changes in viral transcripts across the viral genome, it becomes possible to achieve a more global view of processes such as
transcription, splicing, and editing. The Illumina TruSeq Stranded sequencing system allows for the creation of strand specific libraries to be sequenced at a depth of over 50 million reads. These reads can then be aligned to the viral genome to illustrate the total RNA expression levels at both early and late phases of the viral life cycle. Furthermore, reads spanning splice junctions can be used to observe any changes in the relative abundance of viral isoforms on each strand. Finally, using a pipeline recently developed by Porath et al (2014), we can sort unaligned reads to find clusters of mismatches with a high probability of resulting from A-to-I editing that has been predicted for polyoma transcripts but only observed in a limited way.

**Specific Aim II: Investigate global changes in coding and noncoding host transcripts over the course of a mouse polyomavirus infection.**

Viral proteins, especially the T antigens, have been demonstrated to show a number of significant mitogenic effects on host gene expression. These mitogenic effects result in a number of transcription factors involved in cell cycle control being activated to transition the host cell into S phase and initiate DNA replication. The same sequencing approach used for viral transcripts can also be used to observe changes in all of the detectable host genes simultaneously. By aligning reads from total RNA taken from both infected and uninfected samples at different time points to the mouse genome, we can observe downstream changes in host gene transcription for both coding and noncoding RNA. Both upregulated and downregulated genes can be analyzed based on enrichment for common functions as well as predicted transcription factor binding sites to determine the types of categories being overrepresented. This allows for a much broader look at the
over all changes in the virus infection than could ever be achieved through traditional targeted methods.
II. MATERIALS AND METHODS

A. Cell Culture and Virus Infection

NIH 3T6 mouse fibroblast cells were cultured on 10 cm plates at a density of 600K cells/plate in Gibco Dulbecco’s Modified Eagle Medium (11995-065) supplemented with 100 units of penicillin/streptomycin, 2mM L-Glutamine, and 5% calf serum. Cells were serum starved in 2% calf serum 1 day before infection to synchronize cell cycles. Polyomavirus strain Py59RA was generously provided by Dr. Robert Garcia in 2% Medium at 180 million plaque forming units per ml. Virus was diluted to 15 million PFU/ml to infect cells at an MOI of 50 using 2 ml of virus per plate. Cells were infected by removing media and adding 2 ml of virus for 2 hours before being supplemented with 8ml of 2% media at time point 0. Mock infections were carried out using 2% medium without filtered virus particles. Total RNA was harvested at 12, 18, 24, and 36 hours after infection by removing media, washing with PBS, and adding 4M guanidinium isothiocyanate. For the aphidicolin treated samples, cells were plated at the same density, treated with the same media, and infected at the same MOI for 40 and 48 hours. Aphidicolin positive cells were treated with media containing 2µg/ml aphidicolin.

B. Preparation of TruSeq™ Stranded Library

RNA was isolated by cesium centrifugation through 2 ml of 5.7M cesium chloride at 39000 rpm at 20°C for 18 hours and DNAsate treated with DNAsate 1 from the Ambion DNAse I Kit. 1 µg of RNA was used from each sample for library preparation with the Trusequ Stranded Total RNA kit from Illumina. Ribosomal RNA was depleted using
rRNA binding beads and the RNA was fragmented and primed with random hexamers. First strand synthesis was done using a mixture containing Superscript II and Actinomycin D to synthesize RNA dependent DNA and inhibit DNA dependent RNA respectively. Second strand synthesis was done using dUTP instead of dTTP to create stranded cDNA. Libraries were then adenylated at the 3’ end and adaptors containing identifier sequences and flow-cell binding sequences were ligated to both ends. Finally the cDNA fragments were amplified for 12 cycles instead of 15. The Agilent Technologies 2200 Tapestation was used to verify library size at ~260bp. Sequencing was done on a HiSeq 2000 by Perkin Elmer.

C. Analysis

All alignments were made using Tophat v 2.0.9, Bowtie 2.1.0, and Samtools 1.19. Alignments to polyoma were done using the reference genome for the Py59RA strain of polyomavirus for both time course and aphidicolin experiments. Time course experiments consisted of four conditions, for mock infection and four conditions for polyoma infection representing 12, 18, 24, and 36-hour time points with three biological replicates per condition. Aphidicolin experiments consisted of two conditions Aphi+ and Aphi- with two biological replicates.

Alignment of timecourse reads to the host genome was done using the reference genome from Illumina’s IGENOMES collection, which in turn was taken from USCS March 5, 2013. The annotation file was modified to remove single-base exons due to difficulties tophat had in processing them. Custom scripts were used to postprocess
tophat output to produce strand-specific bedGraph files suitable for use on the UCSC genome browser.

Alignment of reads to the viral origin was done using a custom script that rotated the genome zero position counterclockwise by 149nt to avoid splitting annotated transcripts. The reads were aligned to the “rotated” reference genome using tophat with a new bowtie index and transcriptome index. A custom script was then used to convert the new alignments to the original coordinates and the new alignments were added to the original alignments and displayed on the UCSC genome browser.

Alignment of reads containing tandem repeats of the late leader exon from late gene mRNAs that would not align to the viral genome directly was accomplished by doing a direct Bowtie1 end-to-end alignment to predefined reference sequences allowing for up to 2 mismatches. Three references contained 42 bases of the 5’ UTR of VP1, VP2, or VP3 followed by the 57 base late leader exon and the first 42 bases of the next late leader. One reference contained the last 42 bases of the late leader followed by a full 57 base late leader followed by the first 42 bases of the transcription start region upstream of the late leader exon. The last reference sequence contained a full 57 base late leader flanked by 42 bases of each late leader on either side. This allowed 100-base reads spanning at least one leader-to-leader splice junction that would not align more than once to the same reference sequence to align.

A to G mismatches were identified using custom scripts that split the fastq reads into smaller 25nt fragments and filtered out any that could align to the mouse genome. The remaining fragments were aligned to the Py59RA genome using Bowtie1, allowing for up to 3 mismatches. Another custom postprocessing script was used to interpret
bowtie alignment output format and scan mismatch fields for canonical editing substitutions.

The hyper editing scripts from Porath et al. 2014 were downloaded and run in the context of our cluster pipeline. Custom scripts were used to extract information from output files, summarize, and plot as tracks on the UCSC genome browser.

The relative expression of host genes between time course Mock and PyV infected samples was determined using Cuffdiff v 2.1.1. Dispersion method pooled and dispersion method per condition were used to compare individual replicates and all replicates of a given condition. Custom scripts were used to post process/ filter cuffdiff output to produce time-series instances of significant logFoldchange. Cluster 3.0 was used to produce a k-means clustering of time course gene expression. Genes with a log2 fold change of +/-0.75 representing a 1.5 fold change with an fpkm greater than or equal to 1 were used for gene ontology and transcription factor prediction.

Gene ontology analysis was done using the browser-based application DAVID (Database for Annotation, Visualization, and Integrated Discovery)(http://david.abcc.ncifcrf.gov/home.jsp). Gene lists of upregulated and down regulated genes were each run against a Mus musculus background and gene categories with a false discovery rate below 1E-03 were used with redundant categories removed.

Transcription factor binding prediction was done using the browser-based application Transfind (http://transfind.sys-bio.net/). The same gene lists of up and down regulated genes were used with a 1000nt promoter set (800nt upstream and 200nt downstream TSS) with 1000 genes with the highest predicted factor affinities to take. The TRANSFAC (highest info content for each factor) setting was used to represent the factors.
III. RESULTS

A. RNA expression in NIH 3T6 cells infected with the Py59RA strain of mouse polyomavirus.

Mouse NIH 3T6 cells were serum starved to partially synchronize their cell cycles and then infected without (mock infection) or with filtered Py59RA virus (PyV) at a multiplicity of infection of 50. Total RNA was harvested at 12, 18, 24, and 36-hours post infection from three biological replicates for each condition and was used in the synthesis of stranded cDNA sequencing libraries using the Illumina TruSeq Stranded Total RNA Sample Preparation Kit. This resulted in libraries ~260 base pairs in length (Fig. 6). The libraries were sequenced using the Illumina HiSeq 2000 platform using paired end reads of 100 bases in the forward and reverse directions. Sequencing data were aligned using TopHat and normalized transcript abundance was calculated in fragments per kilobase of exon per million fragments mapped (FPKM) using Cufflinks. Significant differences between the virus and mock infections at each time point were determined using Cuffdiff. Each condition had three biological replicates producing 10-50 million 100 base reads with close to 80% aligning to the host genome and over 1% aligning to the viral genome by 36 hours post infection (Table 1).
Figure 6. Library Sizes

Libraries used for sequencing were run on Agilent Technologies 2200 Tape Station to verify library size. Libraries were made from 3 biological replicates of mock and PyV infected samples at 12, 18, 24, and 36 hour time points. The average size of the libraries was 260 bases.
Library Sizes

Replicate 1

Replicate 2

Replicate 3
Table 1. Viral RNA splicing events during infection

Number of aligned reads spanning viral splice junctions during the time course across three biological replicates. Percentages shown are with respect to total early or total late splice junction alignments
<table>
<thead>
<tr>
<th></th>
<th>Total Reads</th>
<th>Paired Alignment to Mouse Genome</th>
<th>Paired Alignment to Polyomavirus Genome</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mock 12_1</strong></td>
<td>43,640,510</td>
<td>34,532,713 (79.1%)</td>
<td>-</td>
</tr>
<tr>
<td><strong>Mock 12_2</strong></td>
<td>13,407,044</td>
<td>10,084,100 (75.2%)</td>
<td>-</td>
</tr>
<tr>
<td><strong>Mock 12_3</strong></td>
<td>20,726,081</td>
<td>14,381,090 (69.4%)</td>
<td>-</td>
</tr>
<tr>
<td><strong>Py 12_1</strong></td>
<td>44,588,864</td>
<td>33,353,378 (74.8%)</td>
<td>1,710 (&lt;0.01%)</td>
</tr>
<tr>
<td><strong>Py 12_2</strong></td>
<td>21,230,860</td>
<td>18,103,847 (85.3%)</td>
<td>1081 (&lt;0.01%)</td>
</tr>
<tr>
<td><strong>Py 12_3</strong></td>
<td>22,857,265</td>
<td>18,115,603 (79.3%)</td>
<td>1239 (&lt;0.01%)</td>
</tr>
<tr>
<td><strong>Mock 18_1</strong></td>
<td>43,100,950</td>
<td>33,496,251 (77.7%)</td>
<td>-</td>
</tr>
<tr>
<td><strong>Mock 18_2</strong></td>
<td>15,771,155</td>
<td>12,276,902 (77.8%)</td>
<td>-</td>
</tr>
<tr>
<td><strong>Mock 18_3</strong></td>
<td>12,835,580</td>
<td>9,826,513 (76.5%)</td>
<td>-</td>
</tr>
<tr>
<td><strong>Py 18_1</strong></td>
<td>43,139,743</td>
<td>32,662,054 (75.7%)</td>
<td>96,334 (0.22%)</td>
</tr>
<tr>
<td><strong>Py 18_2</strong></td>
<td>26,910,215</td>
<td>21,803,365 (81.0%)</td>
<td>44,146 (0.16%)</td>
</tr>
<tr>
<td><strong>Py 18_3</strong></td>
<td>14,252,449</td>
<td>10,824,627 (75.9%)</td>
<td>29,203 (0.20%)</td>
</tr>
<tr>
<td><strong>Mock 24_1</strong></td>
<td>40,147,080</td>
<td>32,936,666 (82.0%)</td>
<td>-</td>
</tr>
<tr>
<td><strong>Mock 24_2</strong></td>
<td>21,984,888</td>
<td>16,516,224 (75.1%)</td>
<td>-</td>
</tr>
<tr>
<td><strong>Mock 24_3</strong></td>
<td>21,103,516</td>
<td>16,094,651 (76.3%)</td>
<td>-</td>
</tr>
<tr>
<td><strong>Py 24_1</strong></td>
<td>50,787,902</td>
<td>40,004,936 (78.8%)</td>
<td>612,197 (1.2%)</td>
</tr>
<tr>
<td><strong>Py 24_2</strong></td>
<td>23,365,973</td>
<td>18,637,909 (79.8%)</td>
<td>356,237 (1.52%)</td>
</tr>
<tr>
<td><strong>Py 24_3</strong></td>
<td>22,087,645</td>
<td>17,404,983 (78.8%)</td>
<td>409,726 (1.86%)</td>
</tr>
<tr>
<td><strong>Mock 36_1</strong></td>
<td>44,049,053</td>
<td>35,120,311 (79.7%)</td>
<td>-</td>
</tr>
<tr>
<td><strong>Mock 36_2</strong></td>
<td>20,282,845</td>
<td>17,153,722 (84.6%)</td>
<td>-</td>
</tr>
<tr>
<td><strong>Mock 36_3</strong></td>
<td>16,873,257</td>
<td>9,811,783 (58.2%)</td>
<td>-</td>
</tr>
<tr>
<td><strong>Py 36_1</strong></td>
<td>41,028,401</td>
<td>28,135,319 (68.6%)</td>
<td>600,650 (1.46%)</td>
</tr>
<tr>
<td><strong>Py 36_2</strong></td>
<td>16,629,665</td>
<td>11,053,515 (66.5%)</td>
<td>314,095 (1.89%)</td>
</tr>
<tr>
<td><strong>Py 36_3</strong></td>
<td>9,748,905</td>
<td>5,795,074 (59.5%)</td>
<td>198,156 (2.03%)</td>
</tr>
</tbody>
</table>
B. The polyomavirus transcriptome.

In order to visualize changes in viral transcripts at different time points during infection, we used Bowtie and Top Hat to align sequencing reads to a Py59RA reference genome (Fig. 7). At 12 hours after infection, reads mapped to both early and late strands, but with most representing early-strand expression. While this is consistent with a model where the early promoter dominates before the onset of viral DNA replication (Kern 1986), it could also be consistent with a model in which the early-late switch is dependent on relative transcript stability rather than on regulation of transcription initiation (Hyde-DeRuyscher 1988). By 18 hours after infection viral replication has commenced and the infection is entering the late phase. At this point, there is an approximate 100-fold increase in reads aligning to the late strand and these transcripts are beginning to dominate over early-strand transcripts (Fig. 8). Also of note at this time is the appearance of a significant number of transcripts that extend beyond the late polyadenylation site, representing multi-genomic late-strand transcripts (Acheson 1971, Gu 2009). While most fully polyadenylated and spliced late RNAs are cytoplasmic (Kamen 1976, Acheson 1978), these giant RNAs are nuclear and serve as precursors to late mRNAs (Turler 1976, Flavell 1977, Acheson 1978). By 24 and 36 hours after infection the number of late strand alignments increases to around 1000 fold compared to early times. We note that at late times the majority of fully processed viral mRNAs have accumulated in the cytoplasm (Kamen 1975, Beard 1976, Hyde-DeRuyscher 1990).

While the relative levels of early-strand and late-strands change dramatically during infection, we noted no significant changes in patterns of RNA splicing, but did observe several interesting and unexpected changes to early-strand RNA accumulation
Figure 7. Py59RA reference sequence

5' splice sites are indicated in blue. 3' Splice sites are indicated in red. The late leader exon is shown with the black box. The early poly(A) site shown with a blue box. The complement of the late poly(A) site is shown with a red box.
XXX


NIH 3T6 cells were infected with Py59RA at an MOI of 50 or with a mock infection. Total RNA was harvested at 12, 18, 24, and 36 hours and stranded cDNA libraries were prepared for sequencing on the HiSeq 2000 sequencer. Reads were aligned to the Py59RA genome and visualized on the UCSC genome browser. Time course reads aligned to the Py59RA genome unscaled to show differential expression between early (top blue) and late (bottom red) strands.
during infection. First, it is clear that, relative to the bulk of early-strand RNAs, there is elevated alignment to a roughly 200 base pair region just downstream of the early polyadenylation site (Fig. 9). This peak was not seen when DNA replication and the early-late switch is blocked by aphidicolin (Fig. 10). In this experiment, mouse NIH3T6 cells were infected with Py59RA virus in the same manner as the time course except they were treated with either the same media or media with 2 µg/ml of the DNA replication inhibitor aphidicolin to prevent the infection from entering the late phase (Liu 1993). The infected samples, both treated and untreated, were incubated for 40 and 48 hours before RNA was collected for stranded cDNA library synthesis using the same protocol as the time course. This allowed sufficient accumulation of early gene products without the untreated samples reaching the lysis phase of the viral life cycle. Results were only aligned to the viral genome since aphidicolin treatment would likely cause significant alterations to the host transcriptome.

These reads appear to result from poly(A) site read-through and are antisense to VP1. While we do not yet know why or how they accumulate, one possible explanation would be that they represent precursors for viral-encoded microRNA(s). Polyomavirus and SV40 have previously shown viral expression of microRNAs antisense to Large T antigen, which were hypothesized to help avoid a cytotoxic T cell response (Sullivan 2005, Sullivan 2009). Alternatively, the increase in sequences from this region could result from uneven processing/degradation of early-strand transcripts that have read through the early polyadenylation site.

The second unexpected change in early-strand RNA alignments during infection is in the early promoter/origin/noncoding control region. We originally noted changes in
Figure 9. Alignment of the time course reads to the Py59RA genome (scaled)

NIH 3T6 cells were infected with Py59RA at an MOI of 50 or with a mock infection. Total RNA was harvested at 12, 18, 24, and 36 hours and stranded cDNA libraries were prepared for sequencing on the HiSeq 2000 sequencer. Reads were aligned to the Py59RA genome and visualized on the UCSC genome browser. Time course reads aligned to the Py59RA genome scaled to show changes in early strand alignment. Top blue strand is early, bottom red is late. The transcription start site can be seen shifting across the origin to the far right on the early strand. In addition a peak can be seen on the early strand downstream of the polyA site in the center between 3200 and 3500 bases.
NIH 3T6 cells were infected with Py59RA at an MOI of 50 or with a mock infection and treated with or without the DNA replication inhibitor aphidicolin at a concentration of 2µg/ml (5.9µM). Total RNA was harvested at 48 hours and stranded cDNA libraries were prepared for sequencing on the HiSeq 2000 sequencer. Reads were aligned to the Py59RA genome and visualized on the UCSC genome browser. (A) Samples of 48 hr infections with and without the DNA replication inhibitor aphidicolin to block the initiation of late phase aligned to the Py59RA genome unscaled to show differential expression between early and late strands. (B) Scaled to show changes in early strand alignment.
alignments to the edges of our reference genome. By aligning the reads to a linearized reference genome shifted by about 150 bp, we were better able to visualize the nature of these molecules, which span the replication origin. As infection proceeds, there is a clear increase in the alignment of early-strand reads to the origin of replication and enhancer region, strongly suggesting a potential upstream shift in early-strand transcription start sites during infection, and particularly after the onset of viral DNA replication (Fig. 7). We note that even at late times after infection in the aphidicolin treated samples, this shift is not observed (Fig. 8). Such a shift from a strict early transcription start site downstream of a canonical TATA box at early times (e.g., the 12 hour time point) to variable upstream transcription start sites at late times was suggested previously to explain confounding results obtained by S1 nuclease experiments (Kamen 1982). Mapping of the hybridization of 5’ ends of cytoplasmic early-strand transcripts at late times to a region of the polyoma genome containing the noncoding control region and part of the early region showed an apparent increase in usage of upstream start sites (Kamen 1982).

C. Splicing analysis

Using TopHat to map the splice junction reads we were able to determine the relative expression levels of the 5 viral isoforms that contain splice junctions (Table 2). At 12 hours post infection the mRNA for large T represents 75% of early-strand mRNAs, with mRNAs for middle T and small t accounting for 20% and 5%, respectively. We also note that a fourth early mRNA, for “tiny T antigen” encoding just the DnaJ domain (Riley 1997) has been reported and this is seen in our data, although this message is of
Table 2. Viral splicing events during infection

Number of aligned reads spanning viral splice junctions during the time course across three biological replicates. Percentages shown are with respect to total early or total late splice junction alignments.
<table>
<thead>
<tr>
<th>Time</th>
<th>Large T</th>
<th>middle T</th>
<th>small t</th>
<th>tiny t</th>
<th>VP1</th>
<th>VP3</th>
</tr>
</thead>
<tbody>
<tr>
<td>12 Hours</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rep 1</td>
<td>50 (75.8%)</td>
<td>13 (19.7%)</td>
<td>3 (4.5%)</td>
<td>0 (0%)</td>
<td>46 (88.5%)</td>
<td>6 (11.5%)</td>
</tr>
<tr>
<td>Rep 2</td>
<td>40 (74.1%)</td>
<td>9 (16.6%)</td>
<td>5 (9.3%)</td>
<td>0 (0%)</td>
<td>20 (80%)</td>
<td>5 (20%)</td>
</tr>
<tr>
<td>Rep 3</td>
<td>72 (87.7%)</td>
<td>8 (9.8%)</td>
<td>2 (2.5%)</td>
<td>0 (0%)</td>
<td>12 (80%)</td>
<td>3 (20%)</td>
</tr>
<tr>
<td>18 hours</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rep 1</td>
<td>589 (62.9%)</td>
<td>175 (18.7%)</td>
<td>172 (18.4%)</td>
<td>0 (0%)</td>
<td>2735 (89.6%)</td>
<td>319 (10.4%)</td>
</tr>
<tr>
<td>Rep 2</td>
<td>354 (69.3%)</td>
<td>82 (16%)</td>
<td>75 (14.7%)</td>
<td>0 (0%)</td>
<td>1058 (77.3%)</td>
<td>311 (22.7%)</td>
</tr>
<tr>
<td>Rep 3</td>
<td>250 (69.6%)</td>
<td>58 (16.2%)</td>
<td>51 (14.2%)</td>
<td>0 (0%)</td>
<td>863 (80.9%)</td>
<td>204 (19.1%)</td>
</tr>
<tr>
<td>24 Hours</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rep 1</td>
<td>1686 (64.2%)</td>
<td>559 (21.2%)</td>
<td>380 (14.4%)</td>
<td>6 (0.2%)</td>
<td>41245 (90.2%)</td>
<td>4133 (9.1%)</td>
</tr>
<tr>
<td>Rep 2</td>
<td>788 (63.1%)</td>
<td>237 (19%)</td>
<td>218 (17.4%)</td>
<td>7 (0.5%)</td>
<td>14203 (80%)</td>
<td>3553 (20%)</td>
</tr>
<tr>
<td>Rep 3</td>
<td>1074 (65.1%)</td>
<td>306 (18.5%)</td>
<td>262 (15.8%)</td>
<td>10 (0.6%)</td>
<td>17825 (79.6%)</td>
<td>4570 (20.4%)</td>
</tr>
<tr>
<td>36 hours</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rep 1</td>
<td>1822 (63.7%)</td>
<td>589 (20.6%)</td>
<td>442 (15.4%)</td>
<td>10 (0.3%)</td>
<td>39525 (90.2%)</td>
<td>4294 (9.8%)</td>
</tr>
<tr>
<td>Rep 2</td>
<td>630 (59.5%)</td>
<td>250 (23.6%)</td>
<td>169 (16%)</td>
<td>10 (0.9%)</td>
<td>16453 (78.3%)</td>
<td>4555 (21.7%)</td>
</tr>
<tr>
<td>Rep 3</td>
<td>408 (60.6%)</td>
<td>162 (24%)</td>
<td>97 (14.4%)</td>
<td>7 (1%)</td>
<td>11087 (80.1%)</td>
<td>2760 (19.9%)</td>
</tr>
</tbody>
</table>
very low relative abundance. As infection proceeds, the ratios of large, middle and small T antigen mRNAs change little, indicating that early strand splicing is likely not regulated strongly by viral or host proteins that change during infection. Like early mRNAs, late splicing patterns do not change dramatically during infection, although the absolute numbers of alignments increase up to 1000 fold, consistent with the strong upregulation of late gene expression.

D. Leader-to-leader splicing.

Owing to the presence of tandem repeats of late leaders in all three late mRNAs, alternative reference sequences were required to capture reads that span at least one leader-to-leader splice junction (Table 3). The stranded cDNA libraries produce 2x100 base pair reads. The late leader exon is 57 bases long so we required a reference sequence that captured reads that extended at least 1 base past the first leader-to-leader splice junction both downstream and upstream of the late leader on VP1, VP2, and VP3 cDNA as well as the late transcription start sites. We used a reference sequence containing 42 bases upstream of the late leader exon near the transcription start region followed by a full late leader exon and the first 42 bases of the next late leader (Table 3 bottom of first column). We also used a reference sequence of the last 42 bases of late leader, a full late leader, and the first 42 bases of late leader to capture ambiguous leader to leader reads from any mRNA crossing at least 2 leader to leader splice junctions (Table 3 bottom of second column). Reference sequences containing the last 42 bases of a late leader followed by a full late leader and 42 bases of the 5’ UTR of VP1, VP2 or VP3 (Table 3 bottom of third, fourth, and fifth columns). Leader-to-leader splices were
Table 3. Alignment of reads from late leader repeats.

The 100 base reads from time course and aphidicolin samples were realigned to custom reference sequences containing at least one leader to leader splice junction. The reference sequences were 5 variations of a 42-57-42 base splice of late leader exons to each other, the transcription start site, or the 5’UTR of VP1-3. The number of reads aligning to each reference sequence for each condition is shown.
## Late Leader-to-Leader Splicing

<table>
<thead>
<tr>
<th>Time</th>
<th>TSS Late Leader</th>
<th>&gt;2 Late Leaders</th>
<th>VP1</th>
<th>VP2</th>
<th>VP3</th>
</tr>
</thead>
<tbody>
<tr>
<td>12 hours</td>
<td>0</td>
<td>46</td>
<td>15</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>18 hours</td>
<td>20</td>
<td>2687</td>
<td>513</td>
<td>665</td>
<td>111</td>
</tr>
<tr>
<td>24 hours</td>
<td>550</td>
<td>35989</td>
<td>8419</td>
<td>4335</td>
<td>1809</td>
</tr>
<tr>
<td>36 hours</td>
<td>736</td>
<td>36432</td>
<td>8680</td>
<td>4043</td>
<td>1807</td>
</tr>
<tr>
<td>48 Aphi-</td>
<td>3250</td>
<td>149466</td>
<td>25902</td>
<td>8291</td>
<td>6117</td>
</tr>
<tr>
<td>48 Aphi+</td>
<td>4</td>
<td>398</td>
<td>50</td>
<td>21</td>
<td>11</td>
</tr>
</tbody>
</table>
only seen at significant levels in time points later than 18 hours and in cells untreated with aphidicolin, supporting the model that leader to leader splicing occurs after DNA replication.

**E. A-to-I editing of viral RNAs.**

Due to the read-through of early and late transcripts at late times, as well as the genomic overlap of the early and late polyadenylation signals, there is the possibility that if complementary sequences accumulate near one another in the nucleus, they might form double stranded RNAs. As nuclear dsRNAs can be promiscuously edited by ADAR enzymes, which deaminate adenosines to inosines (Bass 2002), viral dsRNAs should also exhibit this fate. We have reported previously that this is indeed the case (Kumar 1997, Gu 2009), but the current study allowed us to examine viral RNA editing in more comprehensive detail, which was done in two ways. Since inosine base pairs like guanosine, we first looked for individual A-G mismatches in the read alignments to search for potential sites for editing. We divided the 100 nucleotide reads into four 25-nucleotide fragments and allowed up to three base mismatches during the alignment step (Zhu 2013). After filtering for any sequences that aligned to the mouse genome, we realigned the sequences to the viral genome to show regions of A-G mismatches (Fig. 11). Such alignment cannot be conclusive for RNA editing, as it could merely reflect sequencing or alignment errors. Indeed, this sort of analysis, when applied to other base transformations, generated similar, though less robust data. We therefore examined viral regions where A-G mismatches greatly outnumbered mismatches caused by any other base changes. The most striking region associated with A-G changes was the
Figure 11. Reads with up to 3 A-G mismatches as potential ADAR editing sites.

Reads from the time course run were broken from 100 bases to 25 bases and realigned with a threshold of 3 mismatches allowed per 25 base read. Reads with A-G mismatches were visualized using the UCSC genome browser to indicate regions with potential editing sites. The polyadenylation sites showed the highest number of A-G mismatched reads at late times of infection.
A-G Mismatches

Small T
Middle T
Large T

12 hpi

18 hpi

24 hpi

36 hpi
overlapping polyadenylation region, which showed a much higher rate of A-G mismatch compared to other combinations (Table 4). This was not surprising since the polyadenylation region has previously shown to be edited at late times during infection (Gu 2009).

As long dsRNA is subject to hyperediting, where up to 50% of the adenosines are converted to inosines (Kumar 1997), it is quite likely that the above analysis missed much or most of the editing that occurs during polyomavirus infection. In order to capture a greater fraction of hyper edited reads that may have been omitted due to having too many A-G mismatches we used a method recently described by Porath et al 2014. These authors developed a pipeline that sorts unaligned reads based on whether they fail to align due to a specific type of mismatch and whether such mismatches occur in the types of clusters associated with hyperediting. The first step was to align all the reads to the mouse genome to reduce the total number of reads and remove most non-viral sequences. Second, we aligned the remaining reads to the viral genome and discarded any that aligned perfectly or with few mismatches. Third, we transformed all As in both the reads and the reference genome to Gs and discarded any transcripts that did not align. This allowed us to only sort for reads that initially had A to G mismatches. Fourth, the mismatches were clustered based on predictions for known A-I editing including adjacent sequences and only reads that clustered in a manner consistent with A-I editing were kept. Finally the bases that were members of clusters were displayed on the UCSC genome browser (Fig. 12). Results showed a time dependent increase in the number of bases that were members of hyperedited clusters, with a large concentration of such
Table 4. Single base mismatches across the polyoma genome.

Base to base mismatches from the 36 hour time point found in the 25 base read alignments allowing for up to 3 mismatches. Mismatches in and around the poly(A) sites show the highest number of A to G mismatch. Early and late poly(A) sites in bold. Each position represents both strands.
<table>
<thead>
<tr>
<th>sample</th>
<th>pos</th>
<th>depth</th>
<th>A&gt;C</th>
<th>A&gt;G</th>
<th>A&gt;T</th>
<th>C&gt;A</th>
<th>C&gt;G</th>
<th>C&gt;T</th>
<th>G&gt;A</th>
<th>G&gt;C</th>
<th>G&gt;T</th>
<th>T&gt;A</th>
<th>T&gt;C</th>
</tr>
</thead>
<tbody>
<tr>
<td>PyV-36</td>
<td>2898</td>
<td>19703</td>
<td>0</td>
<td>71</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>PyV-36</td>
<td>2899</td>
<td>20055</td>
<td>0</td>
<td>12</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>PyV-36</td>
<td>2900</td>
<td>20701</td>
<td>1</td>
<td>169</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>PyV-36</td>
<td>2901</td>
<td>19906</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PyV-36</td>
<td>2902</td>
<td>19666</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>PyV-36</td>
<td>2903</td>
<td>19472</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>PyV-36</td>
<td>2904</td>
<td>19230</td>
<td>0</td>
<td>15</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>PyV-36</td>
<td>2905</td>
<td>18764</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PyV-36</td>
<td>2906</td>
<td>18365</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>PyV-36</td>
<td>2907</td>
<td>17681</td>
<td>1</td>
<td>39</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PyV-36</td>
<td>2908</td>
<td>17089</td>
<td>0</td>
<td>125</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PyV-36</td>
<td>2909</td>
<td>16552</td>
<td>0</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>PyV-36</td>
<td>2910</td>
<td>16587</td>
<td>0</td>
<td>26</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>PyV-36</td>
<td>2911</td>
<td>16586</td>
<td>1</td>
<td>17</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>PyV-36</td>
<td>2912</td>
<td>16584</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>PyV-36</td>
<td>2913</td>
<td>16581</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>PyV-36</td>
<td>2914</td>
<td>16569</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>PyV-36</td>
<td>2915</td>
<td>16508</td>
<td>4</td>
<td>130</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>PyV-36</td>
<td>2916</td>
<td>14582</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>PyV-36</td>
<td>2917</td>
<td>15277</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>PyV-36</td>
<td>2918</td>
<td>4649</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>PyV-36</td>
<td>2920</td>
<td>4667</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>PyV-36</td>
<td>2921</td>
<td>4666</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>PyV-36</td>
<td>2922</td>
<td>4669</td>
<td>4</td>
<td>162</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>PyV-36</td>
<td>2923</td>
<td>4672</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>PyV-36</td>
<td>2924</td>
<td>4089</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>PyV-36</td>
<td>2925</td>
<td>3576</td>
<td>3</td>
<td>31</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>PyV-36</td>
<td>2926</td>
<td>3566</td>
<td>5</td>
<td>43</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>PyV-36</td>
<td>2927</td>
<td>3571</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>PyV-36</td>
<td>2928</td>
<td>3581</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
Figure 12. Accumulation of hyper-editing clusters during the time course.

Reads from the time course were aligned with the host mouse genome and discarded. Reads were then aligned to the Py59RA reference genome and discarded. A’s in reads that did not align to the Py59RA genome were changed to G’s and realigned to a Py59RA reference genome that itself had all A’s changed to G’s. This allowed for reads that did not originally align to the unaltered Py59RA genome due to hyperediting to be captured. The process was repeated for all combinations of single base mismatch. Clusters of A-G mismatches from the time course consistent with hyperediting by ADAR were mapped to the Py59RA genome.
clusters around the overlapping polyadenylation sites. No significant editing was detected in aphidicolin treated samples (data not shown). Together, these data are consistent with our previously reported polyoma editing results (Heiser 1982, Gu 2009). Shortly after the early late switch at 18 hours we see the polyadenylation region (and a cryptic poly(A) site in the early region, just downstream of the stop codon for middle T antigen) showing editing clusters. As the infection progresses the number and size of the clusters increases on both strands. This is consistent with editing resulting from early and late strand hybridization since the over abundance of late sequence compared to early would only produce more clusters if the late strand formed hairpins and was edited in cis. Editing clusters can also be seen in the noncoding control region suggesting that multi-genome length giant transcripts are capable of annealing with one another. While editing has long been predicted due to the long antisense transcripts and poly(A) read-through, this is the first time we have had direct evidence of genomewide editing at late times. We note, however, that while editing is readily and abundantly observed, we cannot conclude from these data whether it is a cause or a consequence of viral gene regulation.

F. Effects of viral infection on host gene expression.

Our deep sequencing data allowed us to not only examine viral gene expression during infection, but also changes in the host transcriptome that result from the infection. Using Cuffdiff to compare gene expression between polyomavirus and mock infected samples, we found 1,216 genes were differentially expressed at least 1.5 fold across all three replicates with 359 genes upregulated and 857 genes downregulated compared to the mock infections at the 36 hour time point (Table 5).
Table 5. Number of host genes differentially expressed between mock and Py59RA infected samples

The number of host transcripts that changed more than 1.5 fold across all three biological replicates was determined using Cuffdiff. Only genes with expression >1 fpkm were counted.
<table>
<thead>
<tr>
<th></th>
<th>12 hours</th>
<th>18 hours</th>
<th>24 hours</th>
<th>36 hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upregulated</td>
<td>0</td>
<td>6</td>
<td>12</td>
<td>359</td>
</tr>
<tr>
<td>Downregulated</td>
<td>0</td>
<td>7</td>
<td>19</td>
<td>857</td>
</tr>
<tr>
<td>Total</td>
<td>0</td>
<td>13</td>
<td>31</td>
<td>1,217</td>
</tr>
</tbody>
</table>
We used the database for annotation, visualization, and integrated discovery (DAVID) (http://david.abcc.ncifcrf.gov/) to determine the gene ontology of the differentially expressed genes that were significantly enriched based on common biological function (Table 6). We also selected genes of high abundance with a strong up or down regulation between mock and virus infected cells at 36 hours for validation by RT-qPCR (Fig. 13 and 14).

For the upregulated genes the categories with the highest enrichment included transcripts coding for ribosomal proteins, chromatin binding proteins, DNA replication, RNA processing, and mitochondrial proteins. These are all consistent with the known increase in growth and metabolism induced by the combined action of the early genes. For example, the transcript for Hist1h1a that codes for the histone 1 linker involved in compacting DNA during replication (van Holde 1988) is upregulated more than 2 fold in infected cells by 36 hours (Fig. 15). The transcript for the mitochondrial inner membrane protein Timm8a1 that is involved in the insertion of cytoplasmic proteins into the mitochondrial inner membrane (Paschen 2000) is also upregulated. Another interesting observation is the transcript coding for the JunB component of AP-1 transcription factors is upregulated. This is of interest because JunB was originally understood to function as an antagonist of the AP-1 protein c-Jun, which is involved in cell cycle progression (Chiu 1989, Schutte 1989). Later observations found this interaction to be more complex and sensitive to cell type and context (Shaulian 2002).

Downregulated genes were enriched in categories such as GTPase regulation, cell adhesion, actin cytoskeleton, extracellular matrix, and transcription regulation. One of the more strongly downregulated transcripts included Col1a2 (Fig. 15) which codes for
Table 6. Gene ontology analysis

The web based application Database for Annotation, Visualization and Integrated Discovery (DAVID) was used to sort lists of significantly upregulated or downregulated genes (>1.5 fold) from the 36 hour time point by common function and enrichment using the Mus musculus background. A false discovery threshold of 1E-03 was used.
<table>
<thead>
<tr>
<th>Pathway</th>
<th>Gene Count</th>
<th>% of upregulated genes</th>
<th>P-value</th>
<th>Fold Enrichment</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Upregulated</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ribosomal Protein</td>
<td>38</td>
<td>11.95</td>
<td>3.60E-30</td>
<td>12.86</td>
<td>4.65E-27</td>
</tr>
<tr>
<td>RNA Processing</td>
<td>31</td>
<td>9.75</td>
<td>3.29E-10</td>
<td>3.9</td>
<td>5.19E-07</td>
</tr>
<tr>
<td>Chromosomal Protein</td>
<td>19</td>
<td>8.5</td>
<td>1.19E-11</td>
<td>8.37</td>
<td>1.53E-08</td>
</tr>
<tr>
<td>Mitochondrial Protein</td>
<td>42</td>
<td>13.2</td>
<td>9.30E-11</td>
<td>3.18</td>
<td>1.20E-07</td>
</tr>
<tr>
<td>DNA Replication</td>
<td>18</td>
<td>5.66</td>
<td>2.16E-09</td>
<td>6.51</td>
<td>3.41E-06</td>
</tr>
<tr>
<td>Ribosome Biogenesis</td>
<td>14</td>
<td>4.4</td>
<td>1.13E-07</td>
<td>6.87</td>
<td>1.79E-04</td>
</tr>
<tr>
<td>Pyrimidine Metabolism</td>
<td>16</td>
<td>5.03</td>
<td>1.50E-09</td>
<td>7.59</td>
<td>1.55E-06</td>
</tr>
<tr>
<td><strong>Downregulated</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GTPase Regulator</td>
<td>58</td>
<td>6.98</td>
<td>1.19E-16</td>
<td>3.49</td>
<td>1.67E-13</td>
</tr>
<tr>
<td>Cell Adhesion</td>
<td>61</td>
<td>7.34</td>
<td>8.60E-10</td>
<td>2.35</td>
<td>1.52E-06</td>
</tr>
<tr>
<td>Actin Cytoskeleton</td>
<td>32</td>
<td>3.85</td>
<td>7.04E-10</td>
<td>3.66</td>
<td>9.73E-07</td>
</tr>
<tr>
<td>Chromosome Organization</td>
<td>44</td>
<td>5.29</td>
<td>2.70E-07</td>
<td>2.36</td>
<td>4.76E-04</td>
</tr>
<tr>
<td>Embryonic Development</td>
<td>48</td>
<td>5.77</td>
<td>1.70E-08</td>
<td>2.47</td>
<td>2.99E-05</td>
</tr>
<tr>
<td>Transcription Regulation</td>
<td>158</td>
<td>19.01</td>
<td>8.99E-09</td>
<td>1.53</td>
<td>1.59E-05</td>
</tr>
<tr>
<td>Urogenital Development</td>
<td>27</td>
<td>3.25</td>
<td>2.60E-09</td>
<td>4</td>
<td>4.59E-06</td>
</tr>
<tr>
<td>Cell Motion</td>
<td>44</td>
<td>5.29</td>
<td>1.73E-08</td>
<td>2.59</td>
<td>3.05E-05</td>
</tr>
<tr>
<td>Extracellular Matrix</td>
<td>39</td>
<td>4.7</td>
<td>3.69E-09</td>
<td>2.96</td>
<td>5.09E-06</td>
</tr>
<tr>
<td>Pathways in Cancer</td>
<td>42</td>
<td>5.05</td>
<td>5.53E-10</td>
<td>2.9</td>
<td>6.27E-07</td>
</tr>
<tr>
<td>Zinc Finger</td>
<td>93</td>
<td>11.2</td>
<td>4.90E-08</td>
<td>1.79</td>
<td>6.79E-05</td>
</tr>
</tbody>
</table>
Figure 13. Validation of upregulated genes.

One noncoding and three coding genes with a high expression and at least 1.5 fold change were selected for validation of upregulated genes by qPCR. FPKM results shown in left column. qRT-PCR results shown in the right column.
Figure 14. Validation of downregulated genes.

One noncoding and three coding genes with a high expression and at least 1.5 fold change were selected for validation of downregulated genes by qPCR. FPKM results shown in left column. qRT-PCR results shown in the right column.
Figure 15. Example of host protein coding transcripts significantly upregulated or downregulated in infected samples compared to mock infection.

Protein coding reads from the time course were aligned to the host genome. (Left) Expression of Hist1ha increases compared to mock infection. (Right) Col1a2 decreases compared to mock infection.
one of the polypeptide components of Collagen type 1 that makes up the extracellular matrix (Vuorio 1990). Another strongly downregulated transcript encodes the steroyl-CoA desaturase 2 (Scd2) enzyme that catalyzes the synthesis of monounsaturated fatty acids, increasing membrane fluidity (Ntambi 1999). Unlike in polyoma infection, Scd2 has been shown to be upregulated in other cancer cells and may be responsible for inducing hepatocellular carcinomas (Muir 2013). It is unclear how the virus would benefit from a reduction in monounsaturated fatty acid synthesis. The transcript for the protein myristolylated alanine-rich C-kinase substrate (Marcks) that is involved in altering the actin cytoskeleton (Laux 2000), cell motility (Brooks 1996), and functions as a tumor suppressor (Bickerboller 2015) was originally shown to be downregulated in mouse 3T3 fibroblast cells expressing v-Src (Joseph 1992). Our data also show significant down regulation of this transcript.

In addition to coding RNAs, we observed significant changes in the expression of noncoding RNAs (Table 7, Fig. 13 and 14). The noncoding small nucleolar RNA Snhg1 was highly expressed and significantly upregulated (Fig 16). Snhg1, also known as the host gene for U22, is expressed from an intron and is required for processing of 18S ribosomal RNA (Tycowski 1994). This is consistent with the increase in the expression of genes involved in ribosome biogenesis and translation in polyoma infected cells. Another noncoding RNA, Metastasis Associated Lung Adenocarcinma Transcript 1 (MALAT1), was significantly downregulated in infected cells (Fig. 16). Though the precise function of MALAT1 is still unclear, it has been shown to be involved in a variety of metastatic cancers (Ji 2003). Its downregulation is consistent with the downregulation of other genes involved in cell motility in polyoma-infected cells.
Table 7. Significant changes in noncoding RNAs.

List of noncoding RNAs by name, feature type, and fold change log2 ratio that change more than 1.5 fold between Mock and PyV infected samples by 36 hours after infection.
<table>
<thead>
<tr>
<th>Upregulated Noncoding RNA</th>
<th>Feature Type</th>
<th>Fold Change Log2 Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1600020E01Rik</td>
<td>lincRNA gene</td>
<td>1.406401543</td>
</tr>
<tr>
<td>2410006H16Rik</td>
<td>Unclassified ncRNA gene</td>
<td>1.091248082</td>
</tr>
<tr>
<td>3300005D01Rik</td>
<td>lincRNA gene</td>
<td>1.682985052</td>
</tr>
<tr>
<td>Gm11974</td>
<td>Unclassified ncRNA gene</td>
<td>1.685511363</td>
</tr>
<tr>
<td>Gm12603</td>
<td>lincRNA gene</td>
<td>2.060374312</td>
</tr>
<tr>
<td>Rmrp</td>
<td>RNase MRP RNA gene</td>
<td>1.78468046</td>
</tr>
<tr>
<td>Rpph1</td>
<td>RNase P RNA gene</td>
<td>2.765659453</td>
</tr>
<tr>
<td>Rprl3</td>
<td>RNase P RNA gene</td>
<td>2.656127479</td>
</tr>
<tr>
<td>Scarna13</td>
<td>snoRNA gene</td>
<td>1.048926376</td>
</tr>
<tr>
<td>Snhg1</td>
<td>Unclassified ncRNA gene</td>
<td>1.389978478</td>
</tr>
<tr>
<td>Terc</td>
<td>Telomerase RNA gene</td>
<td>1.669975844</td>
</tr>
<tr>
<td>Downregulated Noncoding RNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2900097C17Rik</td>
<td>Unclassified ncRNA gene</td>
<td>-0.81491907</td>
</tr>
<tr>
<td>Gm13375</td>
<td>Unclassified ncRNA gene</td>
<td>-1.242099564</td>
</tr>
<tr>
<td>Malat1</td>
<td>lincRNA gene</td>
<td>-1.412442083</td>
</tr>
</tbody>
</table>
Figure 16. Example of host noncoding transcripts significantly upregulated or downregulated in infected samples compared to mock infection.

Noncoding RNA reads from the time course were aligned to the host genome. (Left) Expression of Snhg2 increases compared to mock infection. (Right) Malat1 decreases compared to mock infection.
Taken together, these observations illustrate a change in priority of infected cells from cell-to-cell interaction, movement, and signal transduction to growth, metabolism, and DNA synthesis.

Do differentially expressed genes share common patterns of regulation that may help to explain their response to viral infection? In order to determine common transcription factor binding sites upstream of differentially expressed genes we used the Transfind tool (http://transfind.sys-bio.net/) to determine common transcription factor binding sites among the differentially expressed genes (Kielbasa 2010). Using a threshold of binding motifs common to the upstream regions of <1000 genes (out of 33,837 genes in the mouse database) and a false discovery rate <0.02, we found a number of common transcription factor binding sites (Table 8 and 9). As expected the promoters of upregulated genes were the most enriched for binding sites of the transcription factor E2F, which is a major cell cycle regulator that is released and activated by the binding of Large T antigen to pRB (Bracken 2004). Binding sites for members of the CREB-ATF family of transcription factors were also enriched in the upregulated gene set. This is consistent with CREB Binding Protein (CBP) and EP300 being targets for large T antigen, resulting in CRE containing promoters being activated (Goodman 2000). The nuclear factor kappa B (NFkB), which has been previously shown to be involved in inflammation, transformation, proliferation, angiogenesis, and metastasis shows binding sites upstream of the upregulated genes (Aggarwal 2004). Interestingly, STAT1, which is involved in signal transduction of the interferon pathway, appears to have a common binding site among a number of the upregulated genes. While the interferon pathway does not appear to be active (pathway genes were not noted to be strongly up or down
Table 8. Transfind transcription factor prediction for 36 hour upregulated.

The web based application Transfind was used to predict common transcription factor binding sites among the lists of upregulated and downregulated host genes. Predictions of mouse transcription factor binding sites were made with promoter sets of 1000nt from 800nt upstream to 200nt downstream and compared with 1000 genes with the highest predicted factor affinities among 33,837 total mouse genes. Possible factors were represented by the TRANSFAC matrix.
<table>
<thead>
<tr>
<th>Factor</th>
<th>p-value</th>
<th>FDR</th>
<th>FP</th>
<th># of genes +</th>
<th># of genes -</th>
<th>Background +</th>
<th>Background -</th>
</tr>
</thead>
<tbody>
<tr>
<td>E2F</td>
<td>&lt;1.00E-07</td>
<td>3.00E-09</td>
<td>9.00E-09</td>
<td>45</td>
<td>331</td>
<td>955</td>
<td>32882</td>
</tr>
<tr>
<td>NF-Y</td>
<td>2.74E-07</td>
<td>1.49E-05</td>
<td>5.94E-05</td>
<td>31</td>
<td>345</td>
<td>969</td>
<td>32868</td>
</tr>
<tr>
<td>HIF1</td>
<td>8.45E-07</td>
<td>3.06E-05</td>
<td>1.83E-04</td>
<td>30</td>
<td>346</td>
<td>970</td>
<td>32867</td>
</tr>
<tr>
<td>GABP</td>
<td>2.52E-06</td>
<td>6.83E-05</td>
<td>5.46E-04</td>
<td>29</td>
<td>347</td>
<td>971</td>
<td>32866</td>
</tr>
<tr>
<td>CREB</td>
<td>2.52E-06</td>
<td>6.83E-05</td>
<td>5.46E-04</td>
<td>29</td>
<td>347</td>
<td>971</td>
<td>32866</td>
</tr>
<tr>
<td>ATF</td>
<td>2.52E-06</td>
<td>6.83E-05</td>
<td>5.46E-04</td>
<td>29</td>
<td>347</td>
<td>971</td>
<td>32866</td>
</tr>
<tr>
<td>STAT1</td>
<td>2.00E-05</td>
<td>4.34E-04</td>
<td>4.34E-03</td>
<td>27</td>
<td>349</td>
<td>973</td>
<td>32864</td>
</tr>
<tr>
<td>CREB</td>
<td>2.00E-05</td>
<td>4.34E-04</td>
<td>4.34E-03</td>
<td>27</td>
<td>349</td>
<td>973</td>
<td>32864</td>
</tr>
<tr>
<td>CREBP1</td>
<td>5.32E-05</td>
<td>8.25E-04</td>
<td>1.16E-02</td>
<td>26</td>
<td>350</td>
<td>974</td>
<td>32863</td>
</tr>
<tr>
<td>ATF4</td>
<td>5.32E-05</td>
<td>8.25E-04</td>
<td>1.16E-02</td>
<td>26</td>
<td>350</td>
<td>974</td>
<td>32863</td>
</tr>
<tr>
<td>AHR</td>
<td>5.32E-05</td>
<td>8.25E-04</td>
<td>1.16E-02</td>
<td>26</td>
<td>350</td>
<td>974</td>
<td>32863</td>
</tr>
<tr>
<td>NFkB</td>
<td>7.88E-04</td>
<td>1.07E-02</td>
<td>1.71E-01</td>
<td>23</td>
<td>353</td>
<td>977</td>
<td>32860</td>
</tr>
<tr>
<td>YY1</td>
<td>7.88E-04</td>
<td>1.07E-02</td>
<td>1.71E-01</td>
<td>23</td>
<td>353</td>
<td>977</td>
<td>32860</td>
</tr>
<tr>
<td>ATF2</td>
<td>1.78E-03</td>
<td>1.93E-02</td>
<td>3.86E-01</td>
<td>22</td>
<td>354</td>
<td>978</td>
<td>32859</td>
</tr>
<tr>
<td>CJUN</td>
<td>1.78E-03</td>
<td>1.93E-02</td>
<td>3.86E-01</td>
<td>22</td>
<td>354</td>
<td>978</td>
<td>32859</td>
</tr>
<tr>
<td>C-REL</td>
<td>1.78E-03</td>
<td>1.93E-02</td>
<td>3.86E-01</td>
<td>22</td>
<td>354</td>
<td>978</td>
<td>32859</td>
</tr>
<tr>
<td>IRF1</td>
<td>1.78E-03</td>
<td>1.93E-02</td>
<td>3.86E-01</td>
<td>22</td>
<td>354</td>
<td>978</td>
<td>32859</td>
</tr>
<tr>
<td>ATF3</td>
<td>1.78E-03</td>
<td>1.93E-02</td>
<td>3.86E-01</td>
<td>22</td>
<td>354</td>
<td>978</td>
<td>32859</td>
</tr>
</tbody>
</table>
Table 9. Transfind transcription factor prediction for 36 hour downregulated.

The web based application Transfind was used to predict common transcription factor binding sites among the lists of upregulated and downregulated host genes. Predictions of mouse transcription factor binding sites were made with promoter sets of 1000nt from 800nt upstream to 200nt downstream and compared with 1000 genes with the highest predicted factor affinities among 33,837 total mouse genes. Possible factors were represented by the TRANSFAC matrix.
<table>
<thead>
<tr>
<th>Factor</th>
<th>p-value</th>
<th>FDR</th>
<th>FP</th>
<th># of genes +</th>
<th># of genes -</th>
<th>Background +</th>
<th>Background -</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP1</td>
<td>&lt;1E-09</td>
<td>3.00E-09</td>
<td>2.60E-08</td>
<td>85</td>
<td>781</td>
<td>915</td>
<td>32432</td>
</tr>
<tr>
<td>WT1</td>
<td>&lt;1E-09</td>
<td>3.00E-09</td>
<td>2.60E-08</td>
<td>82</td>
<td>784</td>
<td>918</td>
<td>32429</td>
</tr>
<tr>
<td>AP2</td>
<td>&lt;1E-09</td>
<td>3.00E-09</td>
<td>9.00E-09</td>
<td>81</td>
<td>785</td>
<td>919</td>
<td>32428</td>
</tr>
<tr>
<td>TFIII</td>
<td>&lt;1E-09</td>
<td>3.00E-09</td>
<td>2.60E-08</td>
<td>70</td>
<td>796</td>
<td>930</td>
<td>32417</td>
</tr>
<tr>
<td>MAZ</td>
<td>&lt;1E-09</td>
<td>3.00E-09</td>
<td>2.60E-08</td>
<td>69</td>
<td>797</td>
<td>931</td>
<td>32416</td>
</tr>
<tr>
<td>HIC1</td>
<td>&lt;1E-09</td>
<td>3.00E-09</td>
<td>2.60E-08</td>
<td>66</td>
<td>800</td>
<td>934</td>
<td>32413</td>
</tr>
<tr>
<td>KROX</td>
<td>&lt;1E-09</td>
<td>3.00E-09</td>
<td>9.00E-09</td>
<td>64</td>
<td>802</td>
<td>936</td>
<td>32411</td>
</tr>
<tr>
<td>AHRHIF</td>
<td>6.00E-09</td>
<td>1.10E-07</td>
<td>1.21E-06</td>
<td>58</td>
<td>808</td>
<td>942</td>
<td>32405</td>
</tr>
<tr>
<td>MZF1</td>
<td>3.40E-08</td>
<td>6.15E-07</td>
<td>7.39E-06</td>
<td>56</td>
<td>810</td>
<td>944</td>
<td>32403</td>
</tr>
<tr>
<td>MAZR</td>
<td>2.22E-06</td>
<td>3.44E-05</td>
<td>4.82E-04</td>
<td>51</td>
<td>815</td>
<td>949</td>
<td>32398</td>
</tr>
<tr>
<td>HES1</td>
<td>2.22E-06</td>
<td>3.44E-05</td>
<td>4.82E-04</td>
<td>51</td>
<td>815</td>
<td>949</td>
<td>32398</td>
</tr>
<tr>
<td>SPZ1</td>
<td>4.82E-06</td>
<td>6.97E-05</td>
<td>1.05E-03</td>
<td>50</td>
<td>816</td>
<td>950</td>
<td>32397</td>
</tr>
<tr>
<td>EGR1</td>
<td>1.02E-05</td>
<td>1.31E-04</td>
<td>2.22E-03</td>
<td>49</td>
<td>817</td>
<td>951</td>
<td>32396</td>
</tr>
<tr>
<td>MTF1</td>
<td>1.02E-05</td>
<td>1.31E-04</td>
<td>2.22E-03</td>
<td>49</td>
<td>817</td>
<td>951</td>
<td>32396</td>
</tr>
<tr>
<td>EGR2</td>
<td>2.13E-05</td>
<td>2.43E-04</td>
<td>4.61E-03</td>
<td>48</td>
<td>818</td>
<td>952</td>
<td>32395</td>
</tr>
<tr>
<td>ZIC1</td>
<td>2.13E-05</td>
<td>2.43E-04</td>
<td>4.61E-03</td>
<td>48</td>
<td>818</td>
<td>952</td>
<td>32395</td>
</tr>
<tr>
<td>ZIC2</td>
<td>4.32E-05</td>
<td>4.69E-04</td>
<td>9.38E-03</td>
<td>47</td>
<td>819</td>
<td>953</td>
<td>32394</td>
</tr>
<tr>
<td>AHR</td>
<td>8.60E-05</td>
<td>8.48E-04</td>
<td>1.87E-02</td>
<td>46</td>
<td>820</td>
<td>954</td>
<td>32393</td>
</tr>
<tr>
<td>USF</td>
<td>8.60E-05</td>
<td>8.48E-04</td>
<td>1.87E-02</td>
<td>46</td>
<td>820</td>
<td>954</td>
<td>32393</td>
</tr>
<tr>
<td>CACCCBF</td>
<td>1.67E-04</td>
<td>1.58E-03</td>
<td>3.63E-02</td>
<td>45</td>
<td>821</td>
<td>955</td>
<td>32392</td>
</tr>
<tr>
<td>HIF1</td>
<td>3.18E-04</td>
<td>2.76E-03</td>
<td>6.90E-02</td>
<td>44</td>
<td>822</td>
<td>956</td>
<td>32391</td>
</tr>
<tr>
<td>SMAD4</td>
<td>3.18E-04</td>
<td>2.76E-03</td>
<td>6.90E-02</td>
<td>44</td>
<td>822</td>
<td>956</td>
<td>32391</td>
</tr>
<tr>
<td>PAX3</td>
<td>1.07E-03</td>
<td>8.31E-03</td>
<td>2.33E-01</td>
<td>42</td>
<td>824</td>
<td>958</td>
<td>32389</td>
</tr>
<tr>
<td>NRF1</td>
<td>1.07E-03</td>
<td>8.31E-03</td>
<td>2.33E-01</td>
<td>42</td>
<td>824</td>
<td>958</td>
<td>32389</td>
</tr>
<tr>
<td>AP2a</td>
<td>1.07E-03</td>
<td>8.31E-03</td>
<td>2.33E-01</td>
<td>42</td>
<td>824</td>
<td>958</td>
<td>32389</td>
</tr>
<tr>
<td>USF</td>
<td>1.90E-03</td>
<td>1.42E-02</td>
<td>4.12E-01</td>
<td>41</td>
<td>825</td>
<td>959</td>
<td>32388</td>
</tr>
</tbody>
</table>
regulated during infection), the fact that many upregulated genes share the potential to be affected by it is of interest.
IV. DISCUSSION

We have presented here a global view of polyomavirus infection with respect to both the virus and the host transcriptomes. The data and analyses have allowed us to generate a detailed picture of how viral gene expression changes throughout the life cycle as well as to learn how infection impacts host cell gene expression. We analyzed RNA from a number of individual time points, from relatively early in infection (12 hours post infection) to late times (36 hours post infection) but before extensive cytopathic effects were noted. Although there were several surprises, the expression and processing of viral transcripts was generally consistent with previous literature. At early times in infection, early-strand transcripts dominated over late-strand transcripts. This early pattern is even more dramatic when DNA replication is blocked with aphidicolin and the infection allowed to proceed for 48 hours. At early times, most of these transcripts appear to initiate at a defined site and terminate just downstream of the annotated early polyadenylation signal. These are spliced to produce mRNAs for the three viral early proteins, large, middle and small T antigens. However, by 18 hours, after large T antigen has promoted the onset of viral DNA replication, the pattern of expression reveals a striking shift toward the late phase, where late-strand RNAs greatly dominate over early-strand RNAs. This early-to-late switch is characterized by several distinct and important changes in the pattern of alignments in our data. First, as late messages accumulate rapidly and genomes replicate, we noted an unexpected shift in the position of the 5’-ends of early-strand transcripts. At late times, early-strand RNAs appear to initiate over a broad region upstream of the early TATA box, and throughout the origin of replication.
We do not yet know why this is the case or how it is regulated, but these findings are consistent with an earlier report in which the 5’-ends were mapped using nuclease S1 protection (Kamen 1982). Interestingly, this switch depends on viral DNA replication, since after blocking it with aphidicolin, these upstream 5’-ends were not seen, even after 48 hours.

Another clear change in early-strand expression, and also dependent on DNA replication, was the appearance and accumulation of reads mapping to a region several hundred nucleotides downstream of the polyadenylation site. We do not know the genesis of these reads, and they do not appear to be the result of RNA splicing or a discrete RNA processing event. Further studies are needed to determine whether these represent discrete RNA species or regions of early-strand transcripts that have extended past the polyadenylation site.

Late-strand transcription and processing also changes during infection. While accumulating to only low levels before viral replication, or in its absence, these RNAs accumulate rapidly afterwards, and by late times represent the great majority of viral RNAs in the cell. Their processing may also change with time, although we did not note any discernible change in transcription start sites. As others have reported before, late transcription can proceed many times around the circular viral genome, leading to a heterogeneous collection of giant transcripts that can serve as precursors to mature late mRNAs (Acheson 1971, Aloni 1973, Heiser 1982, Adami 1989, Gu 2009). Processing involves not only alternative splicing to generate messages for VP1, VP2 and VP3, but also the splicing of noncoding late leader exons at their 5’-ends. The number of tandem leader units appearing on late mRNAs reflects the number of transits around the genome.
made by RNA polymerase II before transcription termination and polyadenylation. We noted a striking accumulation of leader-leader splicing events at late times in infection. These were only seen at late times, and after DNA replication.

Late leader splicing requires inefficient late-strand transcription termination and polyadenylation. We have reported previously that this is the result of particular viral genomic features in the early/late polyadenylation region (Gu 2009). These sites overlap on the genome and this affords viral transcripts the opportunity to anneal with one another, leading to dsRNA in the nucleus. dsRNA formation could alter normal 3’-end processing by interfering with the polyadenylation machinery. Also, dsRNA formation could lead to promiscuous editing by ADAR1, leading to RNAs that cannot be recognized by the processing machinery. We have previously found that viruses that cannot generate early-strand and late-strand transcripts that overlap are incapable of the early-late switch (Gu 2009). The consequences of such dsRNA formation are of profound importance for the virus. Inefficient late-strand termination leads to the multigenomic transcripts that serve as precursors to late mRNAs (Acheson 1971, Aloni 1973, Heiser 1982, Adami 1989). Early-strand termination is likely also to become less efficient, but in this case the multigenomic transcripts are not used for mRNA production, but are rather likely degraded in the nucleus owing to unproductive splicing. Our data support these concepts. While our transcriptomic data reveal mostly the accumulated mature mRNAs at late times, close inspection reveals roughly equal levels of poly(A) site readthrough transcripts on each strand. These would be expected to be capable of dsRNA formation, leading to ADAR editing. Indeed, A-to-G mismatches consistent with A-to-I editing were observed to only occur at late times in cells untreated with replication
inhibitor, with robust editing observed in the overlapping polyadenylation sites. The hyperediting pipeline we employed allowed us for the first time to capture polyoma reads that would otherwise be too heavily altered to be identified. The increase in edited clusters aligning along the entire length of the reference genome was predicted but, to our knowledge, has never been directly observed until now. Taken together, these observations confirm previous models of the early-late switch that describe a loss of polyadenylation efficiency concomitant with an increase in editing of the polyadenylation sites resulting in giant multigenomic RNAs, which can be spliced to generate multiple late leader exons on late mRNAs.

Alignment of reads from the time course infection to the host genome showed the greatest change between mock and infected samples by 36 hours, allowing us to observe the accumulated effects of the virus on the host transcriptome. We suspect later time points might yield more genes changing more than 2 fold up or down but time points closer to 72 hours would be entering the lysis phase of the viral life cycle and any observed changes would likely be difficult to separate from those associated with host cell death. At 36 hours post infection, we found 359 genes significantly upregulated and 857 genes significantly downregulated. Many of the genes upregulated were involved in predicted processes such as metabolism, ribosome synthesis, mitochondrial proteins, and cell cycle progression. This is consistent with the action of viral T antigens in stimulating the proliferation of host cells. Many downregulated genes were also largely unsurprising with those involved in cell motility, extracellular matrix, and cell signaling processes prominent on the list. Finally, we noted the upregulation or downregulation of several abundant long noncoding RNAs. There are several lncRNAs of particular interest. One
is Malat1 (downregulated), which has been associated with metastasis and which may play a role in RNA processing (Ji 2003). Another is Terc (upregulated), which is involved in telomere maintenance and cellular lifespan (Artandi 2010). Finally, several lncRNAs associated with RNase P enzymes are also strongly upregulated. We do not yet know the significance of this, but such molecules have been implicated in tRNA and ribosomal biogenesis, which is consistent with the observed increase in host translation in infected samples (Lai 2010).

Here we have presented for the first time a global analysis of both viral and host transcriptional changes from a mouse polyomavirus infection. The viral transcriptome analysis revealed insights into changes in RNA expression, especially with respect to the early strand at late times and confirmed a number of previous observations of the polyoma lifecycle. This is the most direct evidence of widespread hyper-editing of viral RNA predicted from a model of extensive early and late strand transcription. The observations of changes in host transcription also confirm a number of predictions including the increase in genes involved in cell growth, metabolism, and translation. Future investigation into more precise aspects of the lifecycle including later time points, nuclear and cytoplasmic fractionation, and PolyA plus and minus samples may yield further insight into the viral-host interactions.

V. FUTURE DIRECTIONS
The observations made by this investigation into the polyomavirus infection offer some of the most comprehensive insight into the changes in both viral and host transcription. Despite this, a number of questions still remain. What kind of differences would be observed in nuclear vs cytoplasmic or polyA+ vs polyA- RNA from mock and infected samples? We might expect to see more defined viral alignments consistent with mature mRNA in the cytoplasmic fractions and more readthrough from the giant transcripts in the nuclear fractions at late times. PolyA- selection might also reveal more noncoding RNAs in the host. What changes in splicing or editing clusters of host genes are observed between mock and infected samples at early and late times? Perhaps some transcripts that are regulated by processes dependent on the presence of double stranded RNA are affected by virus infection. Do the changes in transcription also produce the same changes in translation? We know of a number of post-translational modifications that are associated with the signaling pathways activated by viral tumor antigens but a more global profile of those changes has not yet been pursued. What is the precise nature of the shift in the 5’ ends of early transcripts at late times? Is it the result of more subtle changes in viral chromatin states beyond initial hyperacetylation? How does DNA replication and the early late switch affect transcription factor binding? Does the process of DNA replication allow transcription factors to better access the late transcription start site contained within the A enhancer? What would happen if the A enhancer was moved to the early side of the origin or the Large T binding sites moved to the late? Is editing a byproduct of the fact that viral DNA forms double stranded RNA or is it necessary for loss of polyadenylation and subsequent production of giant transcripts? These are just a few of the questions that remain to be investigated.
VI. REFERENCES

Acheson NH, Buetti E, Scherrer K, Weil R: Transcription of the Polyoma Virus Genome: Synthesis and Cleavage of Giant Late Polyoma–Specific RNA. Proc Natl Acad Sci USA 1971, 68(9): 2231-2235


Acheson NH: Polyoma virus giant RNAs contain tandem repeats of the nucleotide sequence of the entire viral genome. Proc Natl Acad Sci USA 1978, 75(10): 4754-4758


Aggarwal B: Nuclear factor-kB: The enemy within. Cancer Cell 2004, 6: 203-208


Benjamin TL: Relative Target Sizes for the Inactivation of the Transforming and Reproductive Abilities of Polyoma Virus. *Proc Natl Acad Sci USA* 1965, 54: 121-124

Benjamin TL: Virus-specific RNA in cells productively infected or transformed by polyoma virus. *J Mol Biol* 1966, 16(2): 359-373


Carmichael GG, Schaffhausen BS, Dorsky DI, Oliver DB, Benjamin TL: Carboxy terminus of polyoma middle-sized tumor antigen is required for attachment to membranes, associated protein kinase activities, and cell transformation. *Proc Natl Acad Sci USA* 1982, 79:3579-3583


Chen L, Fluck MM: Role of Middle T-Small T in the Lytic Cycle of Polyomavirus: Control of the Early-to-Late Transcriptional Switch and Viral DNA Replication. *J Virol* 2001, 75: 8380-8389


Chen MH, Benjamin TL: Roles of N-Glycans with alpha 2,6, as well as alpha2,3 Linked Sialic Acid in Infeciton by Polyoma Virus. *Virology* 1997, 233: 440-442


Cogen B, Eckhart W: BALB/3T3 Cells Infected by the ts3 Mutant of Polyoma Virus Fail to Accumulate Virus-Specific Early RNA at the Nonpermissive Temperature. *J Virol* 1977, 24(2): 701-703


Courtneidge SA, Heber A: An 81 kd Protein Complexed with Middle T Antigen and pp60^{c-sr}: A Possible Phosphatidylinositol Kinase. *Cell* 1987, 50: 1031-1037


Crawford LV, Robbins AK, Nicklin PM: Location of the Origin and Terminus of Replication in Polyoma Virus DNA. *J gen Virol* 1974, 25: 133-142


Delos SE, Montross L, Moreland RB, Garcea RL: Expression of the Polyomavirus VP2 and VP3 Proteins in Insect Cells: Coexpression with the Major Capsid Protein VP1 Alters VP2/VP3 Subcellular localization. *Virology* 1993, 194: 393-398


Eckhart W: Complementation between Temperature-sensitive (ts) and Host Range Nontransforming (hr-t) Mutants of Polyoma Virus. *Virology* 1977, 77: 589-597


Eckhart W, Dulbecco R, Burger MM: Temperature-Dependent Surface Changes in Cells Infected or Transformed by a Thermosensitive Mutant of Polyoma Virus. *Proc Natl Acad Sci USA* 1971, 68(2): 283-286 (B)


Flavell AJ, Cowie A, Legon S, Kamen R: Multiple 5’ Terminal Cap Structures in Late Polyoma Virus RNA. *Cell* 1979, 16: 357-371


Fujimura FK, Linney E: Polyoma mutants that productively infect F9 embryonal carcinoma cells do not rescue wild-type polyoma in F9 cells. *Proc Natl Acad Sci USA* 1982, 79: 1479-1483


Hattori J, Carmichael GG, Benjamin TL: DNA Sequence Alterations in Hr-t Deletion Mutants of Polyoma Virus. *Cell* 1979, 16: 505-513


Huerfano S, Zila V, Boura E, Spanielova H, Stokrova J, Forstova J: Minor capsid proteins of mouse polyomavirus are inducers of apoptosis when produced individually but are only moderate contributors to cell death during the late phase of viral infection. *FEBS J* 2010, 277(5): 1270-1283


Hunter T, Hutchinson MA, Eckhart W: Polyoma middle-sized T antigen can be phosphorylated on tyrosine at multiple sites in vitro. *EMBO J* 1984, 3(1): 73-79


Hyde-DeRuyscher R, Carmichael GG: Polyomavirus early-late switch is not regulated at the level of transcription initiation and is associated with changes in RNA processing. *Proc Natl Acad Sci USA* 1988, 85: 8993-8997


Katinka M, Yaniv M: Expression of Polyoma Early Functions in Mouse Embryonal Carcinoma Cells Depends on Sequence Rearrangements in the Beginning of the Late Region. *Cell* 1980, 20: 393-399


Kumar M, Carmichael GG: Nuclear antisense RNA induces extensive adenosine modifications and nuclear retention of target transcripts. *Proc Natl Acad Sci USA* 1997, 94: 3542-3547


Lane DP, Crawford LV: T-antigen is bound to a host protein in SV40-transformed cells. *Nature* 1979, 278: 261-263


Legon S: The Binding of Ribosomes to Polyoma Virus RNA: Possible Role of the Late Leader Region in Initiation Site Recognition. J Mol Biol 1979, 134: 219-240 (A)

Legon S, Flavell AJ, Cowie A, Kamen R: Amplification in the Leader Sequence of Late Polyoma Virus mRNAs. Cell 1979, 16: 373-388 (B)


Ntambi M: Regulation of steroyl-CoA desaturase by polyunsaturated fatty acids and cholesterol. *J Lipid Res* 1999, 40(9): 1549-1558


Pallas DC, Sharik LK, Martin BL, Jaspers S, Miller TB, Brautigan DL, Roberts TM: Polyoma Small and Middle T Antigens and SV40 Small t Antigen Form Stable Complexes with Protein Phosphatase 2A. *Cell* 1990, 60: 167-176


Piette J, Yaniv M: Two different factors bind to the α-domain of the polyoma virus enhancer, one of which also interacts with the SV40 and c-fos enhancers. *EMBO J* 1987, 6:1331-1337


Schaffhausen BS, Benjamin TL: Deficiency in histone acetylation in nontransforming host range mutants of polyoma virus. *Proc Natl Acad Sci USA* 1976, 73(4): 1092-1096


Soeda E, Arrand JR, Smolar N, Griffin BE: Sequence from Early Region of Polyoma Virus DNA Containing Viral Replication Origin and Encoding Small, Middle, and (Part of) Large T Antigens. *Cell* 1979, 17: 357-370


Tsai B, Gilbert JM, Stehle T, Lencer W, Benjamin TL, Rapoport TA: Gangliosides are receptors for murine polyoma virus and SV40. *EMBO J* 2003, 22(17): 4346-4355


Weil R: The Denaturation and The Renaturation of the DNA of Polyoma Virus. *Proc Natl Acad Sci USA* 1963, 49(4): 480-487 (A)

Weil R, Vinograd J: The Cyclic Helix and Cyclic Coil Forms of Polyoma Viral DNA. *Proc Natl Acad Sci USA* 1963, 50: 730-738 (B)


Winocour E, Sachs L: A Plaque Assay for the Polyoma Virus. *Virology* 1959, 8: 397-400


Zullo J, Stiles CD, Garcea RL: Regulation of c-myc and c-fos mRNA levels by polyomavirus: Distinct roles for the capsid protein VP1 and the viral early proteins. *Proc Natl Acad Sci USA* 1987, 84: 1210-1214