

**Single Cell High Content Microscopy Analysis of
BK Polyomavirus Infection**

by

Megan Christine Procario

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Microbiology and Immunology)
in the University of Michigan
2023

Doctoral Committee:

Professor Michael J. Imperiale, Chair
Professor Bethany B. Moore
Assistant Professor Jonathan Z. Sexton
Associate Professor Andrew W. Tai
Associate Professor Jason B. Weinberg
Associate Professor Christiane E. Wobus

Megan C. Procario

mcprocar@umich.edu

ORCID iD: 0000-0002-2501-2548

© Megan C. Procario 2023

Dedication

To Thành and Khôi.

Acknowledgements

I would like to start by thanking my committee, Beth Moore, Andrew Tai, Christiane Wobus, Jason Weinberg, and Jonny Sexton. Their patience, support, guidance, and flexibility has been invaluable to me during this process. Mary O’Riordan has been instrumental in my success, and for that I cannot thank her enough. I would like to thank Beth both as a mentor and a role model. When Beth became chair of the department, she never became less supportive or available. I am grateful to Jason, the scientist who first introduced me to virology. Jason taught me how to be a virologist for six years while I worked as a technician before starting graduate school and is the reason I fell in love with the field. My dissertation wouldn’t be possible without Jonny and his microscopy evangelization. I had no idea that attending a single seminar about cell painting would lead to such a successful collaboration. I am grateful to Jonny for his help and unending enthusiasm about the project. And I am forever indebted to my Ph.D. advisor, Mike Imperiale, who welcomed me into his lab when I needed to find a new home. I am especially grateful to Mike for his guidance in scientific writing.

My graduate school experience would not have been the same without support from others in lab. I want to especially thank Gau Shoua Vue, a talented technician and wonderful friend, as well as the other members of the Imperiale lab who I have worked with over the years. I am very appreciative of the entire Sexton lab, who have always made time to help me in a pinch. A special thanks to Ben Halligan for his help with anything that required coding and to Jesse Wotring for his unfailing moral support.

The experience and wisdom from others currently or previously in academia have been a great source of strength. I greatly appreciate the help seeing the light at the end of the tunnel, especially from Mary McCarthy and Kayla Zochowski, two brilliant scientists I have had the good fortune to work with over the years. Additionally, graduate school would not have been possible if it weren't for M.T. and K.H. Their investment in me and my well-being has been pivotal to my success. I cannot thank them enough for their support.

Finally, I would like to thank my family and friends. I am grateful to my parents for being the antidote to imposter syndrome and believing in me when I didn't believe in myself. I also thank my sisters, Erin and Sarah. Erin, for her writing tips, overall moral support, and providing the most perfect nephews in the world. And Sarah, for showing up whenever I've needed her. And I want to express my gratitude to my long distance 'chosen family', Anne and Juli who have spent countless hours on the phone with me during this process. Many thanks to the Gambers and the Wicks who have both opened their hearts and homes to me and made sure a hungry grad student was fed. I am incredibly fortunate to have too many supportive friends to be able to name them all here, so I will simply thank my SMSP friends, my CBE family, and everyone in between. This experience has been challenging and rewarding, and I couldn't have done it without everyone's help.

Table of Contents

Dedication	ii
Acknowledgements	iii
List of Tables	viii
List of Figures	ix
Abstract	x
Chapter 1 Introduction	1
1.1 BKPyV	1
1.1.1 Discovery	1
1.1.2 Epidemiology	5
1.1.3 Persistence and Disease	6
1.1.4 Structure	9
1.1.5 Genome	9
1.1.6 Life Cycle	12
1.1.7 Nuclear Transactions	18
1.2 PML-NBs	20
1.3 Studying virology <i>in vitro</i>	24
1.4 Goals of this dissertation	25
1.5 Bibliography	28
Chapter 2 Single Cell High Content Microscopy Analysis of BK Polyomavirus Infection	36
2.1 Abstract	36
2.2 Importance	37

2.3 Introduction	38
2.4 Results	41
2.4.1 Assay overview and imaging examples	42
2.4.2 Single cell microscopy as a method for studying BKPvV infection.	44
2.4.3 High DNA content, not time post-infection, distinguishes cells expressing the highest levels of TAg	46
2.4.4 TAg ⁺ cells form clusters defined by features other than timepoint and TAg level	49
2.4.5 Unique subpopulations exist among cells within the same TAg expression levels.....	51
2.4.6 PML-NB characteristics account for differences between TAg expression groups	55
2.5 Discussion	57
2.6 Materials and Methods	63
2.6.1 Cell culture	63
2.6.2 Virus growth and titration	63
2.6.3 Infection and Immunofluorescent staining.....	64
2.6.4 High-Content Imaging.....	65
2.6.5 Image Identification, Feature Extraction, Infection Classification, and Curve Fitting	66
2.6.6 Data processing and UMAP embedding	67
2.6.7 Statistical Testing	67
2.7 Acknowledgements	68
2.8 Tables	69
2.9 Supplemental Figures	75
2.10 Bibliography	78
Chapter 3 Discussion	83

3.1 Bibliography 99

List of Tables

Table 1.1 Human Polyomaviruses	27
Table 2.1 Antibodies used in this study	69
Table 2.2 Major defining features of TAg ⁺ clusters	70
Supplemental Table 2.3 Nuclear measurements obtained by Cell Profiler software*	71
Supplemental Table 2.4 Nuclear measurements obtained by Cell Profiler software [#]	72
Supplemental Table 2.5 PML-NB measurements obtained by Cell Profiler software	73
Supplemental Table 2.6 Percentages of cells at the different time points in each global cluster .	74
Supplemental Table 2.7 Percentages of cells at the different time points in each TAg ⁺ cluster ..	74

List of Figures

Figure 1.1 BKPyV Genome Diagram.....	14
Figure 1.2 PML isoforms.....	21
Figure 2.1 Assay overview and imaging examples.	43
Figure 2.2 Single cell microscopy as a method for studying BKPyV infection.	45
Figure 2.3 High DNA content, not time post-infection, distinguishes cells expressing the highest levels of TAg.	48
Figure 2.4 Appearance of cells with areas of mutual DNA/TAg exclusion.	50
Figure 2.5 TAg ⁺ cells form clusters defined by features other than timepoint and TAg level.	52
Figure 2.6 Unique subpopulations exist among cells within the same TAg expression levels. ...	54
Figure 2.7 PML-NB characteristics account for differences between TAg expression groups....	56
Supplemental Figure 2.8 Cluster C contains more cells at 2 dpi.	75
Supplemental Figure 2.9 Distribution of TAg-positive cells in the clusters.....	76
Supplemental Figure 2.10 Clusters in the high TAg group differ in multiple ways including nuclear area and PML-NB quantity.	77

Abstract

BK polyomavirus (BKPyV) is a non-enveloped, double-stranded DNA virus with a circular genome. By adulthood, the majority of the population is persistently infected with BKPyV. While healthy individuals will not experience disease from BKPyV, people who have received transplants and are on immunosuppressive drugs are at risk. Kidney transplant patients can develop a condition called polyomavirus-associated nephropathy (PVAN), which often requires a reduction in immunosuppressive medication to treat. This puts patients at risk of rejecting the donor kidney. Patients who have undergone a hematopoietic stem cell transplant are at risk for a painful complication that causes bleeding and inflammation of the bladder, hemorrhagic cystitis (HC). Treatment for HC is often supportive and can require blood transfusions. To date there are no effective antivirals to treat or approved vaccines to prevent BKPyV. Most studies of BKPyV have been performed on bulk populations of cells and the dynamics of infection at single-cell resolution have not been explored. As a result, much of our knowledge is based upon the assumption that all cells within a greater population are behaving the same way with respect to infection. This dissertation examines BKPyV infection on a single-cell level using high content microscopy to measure and analyze the viral protein large T antigen (TAg), promyelocytic leukemia protein (PML), DNA, and nuclear morphological features. We observed significant heterogeneity among infected cells, within and across timepoints. We found that the levels of TAg within individual cells did not necessarily increase with time, and that cells with the same TAg levels varied in other ways. These results are discussed in the context of broader field of BKPyV infection studies, including a BKPyV single-cell RNAseq timecourse analysis. By

combining the knowledge established from prior work in bulk cell culture with the insights from single-cell experiments today we propose a more comprehensive view of BKPyV infection.

Chapter 1

Introduction

1.1 BKPyV

1.1.1 Discovery

Polyomaviruses (PyVs) are a family of small, non-enveloped viruses with circular double-stranded DNA genomes. PyVs can infect a range of different hosts, including fish, birds, and mammals. For a given virus, however, that host range is limited. They are known to cause outbreaks of severe disease within avian communities, but don't generally cause illness in other healthy hosts. The earliest discoveries of PyVs were largely those that infect laboratory animals and humans, however as scientific technology has improved this has expanded greatly to include many more species.

The first polyomavirus (PyV) to be discovered is what is now known as mouse polyomavirus or MPyV ¹. It was originally identified by Gross when he was performing serial passages of what he believed to be murine leukemia virus (MLV) through mice and he identified parotid tumors in some of the mice. Gross was able to isolate virus from these tumors and determine that its characteristics were different from those of MLV. MPyV received the name *polyomavirus*, meaning many (poly) tumors (oma) -virus, when Stewart used it to infect neonatal mice and found that it caused multiple tumors ². It was around this same time that Kilham, or K, virus was discovered in suckling mice ³. One of the next PyVs to be identified was Simian Virus 40, or SV40, named vacuolating virus by the authors because of the cytopathic effect it caused in

monkey kidney cells upon infection ⁴. SV40 was discovered when screening poliovirus vaccines, which were grown on African green monkey kidney cells, for viral contaminants. SV40 is resistant to the formalin treatment used for the poliovirus vaccine, which allowed live virus to survive through to the final product. This was concerning because SV40 was determined to be oncogenic in tissue culture and animal models within the next few years ^{5,6}. Fortunately, there is no evidence to date supporting a role for SV40 as a human pathogen or a cause of human cancer ⁷. Shortly after SV40, Rabbit Kidney Vacuolating (RKV) Virus was identified while studying papilloma samples from cottontail rabbits ⁸. The authors were able to determine that RKV was distinct from Shope papillomavirus and was much more similar to SV40.

After these early PyV discoveries came the first two descriptions of Human polyomaviruses (HPyVs). To date there are thirteen identified HPyVs (see Table 1.1). The HPyVs fall into three of the four total genera for PyV ⁹. These three are Alphapolyomavirus, Betapolyomavirus, and Deltapolyomavirus, which all contain mammalian PyVs. The fourth genus is Gammapolyomavirus which is mostly comprised of avian PyVs. The first two HPyVs discoveries were published back-to-back, in the same issue of *The Lancet* in 1971. JC polyomavirus (JCPyV) was isolated from the brain of a patient with Hodgkin's disease and the demyelinating disease, progressive multifocal leukoencephalopathy ¹⁰. Prior to 1971, others had identified virions in brain tissue samples from patients with progressive multifocal leukoencephalopathy but were unable to culture any virus ¹¹. Padgett, et al were able to confirm the presence of SV40-like virions in brain sections, as well as propagate virus from brain tissue in primary human fetal glial cells. The authors confirmed that the agent they had identified was neither SV40 nor human papillomavirus (HPV), and named it using the initials of the patient from whom it had been isolated, J.C. The other virus identified in the early 1970s was BK

polyomavirus (BKPyV), which was also named using the initials of the patient from whom it was isolated ¹². BKPyV was isolated from the urine of a patient who had undergone a kidney transplant from a living donor. The patient presented with acute rejection, which resolved with treatment and was later followed by ureteric obstruction which was treated surgically. BKPyV virions were detected by electron microscopy in urine and ureter tissue samples from the patient. The authors were able to culture virus from the urine in monkey kidney and Vero cells. Low antibody titers were detectable in the sera from the donor and the patient both pre- and early post-transplant. The titers quickly rose later post-transplant, even when virions and culturable virus were no longer detectable in the urine. The authors confirmed that this novel virus was not SV40 by serological testing and that it wasn't HPV by both serological testing and measuring the size of the virions, as papillomaviruses are larger than PyVs. The authors speculated that the newly identified BKPyV was likely present in both the donor and the recipient based on their Ab titers and that ischemic changes in the ureter secondary to organ rejection facilitated viral replication. They postulated that another site of viral replication may be present because removal of the ureter did not halt appearance of virus in the urine. Additionally, the authors suggested that the immune suppression for the transplant may have supported viral replication, but that normal antibody responses occurred in spite of the suppression and disease eventually resolved.

Following JCPyV and BKPyV there was an extended period of time without any new HPyV discoveries until the late 2000s to early 2010s when next-generation sequencing became available. In 2007 two new PyVs were identified in pediatric nasopharyngeal samples. KI polyomavirus (KIPyV) was found during a screen of samples from patients with upper respiratory tract infections in Sweden ¹³. It was found by PCR in approximately 1% of samples and whole genome sequencing revealed limited homology to known polyomaviruses at that time.

The authors did not know the prevalence or pathogenicity of KIPyV. WU polyomavirus (WUPyV) was originally found in a single three-year-old Australian patient with pneumonia using high-throughput DNA sequencing ¹⁴. Following whole-genome sequencing, WUPyV was determined to be the most similar to KIPyV of all the currently identified PyVs. The authors subsequently identified 43 more cases across locations in Australia and the United States. In 2008, researchers at the University of Pittsburgh were searching for potential viral sequences in Merkel cell carcinoma (MCC), an aggressive form of skin cancer ¹⁵. They detected a fusion between a human gene transcript and previously unidentified viral gene transcript. This led to their identification of Merkel Cell polyomavirus (MCPyV). MCPyV sequences were detected in eight of ten MCC tumors, but only 4 of 25 control skin samples, suggesting a role for the virus in MCC. More persuasive, however, was the fact that in six of the eight MCPyV-positive tumors, the viral DNA was integrated into the genome in a monoclonal pattern. This led the authors to conclude that viral gene integration occurred prior to expansion of the tumor cells, which suggests that MCPyV is playing a causal role in development of MCC. This discovery of MCPyV led to interest in other potential human skin polyomaviruses found in healthy individuals. Work from the Buck lab found MCPyV in swabs from skin of healthy participants and also identified two novel polyomaviruses using the same techniques ¹⁶. These two new HPyVs, named HPyV6 and HPyV7, were found either alone or in various combinations with other PyVs. Since then, six more HPyVs have been identified. Many of them are associated with immunocompromised patients, like the Trichodysplasia spinulosa-associated polyomavirus (TSPyV) isolated from a patient who had received a heart transplant and HPyV9 which was isolated from a kidney transplant recipient's serum ^{17,18}. Malawi polyomavirus (MWPyV) was originally identified in a stool sample from a single healthy child, and then identified again in a

screen of stool samples from children with diarrhea ¹⁹. Saint Louis polyomavirus (STLPyV) was also identified in stool specimens, this time from the United States and Gambia ²⁰. Human polyomavirus 12 (HPyV12) was originally identified in gastrointestinal tissue from cancer patients ²¹. When healthy volunteers were screened for seropositivity, the authors found that HPyV12 seroprevalence started at around 12% in toddlers and increased up to 33% into adulthood. New Jersey polyomavirus (NJPyV) was identified in the endothelial cells of a muscle biopsy from a pancreatic transplant patient ²². Although discovered in a human, NJPyV appeared to have the greatest homology with Chimpanzee polyomavirus, leading the authors to wonder about the source of the patient's exposure.

1.1.2 Epidemiology

The focus of this dissertation will be specifically on BK polyomavirus, so from here on, most of the information will be as well. The general seroprevalence for BKPyV in the healthy population increases with age, with maximum reaching as high as >95% positivity for adults ²³. Infection occurs during childhood. The route is potentially via respiratory infection, as suggested by findings of BKPyV in tonsils from an immunocompetent child, though a definite answer remains unclear ²⁴.

BKPyV is classified into four subtypes (I through IV) on the basis of serology and genotyping ^{25,26}. Subtypes I and IV have further been divided into subgroups a, b1, b2, c1, and c2 ²⁷. Subtypes I and IV are the most common, in that order, while subtypes II and III are seen far less frequently. Association of certain subtypes with specific populations and across geographic regions has led some to suggest that BKPyV has co-evolved with human migration ²⁸⁻³⁰. For

example, subtype I is found in Africa, Europe, and Asia, with potential early common ancestors in Africa and Asia. Subtype IV, however, is primarily isolated to Asia, with the exception of subgroup IVc2 which is present in Europe and the Americas.

1.1.3 Persistence and Disease

1.1.3.1 Persistence

Following initial infection, BKPyV persists throughout the life of an individual in the urinary tract. The virus is periodically shed in the urine of healthy (immunocompetent) hosts and shedding is more frequent in during pregnancy, likely due to the pregnant individuals' somewhat-immunocompromised state. The mechanism of BKPyV persistence is not fully understood, but there are pieces of the greater explanation. It is possible that BKPyV could establish a proper latency like the herpesviruses. Alternatively, BKPyV could be replicating at a very low level and viremia and/or viruria are prevented primarily by a T cell response, but this hypothesis remains unproven. The replication of the virus is tightly controlled through multiple mechanisms that will be discussed in greater detail later in this chapter. These include a weak early promoter, viral protein-protein interactions, and a viral miRNA.

1.1.3.2 Polyomavirus-associated nephropathy

When a patient has received a kidney transplant or a hematopoietic stem cell transplant (HSCT), they can become susceptible to BKPyV-induced disease. Kidney transplant patients are at risk of developing polyomavirus associated nephropathy (PVAN). BKPyVAN is defined by the presence of active BKPyV replication within renal tissue, BKPyV DNA in the blood, and

often, inflammation and virus-induced cytopathology in renal tubule cells ³¹. Left untreated, PVAN can lead to severe kidney dysfunction from viral damage. The best treatment is currently reduction of immunosuppression, which allows the patient's own immune system to control the virus, but significantly increases the risk of graft rejection.

1.1.3.3 Hemorrhagic cystitis

The myeloablative therapy that prepares HSCT patients to receive the donor cells leaves patients both immunocompromised and with damage to their urothelium. These patients are at risk for developing BKPyV-associated hemorrhagic cystitis (HC), a painful condition characterized by bleeding and inflammation of the lining of the bladder. HC occurs in 8-25% of pediatric and 7-54% of adult allogeneic HSCT recipients ³². As with PVAN, there are limited effective treatment options for HC.

1.1.3.4 Cancer

Patients who have received kidney transplants are at an increased risk for urinary tract cancers and this risk is further increased for those with PVAN ³³. In spite of this, a direct connection between BKPyV and cancer is elusive. BKPyV is oncogenic in hamsters and has been used to transform cell lines, including BHK21 and fetal human brain cells ³⁴⁻³⁸. However, this transformation only occurs in non-permissive cells, meaning those that cannot be productively infected with the virus. When BKPyV establishes a lytic infection in permissive cells, those cells ultimately die and aren't transformed. This lack of transforming ability in permissive cells is borne out in patient data as well. While tumors isolated from kidney transplant recipients a decade later have been positive for BKPyV DNA and capsid protein, it has

been incredibly rare to find BKPyV DNA integrated into the tumor cell genome ³⁹. A more indirect, “hit and run” model of oncogenesis has been proposed for BKPyV and bladder cancer ^{40,41}. In this model, BKPyV doesn’t directly transform cells, but rather it induces the expression and activation of the cytosine deaminase, APOBEC3B (A3B). Members of the APOBEC3 family primarily deaminate single-stranded DNA (ssDNA) and can act as antiviral restriction factors by catastrophically mutating viral genomes ⁴². APOBEC3B’s mutations do not appear to be harmful to BKPyV’s genome, however. BKPyV’s large tumor antigen (TA_g) and truncTA_g induce A3B and BKPyV’s genome displays signatures of prior mutations by the enzyme ^{43,44}. It has been proposed that BKPyV utilizes A3B to acquire escape mutants from neutralizing antibodies and to expand its cell surface glycan binding repertoire ⁴⁵. A study examining BKPyV DNA in serum and urine samples from patients who later developed renal carcinoma found that the majority of the mutations identified were consistent with an A3B signature. In an expanded cohort of kidney transplant patients, the A3B mutation signature was only seen in BKPyV from patients diagnosed with PVAN, not just BKPyV viremia. However, exome sequencing from the original tumors in the study did not indicate an increase in A3B mutation rates over those of renal tumors in the general population. In contrast, bladder cancers have been demonstrated to have high APOBEC3B expression and mutation frequency ⁴⁶. Baker, et al. have developed a differentiated urothelial cell model that provides strong support for lytic BKPyV infection as the ‘driver’ in the ‘hit and run’ model ⁴⁰. The authors demonstrate virus-induced cell cycle entry, DNA damage, homologous recombination, and APOBEC activity. An additional bladder cancer and mouse tumor model suggests that it may involve β -catenin signaling ⁴⁷.

1.1.4 Structure

BK polyomaviruses are approximately 45 nm non-enveloped particles comprised of two main components: a roughly 5 kilobasepair (Kb) long double-stranded circular DNA genome and a capsid. The BKPyV genome exists in the form of a minichromosome, which contains the DNA as well as cellular histones, H2A, H2B, H3, and H4. The genome also associates with histone H1 during infection of the cell, but not in the mature virion. The capsid is comprised of three proteins, the major capsid protein VP1 and minor capsid proteins VP2 and VP3. The capsid has T=7 icosahedral symmetry and is made up of 72 pentamers. Each pentamer contains five VP1 molecules and either a single VP2 or VP3 molecule in its center⁴⁸. The C-terminus of VP1 molecules from each pentamer interacts with the C-termini of neighboring pentamers, stabilizing the icosahedral structure. The capsid is further stabilized by calcium ions and disulfide bonds between the pentamers. Once assembled, the only exposed protein on the polyomavirus capsid is VP1.

1.1.5 Genome

The BKPyV genome can be divided into three main parts: the early region, the late region, and the non-coding control region (NCCR). The NCCR contains the origin of DNA replication as well as bidirectional promoters, one for each of the other two regions, and transcription factor (TF) binding sites. The early viral gene region (EVGR) encodes early viral proteins: T antigens (large, small, truncated, and Super). The late viral gene region (LVGR) encodes the three viral capsid proteins and agnoprotein.

The NCCR is divided into five sequence blocks, starting with the O block which contains the origin of replication and is flanked by the early gene region on one side and P block on the other. After P, blocks Q, R, and S follow in that order, with S being flanked by the late gene region on the other side. This arrangement of the NCCR defines the “archetype” version of BKPyV. It is the sequence found in healthy individuals and some patients and is considered to be the transmissible form of the virus. It also replicates very poorly in tissue culture. BKPyV with rearrangements in the NCCR, referred to as rearranged variants, maintain the same flanking O and S blocks, but have duplications and/or deletions of the P, Q, and R blocks. These rearranged variants are generally found in patients with severe disease and replicate robustly in tissue culture.

1.1.5.1 Archetype vs. Rearranged

The original strain of BKPyV that many researchers were working with was called “prototype” (or the Gardner strain, for Sylvia Gardner, who had originally isolated it) and was abbreviated as BK(pr). This was virus that had been isolated from a patient’s urine then amplified in Vero cells. At the time, it was considered to be the ‘wild-type’ version of BKPyV, as it was the first found and characterized, but this belief was soon challenged. SV40 was used as a guide for mapping BKPyV’s genome and for understanding potential homologous functions. When BKPyV isolates MM and JM, which were identical to the Gardner virus according to the methods available at the time, were compared to SV40, researchers initially found low nucleotide homology ⁴⁹. However, with improved technology this was proven to be an underestimate. A later study found that BKPyV MM and SV40 share as much as 73 percent amino acid homology and 70 percent nucleotide sequence homology ⁵⁰. This study also

identified differences in repeated regions of the viruses' non-coding control regions. Given the great similarity to SV40, researchers wanted to characterize BKPyV's early promoter to determine if it used a similar enhancer to that used by SV40 to drive transcription ⁵¹. They found that unlike SV40, BK appears to have an overlapping early promoter and enhancer. They also determined that the ability to form plaques is directly proportional to the number of 68-bp repeats (aka, P blocks) in the enhancer of a given BKPyV isolate ^{51,52}. Additionally, transcription of the early gene region is enhanced by transcription factor binding, specifically including Sp1 and NF1 ⁵³. As people continued to explore the differences in the NCCR between viruses, they identified these 68-bp repeats to be important determinants of viral growth in culture. They found that virus isolated directly from patients without propagation in culture first was challenging, if not impossible to grow ⁵⁴. When scientists attempted to characterize the enhancer/promoter region of one of these non-amplified viruses, specifically BK(wv), they found that it had half the early promoter strength of BK(pr) ⁵⁵. The authors found that BK(wv) had only a single copy of the 68-bp repeat found three times in BK(pr), but had an additional block of sequence inserted in its place, the R block ⁵⁶. The authors asserted that this was the archetypal form of BKPyV and that the prototype viruses and those like it were the result of recombinations. They supported this theory with a study in which they demonstrated serial passage of archetype virus in culture consistently led to changes in the NCCR along with increased cytopathology ⁵⁷.

Eventually, when a more standardized, physiologically relevant approach to studying BKPyV in the lab became established, researchers could examine the meaningful differences between archetype and rearranged variants ⁵⁸. This, however, proved to be quite challenging, as archetype virus was still remarkably difficult to amplify, especially without introducing mutations. It was

only upon supplying large amounts of SV40 T antigen *in trans* that archetype would propagate well ⁵⁹. Once a method for propagating archetype BKPyV was available, research could be done to better understand the differences between archetype and variants.

As mentioned earlier, our understanding that there even were two types of NCCR arrangements came into existence due to the heterogeneity of patient samples. This holds true when looking at clinical samples with the more sophisticated techniques of today. As a general rule, rearranged variants are isolated primarily from patients with PVAN and/or those with severe systemic disease. Samples obtained from urine in patients with HC or from healthy patients tend to contain archetype virus.

1.1.6 Life Cycle

1.1.6.1 Binding/Entry

BKPyV enters the cell by binding of the major capsid protein, VP1, to ganglioside receptors GT1b/GD1b, which triggers internalization of the virus ⁶⁰. This internalization step occurs via a caveolin- and clathrin-independent pathway ⁶¹. Following internalization, the virus traffics to the ER through an endolysosomal pathway, during which acidification occurs ⁶². Within the ER, the viral capsid continues to undergo conformational changes, exposing the minor capsid proteins and destabilizing the icosahedral shell. The virion is then transported back to the cytoplasm via the ERAD pathway prior to nuclear entry of the genome ⁶³. Nuclear localization sequences (NLS) found on the capsid proteins target the genome to importins, facilitating its transport through nuclear pore complexes (NPCs) ⁶⁴.

1.1.6.2 Transcription

BK polyomavirus gene transcription progresses via RNA polymerase II from the bidirectional promoters in the NCCR, with the early and late regions being transcribed in opposite directions (figure 1.1). Early mRNA is produced by post-transcriptional processing at a poly(A) site, located approximately halfway around the genome, and intron splicing. BKPyV encodes four proteins produced from alternatively spliced transcripts. These proteins are large T antigen (TAg), small t antigen (tAg), truncated T antigen (truncT), and Super T antigen (SuperT) ⁶⁵. Late transcription proceeds in the opposite direction to generate two mature mRNAs. Like early transcripts, these are also post-translationally modified at a poly(A) site (located approximately halfway around the genome) and alternatively spliced. The first mRNA encodes agnoprotein and VP1 and the second encodes VP2/VP3. VP3 is translated from the same open reading frame as VP2 but utilizes a downstream start codon ⁶⁶. Transcription of the LVGR continues beyond the 3' poly(A) site to produce a microRNA (miRNA) ⁶⁷. The miRNA is complementary to the early mRNAs, targeting them for degradation ⁶⁸. The miRNA can function to regulate TAg expression, and thus, viral replication.

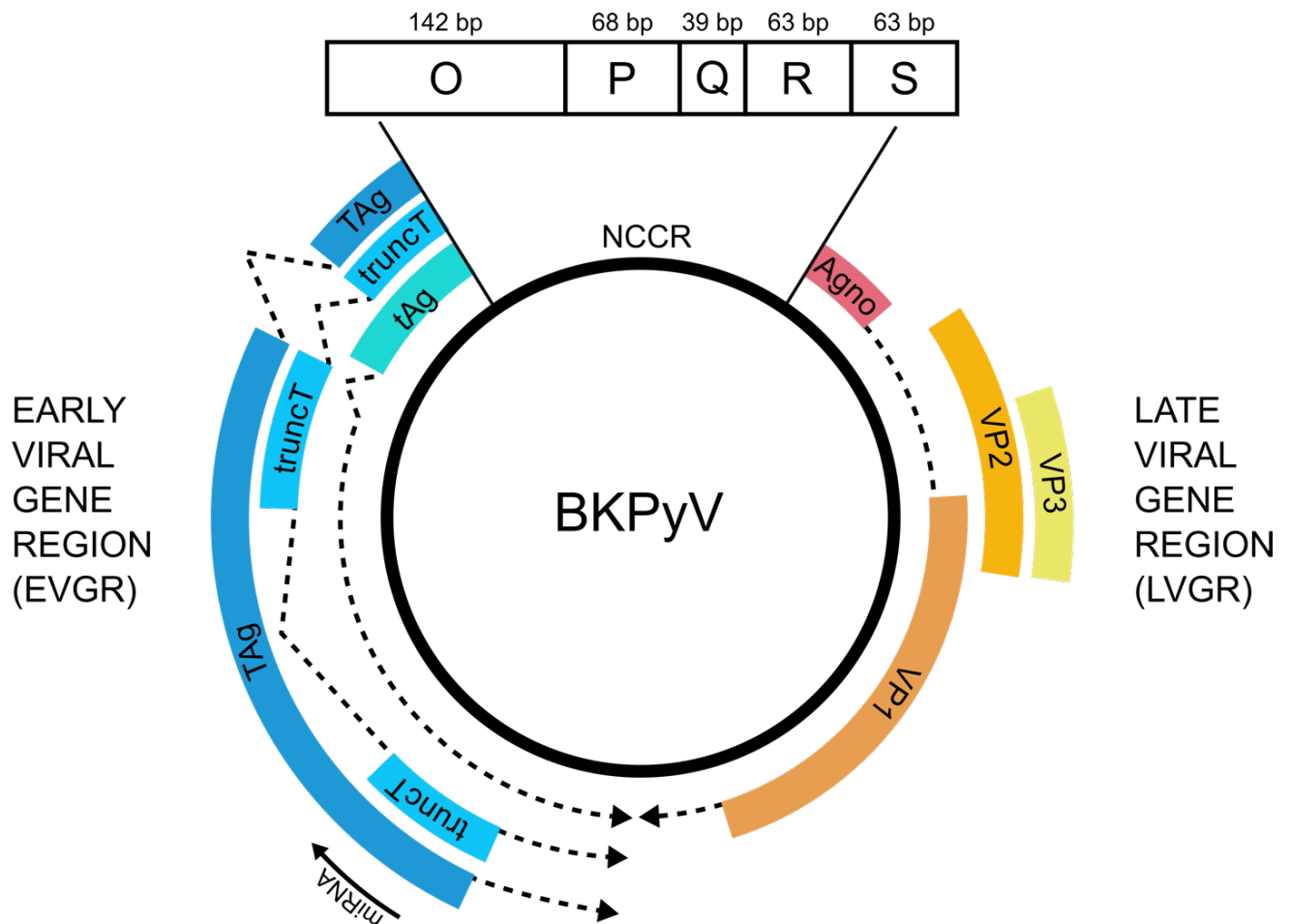


Figure 1.1 BKPyV Genome Diagram

The non-coding control region (NCCR) is magnified to show the sizes of the individual sequence blocks at the center of the genome. The coding sequences for the viral proteins are represented as solid bars, with the arrows indicating the direction of transcription. The miRNA is drawn to indicate roughly where it complements the early viral transcripts.

1.1.6.3 Early proteins

Large tumor antigen (TA_g) is a 695 residue-long nuclear phosphoprotein that plays multiple key roles in the polyomavirus life cycle, including recruiting host factors for viral replication, regulating viral replication, and antagonizing tumor suppressor proteins. It can be post-translationally modified by O-glycosylation, acylation, poly(ADP)-ribosylation, and acetylation, in addition to phosphorylation. TA_g performs its many duties via its multiple domains, from its

N-terminal J domain (named for its homology to the HSP40 DnaJ domain) to the C-terminal host-range domain. Much of what we know about these motifs comes from studies of SV40. The TAg J-domain is known to bind to and activate the helicase activity of HSC70, while its DNA-binding domain (DBD) binds to replication protein A (RPA). TAg acts as an initiation factor for DNA synthesis. Its helicase domain binds to the DNA polymerase alpha/primase complex, recruiting it to the site of viral DNA replication. TAg uses its helicase function to unwind the DNA at the origin of replication to allow for initiation of DNA synthesis. The helicase domain also binds to p53, sequestering the tumor suppressor protein and preventing it from inducing apoptosis or preventing entry into S phase. Additional TAg functional motifs include an NLS, which is required for efficient replication, and an LXCXE motif that binds directly to Rb family proteins, which in turn facilitates the de-repression of E2F transcription factors needed for G₁/S transition. SuperT is the product of wraparound transcription, resulting in a protein that is larger than TAg and contains two LXCXE motifs ⁶⁵.

Small tumor antigen (tAg) is a cysteine rich 145 amino acid protein. It has the same 82 N-terminal amino acids as TAg, including all of TAg's first exon (and therefore its N-terminal J domain), but has a different C-terminus. This C-terminus contains a highly conserved region that is rich in cysteine and histidine residues, which allows for binding to two zinc molecules. This zinc binding domain is key to tAg's ability to bind the host protein phosphatase 2A (PP2A). PP2A plays important roles in regulation of cellular processes, including cell proliferation and preventing cellular transformation. The PP2A is a complex comprised of three main subunits, each with a different role and two or more isoforms. SV40's tAg binds to subunit A, which functions as a scaffold for PP2A, and this displaces the regulatory B subunit, but leaves the catalytic C subunit in the complex ⁶⁹. BKPyV's tAg also interacts with PP2A subunits A and C

and has recently been shown to interact with the B subunit members striatin (STRN) and striatin 3 (STRN3)⁷⁰. BKPyV's tAg inhibits TAg gene expression and viral DNA replication, in part through tAg's interactions with PP2A. Importantly, the inhibition of DNA replication is only seen in archetype virus. Although tAg binding to PP2A disrupts the primary cellular function of the protein, PP2A may retain some activity.

1.1.6.4 miRNA

BKPyV can regulate its early gene products post-transcriptionally, via viral miRNAs⁷¹. Specifically, it can regulate expression of TAg and tAg via two miRNAs that are perfectly complementary to the early gene region. These miRNAs are encoded beyond the 3' poly(A) site of the late gene region. The amount of miRNA produced by BKPyV is dependent on the NCCR, because it is thought to be transcribed from the late promoter. This means that archetype virus, which has a relatively weaker early promoter and stronger late promoter, produces more miRNA and as a result, further decreases the low amount of TAg it's already produced. And in turn, rearranged virus, which has a strong early promoter and weaker late promoter, makes far more TAg mRNA than miRNA. In fact, it was originally believed that BKPyV's miRNA didn't have a function because it was only studied in the context of rearranged BKPyV prior to the work of Broekema, et al^{67,71}.

1.1.6.5 DNA replication

Replication of the viral genome is orchestrated, in large part, by TAg. TAg binds to the origin of replication in the viral genome, where it uses its helicase activity to unwind the double-stranded DNA (dsDNA) to facilitate replication. TAg also recruits the cellular DNA replication

machinery including RPA, DNA polymerase alpha/primase, and topoisomerase I and II. In addition to initiating replication, TAg is also required for elongation.

1.1.6.6 Assembly/egress

Assembly is initiated by transport of capsid proteins VP1, 2, and 3 from the cytosol into the nucleus, where they form capsomeres. Although all three capsid proteins contain an NLS, VP2 and 3 with a mutation in the NLS will still accumulate in the nucleus if provided wild-type VP1⁶⁴. This suggests that the proteins enter as a complex, rather than individually. Purified recombinant VP1 can even self-assemble into capsomeres and form empty capsids, demonstrating that VP2/3 and viral DNA are unnecessary for capsid formation. Recombinant purified VP1 protein is used create capsomeres for use in serology testing⁷². VP1 proteins from three BKPyV genotypes and one from JCPyV have been made into virus-like particles for use as a potential multivalent polyomavirus vaccine⁷³. It is not currently known if the capsid assembles around a nascent minichromosome or if it assembles first, then the genome is incorporated.

The mechanism used by BKPyV to exit cells is not clear. While rearranged variants of BKPyV induce obvious cytopathic effect in tissue culture, no mechanism for lysis has been defined. There are some studies implicating agnoprotein, a 66 amino acid primarily cytoplasmic phosphoprotein. Numerous hypotheses about the role(s) of polyomavirus agnoproteins have been made, but only a few have been confirmed in multiple studies. The interaction between agnoprotein and a membrane fusion protein called α -soluble N-ethylmaleimide-sensitive fusion (NSF) attachment protein (α -SNAP) is one of the latter^{74,75}. Some work has suggested that this interaction with α -SNAP is essential for nuclear egress into the cytoplasm, however this still doesn't explain cellular egress.⁷⁵ Other work suggests a non-lytic route of egress from cells

using a late endosomal pathway ⁷⁶. There are clearly still many gaps in knowledge in this part of the field.

1.1.7 Nuclear Transactions

1.1.7.1 S phase entry

Upon infecting a cell, BKPyV requires use of the host DNA replication machinery, which is only available during S phase. As the virus usually infects cells that are either quiescent (in G₀) or in G₁, it must induce S phase entry itself. To facilitate this, TAg's LXCXE motif binds RB-E2F complexes, which releases RB-mediated repression of E2F. Because RB-E2F complexes repress genes needed for G₁/S transition and DNA synthesis, removal of the repression induces entry into S phase. TAg plays a further role to support entry into S phase and staying there by binding p53. The act of forcing entry into S phase triggers p53, which could in turn induce growth arrest or apoptosis ⁷⁷. By sequestering p53, TAg prevents it from activating its target genes. One of those target genes is p21, a cyclin kinase inhibitor (CKI) that ultimately increases RB-E2F complex formation. In this way, TAg's prevention of p53 activity helps maintain the cell in S phase.

1.1.7.2 G₂/M arrest and DDR

Following entry into S phase, BKPyV-infected cells continue on in a state of dysregulation. Multiple rounds of DNA replication can occur within a single cycle in each cell. This leads to infected cells with >G₂ DNA content, but that do not progress to mitosis ⁷⁸. These cells instead remain arrested at the G₂/M checkpoint.

One regulator of growth arrest at the G₂/M checkpoint is Chk1, which is phosphorylated by part of the DNA Damage Response (DDR). The DDR is comprised of two signaling arms, named for their kinases: ataxia telangiectasia mutated (ATM), and ATM and Rad3-related (ATR). The ATM kinase detects dsDNA breaks using the Mre11-Rad50-NBS1 complex (MRN) while the ATR kinase responds to ssDNA breaks. Both phosphorylate substrates, including Chk1, Chk2, and γ H2AX. It is known from studies of SV40 that the process of viral genome replication induces activation of the DDR which mediates both repair of replication-induced double-stranded breaks (DSB)s and maintenance of replication fork integrity ^{79,80}. In the context of G₂/M arrest, phosphorylation of Chk1 facilitates pChk1 to phosphorylate the phosphatase Cdc25 and target it for degradation. This maintains inhibition of the cyclin B-Cdc2 complex by Wee1 that would otherwise signal cells to enter mitosis ^{81,82}.

BKPyV infection activates both the ATM and ATR signaling pathways, which both contribute to the virus' replication and growth ⁷⁸. Active viral DNA replication is required for this induction of the DDR ⁸³. Specifically, BKPyV's activation of ATM induces Chk2 phosphorylation and activation of ATR induces phosphorylation of Chk1 and γ H2AX. The ATR pathway is activated, at least in part, via a component of the mismatch repair (MMR) pathway ⁸⁴. MMR is an excision repair pathway known to activate ATR. The protein complex MutS α is half of a DNA damage sensor that initiates the MMR pathway. MutS α is required for robust activation ATR and subsequently, G₂/M arrest. The roles of ATM and ATR during BKPyV infection have a temporal component. ATM is important early during infection for entry into S-phase, while ATR's role is to prevent entry into mitosis later in the infection ⁸². Overall, this prolongs the time spent in S-phase which allows for optimal viral DNA replication.

1.2 PML-NBs

An important host factor that has been explored in the context of HPyVs is promyelocytic leukemia (PML) protein. PML, also known as TRIM19, is a member of the TRIM (tripartite motif) family of proteins. Members of the TRIM family share a common RBCC motif in their N-terminus, with unique C-termini. This RBCC motif is comprised of a RING-finger domain, two B-boxes, and a coiled-coil domain.

The *PML* gene was identified because of a chromosomal translocation t(15;17) found in acute promyelocytic leukemia (APL) which leads to expression of the fusion oncoprotein PML-RAR α ⁸⁵. The *PML* gene contains 9 exons and encodes at least seven transcriptional isoforms. These isoforms all share at least the same first three exons, which include the RBCC motif, with their C-termini differing to various extents (figure 1.2). For example, while isoforms I through VI all contain an NLS, isoform VII does not, and is therefore restricted to the cytoplasm. These different C-terminal sequences appear to dictate the functions of individual isoforms within a wide range of contexts.

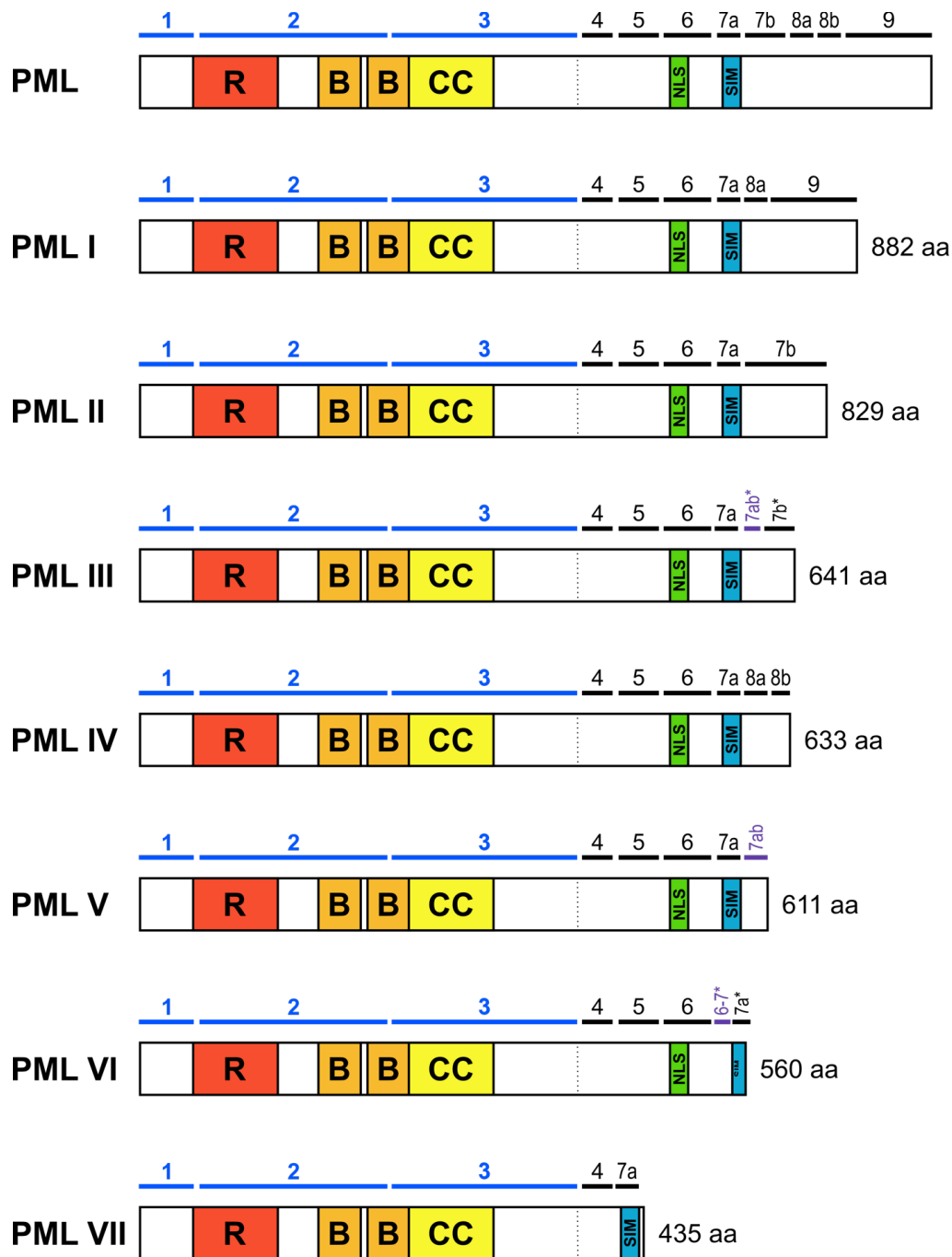


Figure 1.2 PML isoforms

Diagram describing PML transcriptional variants. The first, labeled PML, is a visualization of all available exons in the *PML* gene. The following six sequences represent the actual isoforms. The exons that make up the RBCC domain are in blue, followed by the those involved in splice variants in black. Retained introns are indicated in purple, while partial sequences are indicated with an asterisk. (Abbreviations: nuclear localization sequence = NLS; SUMO interacting motif = SIM)

PML protein is the core component of sub-nuclear protein complexes called promyelocytic leukemia nuclear bodies (PML-NBs). The membraneless organelles are dynamic in nature, containing the constitutive component proteins Sp100 and small ubiquitin-like modifier (SUMO) as well as any of many transient proteins depending on the circumstance and where the cell is in the cell cycle. The only time that PML-NBs are naturally without other proteins is during mitosis. At this point, PML becomes de-SUMOylated and temporarily dissociates from other proteins, including Sp100, but maintains self-associated due to its RBCC motif. Upon returning to interphase, PML is once again SUMOylated and associates with other components.

Cellular stressors are highly influential on the quantity, size, and composition of PML-NBs. These stressors include, but are not limited to, senescence, DNA damage, cytokines/IFNs, and viral infections. PML-NBs play a meaningful role in regulation of tumor suppression and the cell cycle, as well as being an integral component of both the intrinsic and innate immune response. The specific PML isoforms that are expressed in response to stress may differ depending on the cause. For example, isoform II expression is induced by type-I and -II interferon (IFN) signaling and the PML protein itself associates with transcription factors at the IFN-stimulated response element (ISRE) to induce expression of IFN-stimulated genes as well as IFN- β ⁸⁶. This isoform is more likely to be enriched in response to a viral infection than UV-radiation. Sometimes multiple isoforms may be expressed but only one is able to perform a given function. An example here would be isoform IV, which is the only form of PML known to induce senescence in primary cells⁸⁷. Given their critical role in the processes of host cells, it is unsurprising that PML-NBs are both a target of evasion and usurpation by viruses.

PML-NBs have been implicated in the pathogenesis of both RNA and DNA viruses. Both RNA and DNA viruses are inhibited by and reliant on PML-NBs during their life cycle. In many

cases the PML protein itself is key for interactions with the virus, but depending on the virus there may be different interactions between different viral proteins and PML-NB proteins. For example, human adenoviruses (HAdVs) have many and varied associations with PML and other component proteins that make up PML-NBs ⁸⁸. One of the most well-studied interactions is between E4 gene open reading frame 3 protein (E4Orf3) and PML isoform II ⁸⁹. E4orf3 directly binds to PML II, reorganizing PML-NBs into track-like formations. Another interaction is between the HAdV capsid protein, pVI, and the PML-NB associated protein, Daxx ⁹⁰. HAdV pVI displaces Daxx from PML-NBs which disrupts Daxx's function as a repressor of viral gene expression.

Studies on SV40 suggest that viral DNA replication occurs at PML-NBs ^{91,92}. A study of MCPyV found that PML-NBs changed in number and size in response to actively replicating virus ⁹³. BKPyV infection of renal proximal tubule epithelial (RPTE) cells functionally reorganizes PML-NBs ⁹⁴. Upon infection with BKPyV, the number of nuclear bodies decreases, and the size of the individual bodies increases. TAg expression from a plasmid is not sufficient to induce this response. Although the mechanism of NB rearrangement is not known, PML-NB function is impacted in the process. BKPyV infection and subsequent PML-NB reorganization renders cells equally susceptible to both WT herpes simplex virus 1 (HSV-1) and a mutant virus normally inhibited by functional PML. PML does not appear to be essential for rearranged BKPyV infection in RPTE cells, but further research is needed to better understand its role during infection.

1.3 Studying virology *in vitro*

Since the 1950's scientists have been using tissue culture lines to routinely propagate and study animal viruses ⁹⁵. And up until quite recently these studies were performed on a bulk, or population, level. This meant that data collected for a group of cells was often considered representative of any given cell within that group. Or if one were not to make such an assumption, then the group data would simply be an average of what could potentially be a wide range of values. Taking an average or assuming all cells have the same value is suboptimal, however, because it is rare that every cell within a well or a flask is infected or infected at the same time, especially when using a low multiplicity of infection. Additionally, different cells may have different expression levels of genes. These could change depending on the infection state of the cells around them and the density of the cells around them. Any of these variables within any given individual cell become potential variables during viral infection. One could potentially increase the homogeneity of the cell population by increasing the multiplicity of infection, and thus, the percentage of cells infected at once. This increased homogeneity might allow the measurements of the bulk culture to better reflect the measurements of the individual cells, but may also fundamentally change the experiment.

As the technology has become available, scientists have been developing techniques to account for these variables by studying viral infection at a single-cell level. Early studies were relatively low throughput, with methods often focused on measuring viral output, such as plaque assays and RT-qPCR ⁹⁶⁻⁹⁸. Later work included higher-throughput methods, such as transcriptomics. Single-cell RNA sequencing (scRNA-seq) studies have become common in many fields, including virology. Utilizing scRNA-seq has allowed researchers to study viral infection at a higher resolution than before. For example, a scRNA-seq analysis of HSV-1

infection of primary human fibroblasts identified the transcription factor, NRF2, as a restriction factor for HSV-1 infection ⁹⁹. The analyses facilitated by single-cell transcriptomics allowed the authors to detect NRF2 target genes that had not appeared in bulk RNAseq studies. Many other viruses have been studied using scRNA-seq as well. While there are numerous examples, a few include the Epstein-Barr virus, influenza A, Zika and Dengue viruses, and HIV-1 ^{100–103}. The only scRNA-seq study of BKPyV to date was carried out by the Pipas lab ¹⁰⁴. This study demonstrated that BKPyV gene expression levels did not progress in a linear fashion for all infected cells. While there was a population of cells with high viral gene expression, it was not large and did not increase in size substantially over time. There were also populations of cells with low and medium viral expression levels that remained relatively consistent over time. The authors proposed that the presence of these persistent low and medium expression populations indicates that cellular responses to infection are determined early or prior to infection. Although quite common, RNAseq experiments are not the only single-cell virus studies that exist. Other techniques include genomics, metabolomics, and microscopy ^{105–108}.

1.4 Goals of this dissertation

In this dissertation I examine BKPyV infection at a finer resolution using a single-cell high content microscopy analysis. Chapter Two will show that studying BKPyV on a single-cell level is an effective and efficient method of gaining insight into the heterogeneity of a population of BKPyV-infected cells. Chapter Two will also demonstrate there are differences among groups of cells even with the same levels of TAg and that some of these differences are in the cells' PML-NBs. In Chapter Three I will synthesize these results in the context of what is known about

BKPyV infection as well as other single-cell applications. I will also provide broad hypotheses for some of the unanswered questions and propose future experiments.

Table 1.1 Human Polyomaviruses

<u>Genera</u>	<u>Name</u>	<u>Species</u>	<u>Abbreviation</u>	<u>Year Published</u>	<u>Reference</u>
Beta	BK	HPyV1	BKPyV	1971	Gardner 1971 ¹²
Beta	JC	HPyV2	JCPyV	1971	Padgett 1971 ¹⁰
Beta	KI	HPyV3	KIPyV	2007	Allander 2007 ¹³
Beta	WU	HPyV4	WUPyV	2007	Gaynor 2007 ¹⁴
Alpha	Merkel cell	HPyV5	MCPyV	2008	Feng 2008 ¹⁵
Delta	HPyV6	HPyV6	HPyV6	2010	Schowalter 2010 ¹⁶
Delta	HPyV7	HPyV7	HPyV7	2010	Schowalter 2010 ¹⁶
Alpha	Trichodysplasia spinulosa	HPyV8	TSPyV	2010	Meijden 2010 ¹⁷
Alpha	HPyV9	HPyV9	HPyV9	2011	Scuda 2011 ¹⁸
Delta	Malawi	HPyV10	MWPyV	2012	Siebrasse 2012 ¹⁹
Delta	St. Louis	HPyV11	STLPyV	2013	Lim 2013 ²⁰
Alpha	HPyV12	HPyV12	HPyV12	2013	Korup 2013 ²¹
Alpha	New Jersey	HPyV13	NJPyV	2014	Mishra 2014 ²²

1.5 Bibliography

1. Gross, L. A Filterable Agent, Recovered from Ak Leukemic Extracts, Causing Salivary Gland Carcinomas in C3H Mice. *P Soc Exp Biol Med* **83**, 414–421 (1953).
2. Stewart, S. E., Eddy, B. E. & Borgese, N. Neoplasms in Mice Inoculated with a Tumor Agent Carried in Tissue Culture². *Jnci J National Cancer Inst* **20**, 1223–1243 (1958).
3. Kilham, L. & Murphy, H. W. A Pneumotropic Virus Isolated from C3H Mice Carrying the Bittner Milk Agent. *P Soc Exp Biol Med* **82**, 133–137 (1953).
4. Sweet, B. H. & Hilleman, M. R. The Vacuolating Virus, S.V.40. *P Soc Exp Biol Med* **105**, 420–427 (1960).
5. Rabson, A. S. & Kirschstein, R. L. Induction of Malignancy in vitro in Newborn Hamster Kidney Tissue Infected with Simian Vacuolating Virus (SV40). *P Soc Exp Biol Med* **111**, 323–328 (1962).
6. Black, P. H., Rowe, W. P. & Cooper, H. L. An Analysis of SV40-induced Transformation of Hamster Kidney Tissue in vitro, II. Studies of Three Clones Derived from a Continuous Line of Transformed Cells. *Proc National Acad Sci* **50**, 847–854 (1963).
7. Poulin, D. L. & DeCaprio, J. A. Is There a Role for SV40 in Human Cancer? *J Clin Oncol* **24**, 4356–4365 (2006).
8. Hartley, J. W. & Rowe, W. P. New Papovavirus Contaminating Shope Papillomata. *Science* **143**, 258–260 (1964).
9. Moens, U. *et al.* ICTV Virus Taxonomy Profile: Polyomaviridae. *J Gen Virol* **98**, 1159–1160 (2017).
10. Padgett, B. L., Zurhein, G. M., Walker, D. L., Eckroade, R. J. & Dessel, B. H. Cultivation of Papova-like Virus from Human Brain with Progressive Multifocal Leucoencephalopathy. *Lancet* **297**, 1257–1260 (1971).
11. Rhein, G. M. Z. & Chou, S.-M. Particles Resembling Papova Viruses in Human Cerebral Demyelinating Disease. *Science* **148**, 1477–1479 (1965).
12. Gardner, S. D., Field, A. M., Coleman, D. V. & Hulme, B. New Human Papovirus (B.K.) Isolated from Urine after Renal Transplantation. *Lancet* **297**, 1253–1257 (1971).
13. Allander, T. *et al.* Identification of a Third Human Polyomavirus. *J Virol* **81**, 4130–4136 (2007).

14. Gaynor, A. M. *et al.* Identification of a Novel Polyomavirus from Patients with Acute Respiratory Tract Infections. *Plos Pathog* **3**, e64 (2007).
15. Feng, H., Shuda, M., Chang, Y. & Moore, P. S. Clonal Integration of a Polyomavirus in Human Merkel Cell Carcinoma. *Science* **319**, 1096–1100 (2008).
16. Schowalter, R. M., Pastrana, D. V., Pumphrey, K. A., Moyer, A. L. & Buck, C. B. Merkel Cell Polyomavirus and Two Previously Unknown Polyomaviruses Are Chronically Shed from Human Skin. *Cell Host Microbe* **7**, 509–515 (2010).
17. Meijden, E. van der *et al.* Discovery of a New Human Polyomavirus Associated with Trichodysplasia Spinulosa in an Immunocompromized Patient. *Plos Pathog* **6**, e1001024 (2010).
18. Scuda, N. *et al.* A Novel Human Polyomavirus Closely Related to the African Green Monkey-Derived Lymphotropic Polyomavirus. *J Virol* **85**, 4586–4590 (2011).
19. Siebrasse, E. A. *et al.* Identification of MW Polyomavirus, a Novel Polyomavirus in Human Stool. *J Virol* **86**, 10321–10326 (2012).
20. Lim, E. S. *et al.* Discovery of STL polyomavirus, a polyomavirus of ancestral recombinant origin that encodes a unique T antigen by alternative splicing. *Virology* **436**, 295–303 (2013).
21. Korup, S. *et al.* Identification of a Novel Human Polyomavirus in Organs of the Gastrointestinal Tract. *Plos One* **8**, e58021 (2013).
22. Mishra, N. *et al.* Identification of a Novel Polyomavirus in a Pancreatic Transplant Recipient with Retinal Blindness and Vasculitic Myopathy. *J Infect Dis* **210**, 1595–1599 (2014).
23. Kamminga, S., Meijden, E. van der, Feltkamp, M. C. W. & Zaaijer, H. L. Seroprevalence of fourteen human polyomaviruses determined in blood donors. *Plos One* **13**, e0206273 (2018).
24. Goudsmit, J., Dillen, P. W., Strien, A. van & Noordaa, J. van der. The role of BK virus in acute respiratory tract disease and the presence of BKV DNA in tonsils. *J. Med. Virol.* **10**, 91–99 (1982).
25. Jin, L., Gibson, P. E., Booth, J. C. & Clewley, J. P. Genomic typing of BK virus in clinical specimens by direct sequencing of polymerase chain reaction products. *J. Med. Virol.* **41**, 11–17 (1993).
26. Knowles, W. A., Gibson, P. E. & Gardner, S. D. Serological typing scheme for bk-like isolates of human polyomavirus. *J. Med. Virol.* **28**, 118–123 (1989).
27. Torres, C. Evolution and molecular epidemiology of polyomaviruses. *Infect Genetics Evol* **79**, 104150 (2020).

28. Yogo, Y., Sugimoto, C., Zhong, S. & Homma, Y. Evolution of the BK polyomavirus: epidemiological, anthropological and clinical implications. *Rev Med Virol* **19**, 185–199 (2009).
29. Zheng, H.-Y. *et al.* Relationships between BK virus lineages and human populations. *Microbes Infect* **9**, 204–213 (2007).
30. Takasaka, T. *et al.* Subtypes of BK virus prevalent in Japan and variation in their transcriptional control region. *J Gen Virol* **85**, 2821–2827 (2004).
31. Imlay, H. *et al.* Consensus Definitions of BK Polyomavirus Nephropathy in Renal Transplant Recipients for Clinical Trials. *Clin Infect Dis* (2022) doi:10.1093/cid/ciac071.
32. Cesaro, S. *et al.* ECIL guidelines for the prevention, diagnosis and treatment of BK polyomavirus-associated haemorrhagic cystitis in haematopoietic stem cell transplant recipients. *J Antimicrob Chemoth* **73**, 12–21 (2017).
33. Li, Y.-J. *et al.* High Incidence and Early Onset of Urinary Tract Cancers in Patients with BK Polyomavirus Associated Nephropathy. *Viruses* **13**, 476 (2021).
34. Takemoto, K. K. & Martin, M. A. Transformation of Hamster Kidney Cells by BK Papovavirus DNA. *J Virol* **17**, 247–253 (1976).
35. Major, E. O. & Mayorca, G. D. Malignant Transformation of BHK21 Clone 13 Cells by BK Virus—A Human Papovavirus. *Proc National Acad Sci* **70**, 3210–3212 (1973).
36. Portolani, M., Barbanti-Brodano, G. & Placa, M. L. Malignant transformation of hamster kidney cells by BK virus. *J Virol* **15**, 420–422 (1975).
37. Takemoto, K. K., Linke, H., Miyamura, T. & Fareed, G. C. Persistent BK papovavirus infection of transformed human fetal brain cells. I. Episomal viral DNA in cloned lines deficient in T-antigen expression. *J Virol* **29**, 1177–1185 (1979).
38. Nakshatri, H., Pater, M. M. & Pater, A. Functional role of BK virus tumor antigens in transformation. *J Virol* **62**, 4613–4621 (1988).
39. Alexiev, B. A. *et al.* BK virus-associated urinary bladder carcinoma in transplant recipients: report of 2 cases, review of the literature, and proposed pathogenetic model. *Hum Pathol* **44**, 908–917 (2013).
40. Baker, S. C. *et al.* Induction of APOBEC3-mediated genomic damage in urothelium implicates BK polyomavirus (BKPyV) as a hit-and-run driver for bladder cancer. *Oncogene* **41**, 2139–2151 (2022).
41. Starrett, G. J. & Buck, C. B. BK polyomavirus is a cause of bladder cancer. *Curr Opin Virol* **39**, 8–15 (2019).

42. Harris, R. S. & Dudley, J. P. APOBECs and virus restriction. *Virology* **479**, 131–145 (2015).
43. Verhalen, B., Starrett, G. J., Harris, R. S. & Jiang, M. Functional Upregulation of the DNA Cytosine Deaminase APOBEC3B by Polyomaviruses. *J Virol* **90**, 6379–6386 (2016).
44. Starrett, G. J. *et al.* Polyomavirus T Antigen Induces APOBEC3B Expression Using an LXCXE-Dependent and TP53-Independent Mechanism. *Mbio* **10**, e02690-18 (2019).
45. Peretti, A. *et al.* Characterization of BK Polyomaviruses from Kidney Transplant Recipients Suggests a Role for APOBEC3 in Driving In-Host Virus Evolution. *Cell Host Microbe* **23**, 628–635.e7 (2018).
46. Burns, M. B., Temiz, N. A. & Harris, R. S. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat Genet* **45**, 977–983 (2013).
47. Zeng, Y., Sun, J., Bao, J. & Zhu, T. BK polyomavirus infection promotes growth and aggressiveness in bladder cancer. *Virol J* **17**, 139 (2020).
48. Griffith, J. P., Griffith, D. L., Rayment, I., Murakami, W. T. & Caspar, D. L. D. Inside polyomavirus at 25-Å resolution. *Nature* **355**, 652–654 (1992).
49. Howley, P. M., Khoury, G., Byrne, J. C., Takemoto, K. K. & Martin, M. A. Physical map of the BK virus genome. *J Virol* **16**, 959–973 (1975).
50. Yang, R. C. A. & Wu, R. BK Virus DNA: Complete Nucleotide Sequence of a Human Tumor Virus. *Science* **206**, 456–462 (1979).
51. Deyerle, K. L., Cassill, J. A. & Subramani, S. Analysis of the early regulatory region of the human papovavirus BK. *Virology* **158**, 181–193 (1987).
52. Watanabe, S., Kotake, S., Nozawa, A., Uchida, S. & Muto, T. Tumorigenicity of human bk papovavirus plaque isolates, wild-type and plaque morphology mutant, in hamsters. *Int. J. Cancer* **29**, 583–586 (1982).
53. Deyerle, K. L. & Subramani, S. Linker scan analysis of the early regulatory region of human papovavirus BK. *J Virol* **62**, 3378–3387 (1988).
54. Mew, R. T., Lecatsas, G., Prozesky, O. W. & Harley, E. H. Characteristics of BK Papovavirus DNA Prepared Directly from Human Urine. *Intervirology* **16**, 14–19 (1981).
55. Rubinstein, R., Pare, N. & Harley, E. H. Structure and function of the transcriptional control region of nonpassaged BK virus. *J Virol* **61**, 1747–1750 (1987).
56. Rubinstein, R. & Harley, E. H. BK virus DNA cloned directly from human urine confirms an archetypal structure for the transcriptional control region. *Virus Genes* **2**, 157–165 (1989).

57. Rubinstein, R., Schoonakker, B. C. & Harley, E. H. Recurring theme of changes in the transcriptional control region of BK virus during adaptation to cell culture. *J Virol* **65**, 1600–1604 (1991).
58. Low, J., Humes, H. D., Szczypka, M. & Imperiale, M. BKV and SV40 infection of human kidney tubular epithelial cells in vitro. *Virology* **323**, 182–188 (2004).
59. Broekema, N. M. & Imperiale, M. J. Efficient propagation of archetype BK and JC polyomaviruses. *Virology* **422**, 235–241 (2012).
60. Low, J. A., Magnuson, B., Tsai, B. & Imperiale, M. J. Identification of Gangliosides GD1b and GT1b as Receptors for BK Virus. *J Virol* **80**, 1361–1366 (2006).
61. Zhao, L., Marciano, A. T., Rivet, C. R. & Imperiale, M. J. Caveolin- and clathrin-independent entry of BKPyV into primary human proximal tubule epithelial cells. *Virology* **492**, 66–72 (2016).
62. Jiang, M., Abend, J. R., Tsai, B. & Imperiale, M. J. Early Events during BK Virus Entry and Disassembly. *J Virol* **83**, 1350–1358 (2009).
63. Bennett, S. M., Jiang, M. & Imperiale, M. J. Role of Cell-Type-Specific Endoplasmic Reticulum-Associated Degradation in Polyomavirus Trafficking. *J Virol* **87**, 8843–8852 (2013).
64. Bennett, S. M., Zhao, L., Bosard, C. & Imperiale, M. J. Role of a nuclear localization signal on the minor capsid Proteins VP2 and VP3 in BKPyV nuclear entry. *Virology* **474**, 110–116 (2015).
65. Nomburg, J. *et al.* Long-read sequencing reveals complex patterns of wraparound transcription in polyomaviruses. *Plos Pathog* **18**, e1010401 (2022).
66. Seif, I., Khoury, G. & Dhar, R. The genome of human papovavirus BKV. *Cell* **18**, 963–977 (1979).
67. Seo, G. J., Fink, L. H. L., O’Hara, B., Atwood, W. J. & Sullivan, C. S. Evolutionarily Conserved Function of a Viral MicroRNA. *J Virol* **82**, 9823–9828 (2008).
68. Grundhoff, A. & Sullivan, C. S. Virus-encoded microRNAs. *Virology* **411**, 325–343 (2011).
69. Cho, U. S. *et al.* Structural Basis of PP2A Inhibition by Small t Antigen. *Plos Biol* **5**, e202 (2007).
70. Zou, W. & Imperiale, M. J. Regulation of Virus Replication by BK Polyomavirus Small T Antigen. *J Virol* e00077-23 (2023) doi:10.1128/jvi.00077-23.
71. Broekema, N. M. & Imperiale, M. J. miRNA regulation of BK polyomavirus replication during early infection. *Proc National Acad Sci* **110**, 8200–8205 (2013).

72. Kean, J. M., Rao, S., Wang, M. & Garcea, R. L. Seroepidemiology of Human Polyomaviruses. *Plos Pathog* **5**, e1000363 (2009).
73. Peretti, A. *et al.* A multivalent polyomavirus vaccine elicits durable neutralizing antibody responses in macaques. *Vaccine* **41**, 1735–1742 (2023).
74. Johannessen, M. *et al.* BKV Agnoprotein Interacts with α -Soluble N-Ethylmaleimide-Sensitive Fusion Attachment Protein, and Negatively Influences Transport of VSVG-EGFP. *Plos One* **6**, e24489 (2011).
75. Panou, M.-M. *et al.* Agnoprotein Is an Essential Egress Factor during BK Polyomavirus Infection. *Int J Mol Sci* **19**, 902 (2018).
76. Evans, G. L., Caller, L. G., Foster, V. & Crump, C. M. Anion homeostasis is important for non-lytic release of BK polyomavirus from infected cells. *Open Biol* **5**, 150041 (2015).
77. Engeland, K. Cell cycle regulation: p53-p21-RB signaling. *Cell Death Differ* **29**, 946–960 (2022).
78. Jiang, M., Zhao, L., Gamez, M. & Imperiale, M. J. Roles of ATM and ATR-Mediated DNA Damage Responses during Lytic BK Polyomavirus Infection. *Plos Pathog* **8**, e1002898 (2012).
79. Sowd, G. A., Li, N. Y. & Fanning, E. ATM and ATR Activities Maintain Replication Fork Integrity during SV40 Chromatin Replication. *Plos Pathog* **9**, e1003283 (2013).
80. Sowd, G. A. *et al.* SV40 Utilizes ATM Kinase Activity to Prevent Non-homologous End Joining of Broken Viral DNA Replication Products. *Plos Pathog* **10**, e1004536 (2014).
81. Heald, R., McLoughlin, M. & McKeon, F. Human wee1 maintains mitotic timing by protecting the nucleus from cytoplasmically activated cdc2 kinase. *Cell* **74**, 463–474 (1993).
82. Justice, J. L., Needham, J. M. & Thompson, S. R. BK Polyomavirus Activates the DNA Damage Response to Prolong S Phase. *J Virol* **93**, (2019).
83. Verhalen, B., Justice, J. L., Imperiale, M. J. & Jiang, M. Viral DNA Replication-Dependent DNA Damage Response Activation during BK Polyomavirus Infection. *J Virol* **89**, 5032–5039 (2015).
84. Justice, J. L., Needham, J. M., Verhalen, B., Jiang, M. & Thompson, S. R. BK Polyomavirus Requires the Mismatch Repair Pathway for DNA Damage Response Activation. *J Virol* **96**, e02028-21 (2022).
85. Longo, L. *et al.* Rearrangements and aberrant expression of the retinoic acid receptor alpha gene in acute promyelocytic leukemias. *J Exp Medicine* **172**, 1571–1575 (1990).

86. Chen, Y., Wright, J., Meng, X. & Leppard, K. N. Promyelocytic Leukemia Protein Isoform II Promotes Transcription Factor Recruitment to Activate Interferon Beta and Interferon-Responsive Gene Expression. *Mol Cell Biol* **35**, 1660–1672 (2015).
87. Bischof, O. *et al.* Deconstructing PML-induced premature senescence. *Embo J* **21**, 3358–3369 (2002).
88. Hofmann, S., Stubbe, M., Mai, J. & Schreiner, S. Double-edged role of PML nuclear bodies during human adenovirus infection. *Virus Res* **295**, 198280 (2021).
89. Hoppe, A., Beech, S. J., Dimmock, J. & Leppard, K. N. Interaction of the Adenovirus Type 5 E4 Orf3 Protein with Promyelocytic Leukemia Protein Isoform II Is Required for ND10 Disruption. *J Virol* **80**, 3042–3049 (2006).
90. Schreiner, S. *et al.* Transcriptional Activation of the Adenoviral Genome Is Mediated by Capsid Protein VI. *Plos Pathog* **8**, e1002549 (2012).
91. Tang, Q., Bell, P., Tegtmeyer, P. & Maul, G. G. Replication but Not Transcription of Simian Virus 40 DNA Is Dependent on Nuclear Domain 10. *J Virol* **74**, 9694–9700 (2000).
92. Ishov, A. M. & Maul, G. G. The periphery of nuclear domain 10 (ND10) as site of DNA virus deposition. *J Cell Biology* **134**, 815–826 (1996).
93. Neumann, F. *et al.* Replication of Merkel cell polyomavirus induces reorganization of promyelocytic leukemia nuclear bodies. *J Gen Virol* **97**, 2926–2938 (2016).
94. Jiang, M., Entezami, P., Gamez, M., Stamminger, T. & Imperiale, M. J. Functional Reorganization of Promyelocytic Leukemia Nuclear Bodies during BK Virus Infection. *Mbio* **2**, e00281-10 (2011).
95. Oldstone, M. B. A. & Levine, A. J. Virology in the Next Millennium. *Cell* **100**, 139–142 (2000).
96. Zhu, Y., Yongky, A. & Yin, J. Growth of an RNA virus in single cells reveals a broad fitness distribution. *Virology* **385**, 39–46 (2009).
97. Timm, A. & Yin, J. Kinetics of virus production from single cells. *Virology* **424**, 11–17 (2012).
98. Schulte, M. B. & Andino, R. Single-Cell Analysis Uncovers Extensive Biological Noise in Poliovirus Replication. *J. Virol.* **88**, 6205–6212 (2014).
99. Wyler, E. *et al.* Single-cell RNA-sequencing of herpes simplex virus 1-infected cells connects NRF2 activation to an antiviral program. *Nat Commun* **10**, 4878 (2019).

100. Yan, B. *et al.* A comprehensive single cell data analysis of lymphoblastoid cells reveals the role of super-enhancers in maintaining EBV latency. *J. Méd. Virol.* **95**, e28362 (2023).
101. Ramos, I. *et al.* Innate Immune Response to Influenza Virus at Single-Cell Resolution in Human Epithelial Cells Revealed Paracrine Induction of Interferon Lambda 1. *J Virol* **93**, e00559-19 (2019).
102. Zanini, F., Pu, S.-Y., Bekerman, E., Einav, S. & Quake, S. R. Single-cell transcriptional dynamics of flavivirus infection. *Elife* **7**, e32942 (2018).
103. Wiegand, A. *et al.* Single-cell analysis of HIV-1 transcriptional activity reveals expression of proviruses in expanded clones during ART. *Proc. Natl. Acad. Sci.* **114**, E3659–E3668 (2017).
104. An, P. *et al.* Single-Cell Transcriptomics Reveals a Heterogeneous Cellular Response to BK Virus Infection. *J Virol* **95**, (2021).
105. Ciuffi, A., Rato, S. & Telenti, A. Single-Cell Genomics for Virology. *Viruses* **8**, 123 (2016).
106. Kumar, R., Ghosh, M., Kumar, S. & Prasad, M. Single Cell Metabolomics: A Future Tool to Unmask Cellular Heterogeneity and Virus-Host Interaction in Context of Emerging Viral Diseases. *Front Microbiol* **11**, 1152 (2020).
107. Francis, A. C. & Melikyan, G. B. Live-Cell Imaging of Early Steps of Single HIV-1 Infection. *Viruses* **10**, 275 (2018).
108. Guo, F. *et al.* Single-Cell Virology: On-Chip Investigation of Viral Infection Dynamics. *Cell Reports* **21**, 1692–1704 (2017).

Chapter 2

Single Cell High Content Microscopy Analysis of BK Polyomavirus Infection

Megan C. Procario¹, Jonathan Z. Sexton^{2,3,4}, Benjamin S. Halligan², Michael J. Imperiale^{1,5,*}

¹Department of Microbiology and Immunology, Medical School

²Department of Internal Medicine, Medical School

³Department of Medicinal Chemistry, College of Pharmacy

⁴Center for Drug Repurposing

⁵Rogel Cancer Center

University of Michigan, Ann Arbor, MI 48109

* Corresponding author

2.1 Abstract

By adulthood, the majority of the population is persistently infected with BK polyomavirus (BKPyV). Only a subset of the population, generally transplant recipients on immunosuppressive drugs, will experience disease from BKPyV, but those who do have few treatment options and frequently poor outcomes because to date there are no effective antivirals to treat or approved vaccines to prevent BKPyV. Most studies of BKPyV have been performed on bulk populations of cells and the dynamics of infection at single-cell resolution have not been explored. As a result, much of our knowledge is based upon the assumption that all cells within a greater population are behaving the same way with respect to infection. The present study examines BKPyV infection on a single-cell level using high content microscopy to measure and analyze the viral protein large T antigen (TAg), promyelocytic leukemia protein (PML), DNA, and

nuclear morphological features. We observed significant heterogeneity among infected cells, within and across timepoints. We found that the levels of TAg within individual cells did not necessarily increase with time, and that cells with the same TAg levels varied in other ways. Overall, high-content single-cell microscopy is a novel approach to studying BKPyV that enables experimental insight into the heterogenous nature of the infection.

2.2 Importance

BK polyomavirus (BKPyV) is a human pathogen that infects nearly everyone by adulthood and persists throughout a person's life. Only people with significant immune suppression develop disease from the virus, however. Until recently the only practical means of studying many viral infections was to infect a group of cells in the laboratory and measure the outcomes in that group. However, interpreting these bulk population experiments requires the assumption that infection influences all cells within a group similarly. This assumption has not held for multiple viruses tested so far. Our study establishes a novel single-cell microscopy assay for BKPyV infection. Using this assay, we discovered differences among individual infected cells that have not been apparent in bulk population studies. The knowledge gained in this study and the potential for future use demonstrate the power of this assay as a tool for understanding the biology of BKPyV.

2.3 Introduction

BK polyomavirus (BKPyV) is a small double stranded DNA tumor virus. It is one of approximately fourteen human polyomaviruses (HPyV). BKPyV infection, while usually asymptomatic, is very common and persists throughout a person's lifetime. Greater than 90% of individuals are seropositive for BKPyV by the time they reach adulthood ^{1,2}. Despite its ubiquity, BKPyV's disease burden is limited to patients post-transplant (reviewed in ³. Any person with significant immune defects may suffer from BKPyV-associated disease, but the most common are patients undergoing bone marrow, hematopoietic stem cell, or solid organ transplants. One complication that can occur in kidney transplant patients is polyomavirus associated nephropathy (PVAN) because of new or reactivated BKPyV infection ⁴. The same immunosuppressant medications used to prevent rejection of the donor organs can prevent a sufficient immune response to clear the virus. There are currently no effective antivirals or approved vaccines to treat or prevent BKPyV, respectively ⁵⁻⁸. In these circumstances, clinicians must decide between decreasing the immunosuppression (and risk losing the graft) and maintaining immunosuppression (and risk uncontrolled viremia and destruction of the graft) ^{9,10}.

The BKPyV genome is circular and transcription is driven by bidirectional promoters within a noncoding control region (NCCR), which also contains the origin of DNA replication as well as transcription factor binding sites. One promoter drives early gene expression, which includes large (T), small (t), truncated (truncT) tumor antigen (Ag), and SuperT ¹¹⁻¹³. The other drives late gene expression of agnoprotein and the capsid proteins, VP1, VP2, and VP3. Upon entering the nucleus, the virus begins transcription of the early region, followed by translation and nuclear translocation of the proteins to facilitate viral DNA replication. They do so by stimulating the cell's entry into S phase, recruiting host DNA replication machinery to the viral genome, and

preventing apoptosis. BKPyV relies on manipulation of cell cycle regulation to facilitate efficient replication, and both induces and relies upon the DNA damage response (DDR) for productive infection ^{14,15}. In conjunction with the onset of DNA replication, late genes are transcribed from the opposite promoter and the capsid proteins are translocated to the nucleus where they self-assemble and package the nascent viral DNA to create progeny virions ¹³.

As demonstrated by its tumor antigens, BKPyV influences the behavior of many nuclear factors during the course of infection. These include promyelocytic leukemia nuclear bodies (PML-NBs) ¹⁶. PML-NBs are dynamic, punctate, sub-nuclear structures found in all mammalian cells. They are multi-protein complexes containing a scaffold of PML protein, the other constitutive component proteins Sp100 and Daxx, and a number of other transiently associated proteins as circumstances require ^{17–20}. These nuclear bodies are highly sumoylated and interact closely with cellular chromatin ^{21–23}. PML-NBs respond to many forms of stress in the cell, ranging from metabolic deficiencies to senescence. Of particular relevance to BKPyV are their roles in both tumor suppression and antiviral pathways. Additionally, PML-NBs interact with the DDR pathway ^{24–26}. BKPyV infection, along with that of many other viruses, has been demonstrated to significantly alter the structure and function of PML-NBs (reviewed in ^{27–30}).

The current *in vitro* model of primary renal proximal tubule epithelial (RPTE) cell infection has allowed many discoveries about BKPyV including the viral life cycle, genetic regulation, and virus-host interactions (including PML-NBs) ^{16,31–42}. Almost all of these studies have been performed at the cell population level and the results have generally been interpreted as being representative of any given individual cell in that population. However, a recent single cell RNAseq study from the Pipas lab ⁴³ has indicated that BKPyV infection is heterogeneous, with several discrete cellular states. In the present study, we sought to similarly examine single cell

phenotypes across the course of BKPyV infection using high-content microscopy. This morphological examination focused on three parameters: TAg, to identify BKPyV-positive cells and characterize the amount and distribution of the viral protein; PML, to identify structural changes in NBs; and DNA, as a marker of the nucleus as well as to inform about impact of infection on DNA and the nucleus in general. The high resolution and quantity of images allowed for an advanced microscopic analysis of single cell BKPyV infection that has not previously been attempted. These data provide insights into the heterogeneity of individual cells during the course of infection and offer suggestions for future mechanistic studies of BKPyV.

2.4 Results

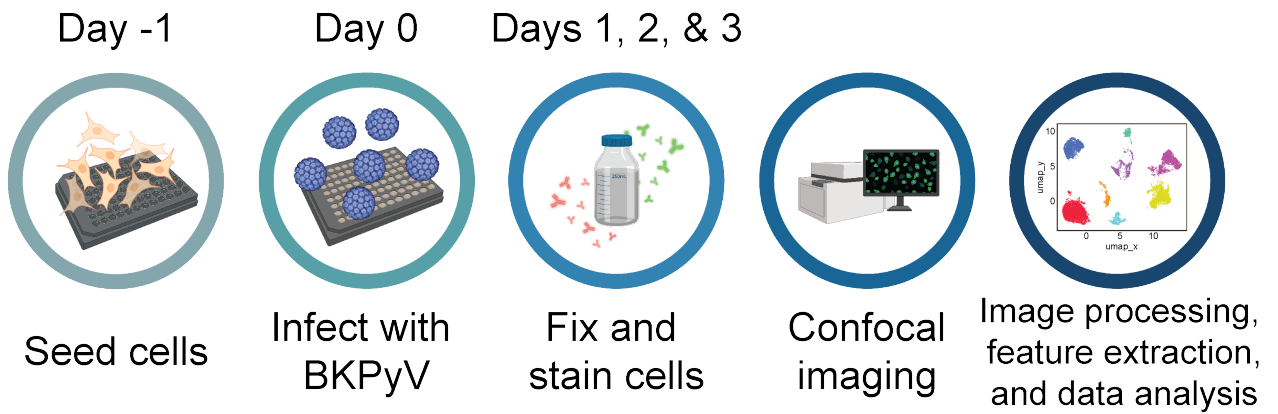
The vast majority of our understanding of the biology of BKPyV has come from bulk cell population level studies, many using the *in vitro* model of renal proximal tubule epithelial (RPTE) cells. Timecourse experiments have demonstrated these infections to be a linear progression from cells expressing low levels of viral proteins at early timepoints to much greater amounts at later timepoints¹³. However, a recent study examining BKPyV infection on a single-cell transcriptomic level has identified distinct sub-populations of infected cells⁴³. Cells displayed widely variable levels of viral transcripts within the population and these differences were maintained as the infection progressed. Viral gene expression in cells did not always increase from low to high over time, further highlighting distinct sub-populations. Our study complements and expands on this newly-emerging understanding of BKPyV infection at the single-cell level by focusing on morphological measurements using high-content confocal microscopy.

For our single-cell analysis of BKPyV infection, we chose to focus solely on nuclear features because BKPyV replicates and assembles in the nucleus. During the course of infection, nuclear host proteins including PML are modified and/or regulated. We sought to expand on our previous work that identified BKPyV infection as a cause of both structural and functional changes in PML-NBs¹⁶. Specifically, we aimed to characterize the morphotype of PML-NB on a single-cell level throughout the course of BKPyV infection, paying particular attention to any correlation with TAg expression.

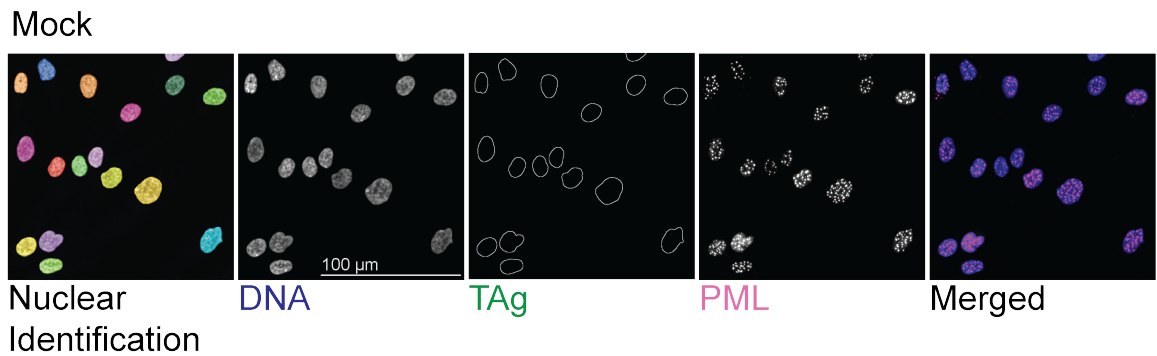
2.4.1 Assay overview and imaging examples

To collect the image data, cells were infected at an MOI of 0.5 FFU/cell or mock-infected, then fixed and stained after 1, 2, or 3 days (Figure 2.1A). The early viral protein TAg, and the principal component of PML-NBs, PML protein, were detected by immunofluorescence and DNA was stained using Hoechst 33342 (Figure 2.1B and Figure 2.1C). We measured features relating to the objects' size, shape, and area as well as the magnitude, variation, and spatial distribution of signal intensity within each of the three channels (Supplemental Table 2.3, Supplemental Table 2.4, Supplemental Table 2.5). PML-NBs in each nucleus were identified and select measurements were made for these as well.

A



B



C

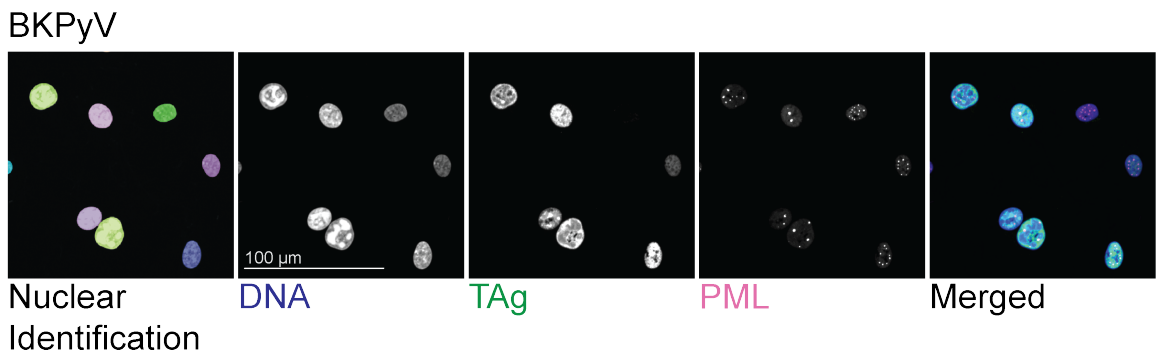


Figure 2.1 Assay overview and imaging examples.

A. Timeline illustrating the experimental protocol, the details for which can be found in Materials and Methods. B and C. Representative images of cells at 2 days post-infection. The first panel indicates the identified nuclei as identified using the CellPose software, pseudo-colored arbitrarily to distinguish individual objects. The next three panels are the individual channels, while the final panel is pseudo-colored (DNA = blue, TAg = green, PML = magenta) and merged. B. Cells from a mock infected well. Nuclei are outlined in the TAg panel due to lack of visible signal. C. Cells from a BKPvV-infected well.

2.4.2 Single cell microscopy as a method for studying BKPyV infection.

To validate the single-cell microscopy assay, we first examined a few well-established characteristics of BKPyV infection. Infection induces host cells to enter into S phase of the cell cycle and causes G2/M arrest, which increases the DNA content of infected cells ^{14,43}. BKPyV-infected cells have 4N DNA content or greater once they have progressed beyond S phase ⁴⁴. To determine if our model demonstrated such an effect, we plotted histograms of the total DNA pixel intensity measurements from either the mock-infected cells or TAg⁺ cells at each time point and examined the distribution of cells within each population. For mock-infected cells, the percentage of cells with > 2N DNA content was approximately 21% for all three days. However, in the TAg⁺ cell population, this fraction of cells increased over time: 30%, 45%, and 56% at 1, 2, and 3 DPI, respectively (Figure 2.2A).

As an additional validation point, we quantified the number of PML-NBs per nucleus. BKPyV infection induces PML-NB rearrangement, which includes a reduction in the number of bodies ¹⁶. We plotted histograms of the distribution of numbers of PML-NB per nucleus in the mock-infected and TAg⁺ cells at all time points (Figure 2.2B). We found that 10% of mock-infected cells fell into the smallest bin of the histogram, ≤ 8 nuclear bodies per nucleus, compared to 29% of TAg⁺ cells. This finding confirms our previous work but at a larger scale as here we have analyzed roughly 40,000 cells, orders of magnitude greater than what was feasible without high-content analysis ¹⁶.

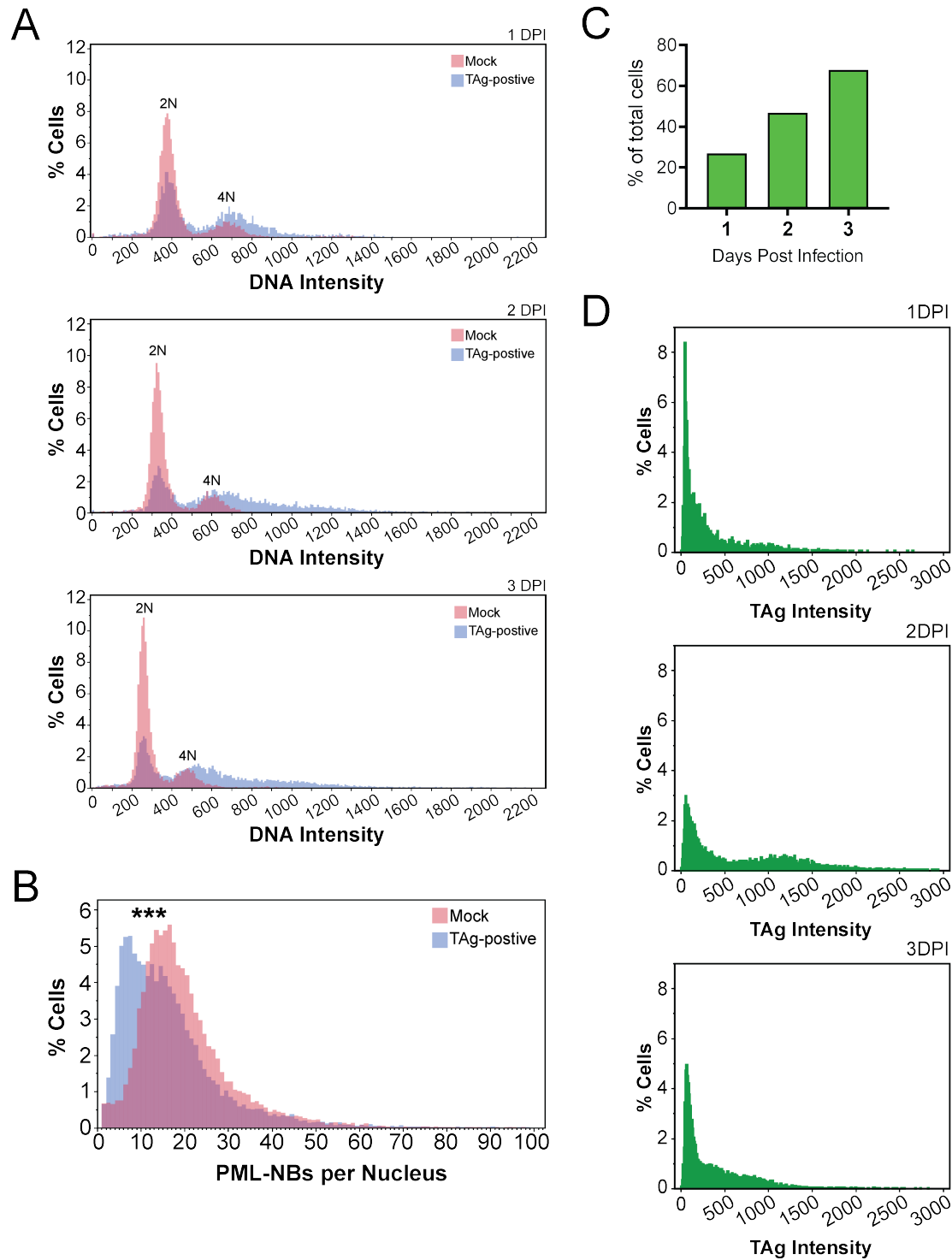


Figure 2.2 Single cell microscopy as a method for studying BKPvV infection.

A. Histograms depicting the DNA content of mock-infected (pink) and TAG-positive (blue) cells at the three timepoints. B. As in A but depicting quantities of PML-NBs per nucleus in mock-infected (pink) and TAG-positive (blue) cells. C. Percentage of TAG-positive cells out of total population of cells at the indicated timepoints. D. Histograms of only TAG-positive cells depicting the distribution of cells with different levels of TAG at 1, 2, and 3-days post infection. *** $p < 0.0001$ Wilcoxon one-way analysis

Having validated that we were able to capture these two known host cell properties that change during infection, we next sought to focus on the characteristics of infected, TAg⁺ cells. First, we quantified the percentage of TAg⁺ cells at each timepoint and found that, as expected, this value increased with time (Figure 2.2C). Next, we quantified the TAg signal intensities for each day post infection (Figure 2.2D). Consistent with the report from the Pipas laboratory, we found a wide range of TAg expression at all timepoints⁴³. While the data from days 1 and 3 are similar, with the majority of cells falling into low TAg intensity bins, the plot from day 2 had a bimodal distribution and a wider range. Unexpectedly, a greater percentage of cells at 2 dpi had overall higher TAg expression than those at 1 or 3 dpi.

2.4.3 High DNA content, not time post-infection, distinguishes cells expressing the highest levels of TAg

We next wanted to characterize the heterogeneity in infected cells. To visualize distinct morphologic cell classes, we employed dimensionality reduction and clustering using Uniform Manifold Approximation and Projection (UMAP) embedding. We started with a global UMAP, including all the cells regardless of infection or TAg status, and labeled the major clusters of cells, alphabetically (Figure 2.3A). It was immediately apparent that multiple distinct subgroups existed within the population of cells. Therefore, explore phenotypic features of all cells and differentiating factors in the subset of infected cells. First, to determine if exposure to viral inoculum defined the clusters, we highlighted the cells that received mock inoculum (Figure 2.3B) and BKPyV (Figure 2.3C). Apart from Cluster C, which was comprised only of cells that had received viral inoculum, there did not appear to be any impact of infection condition on which cluster a cell was in. Second, we highlighted just the TAg⁺ cells to see if that better

explained the clustering (Figure 2.3D). Not all cells that received the viral inoculum expressed detectable TAg, and TAg⁺ cells were found in all the clusters. Interestingly, Cluster C did not just exclusively contain infected cells, but also almost all the cells were TAg⁺. We next asked if time post infection was dictating some of the clustering. We colored cells on the UMAP by their day post infection (Figure 2.3E and Supplemental Figure 2.8A). Cells from each time point were present in every cluster, including Cluster C, which has more cells at 2 days post infection than at 1 or 3 (Supplemental Table 2.6). To determine which features in addition to TAg made Cluster C distinct from the others, we used predictive modeling to identify the main contributors. The most significant feature was integrated DNA intensity, which is proportional to the amount of DNA in the nucleus (Supplemental Figure 2.8B). Although other clusters had TAg⁺ cells and a range of DNA intensity, Cluster C was distinct in having the highest DNA intensity overall. We assembled image montages of random cells from each of the clusters (Figure 2.3F). These montages revealed patterns shared among cells that an observer may not notice if simply scanning a slide by eye. For example, we identified a phenotype in some cells in Cluster C in which there appeared to be areas of very intense DNA signal within the nucleus. These areas of DNA intensity appeared to exhibit a much lower TAg signal than the rest of the nucleus, and the inverse was true as well. To examine this in more detail we plotted the TAg and DNA intensities along a line drawn across the longest axis of a representative cell from each timepoint (Figure 2.4). We observed that in cells with these areas of exclusion, the regions of high DNA intensity were anti-correlated with regions of high TAg intensity. This was especially noticeable at 2 and 3 DPI. PML-NBs were rarely identified within the intense DNA signal, but rather were found to be decorating the perimeter of these areas.

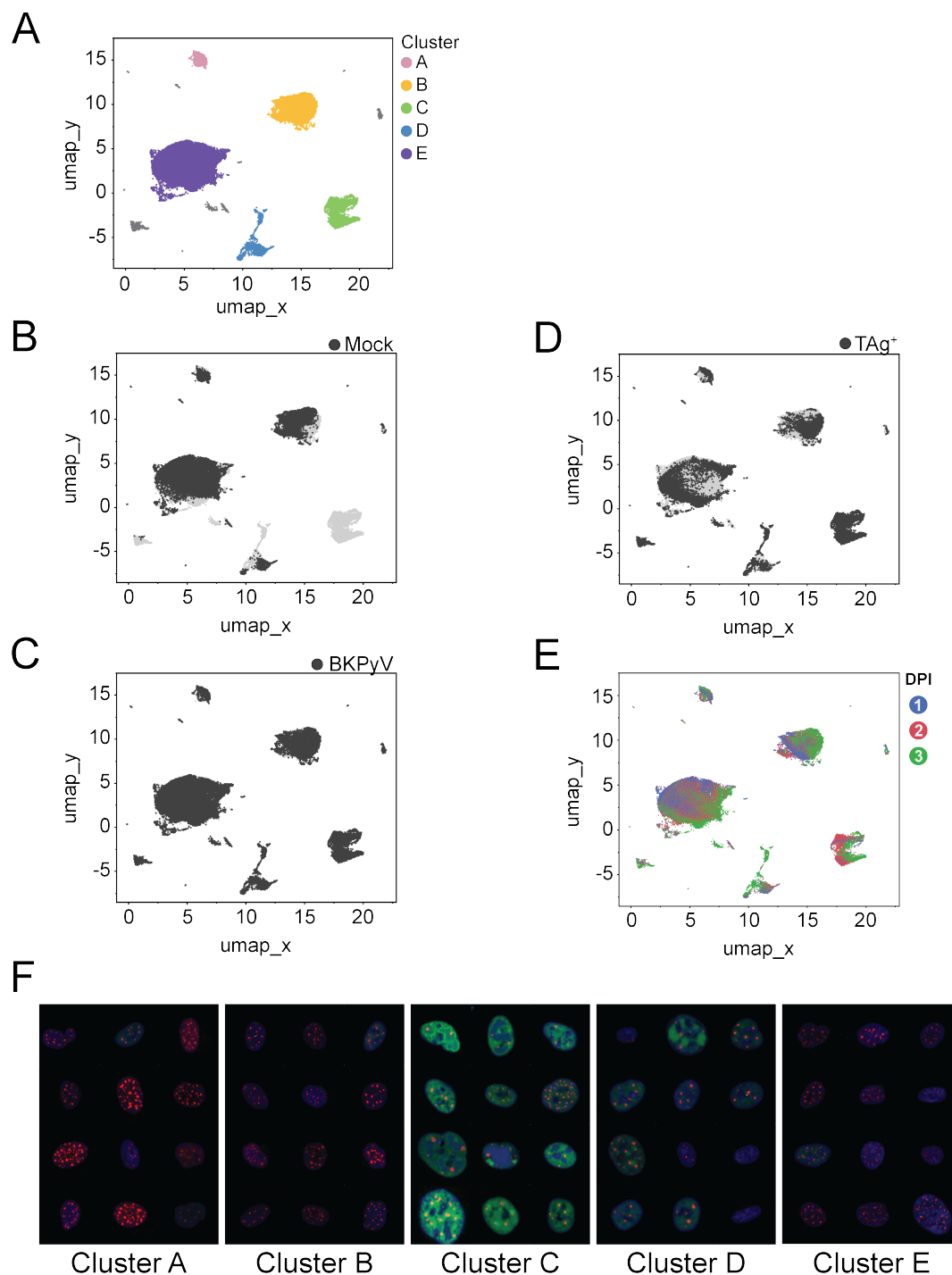


Figure 2.3 High DNA content, not time post-infection, distinguishes cells expressing the highest levels of TAG.

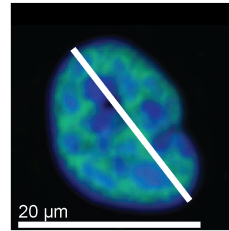
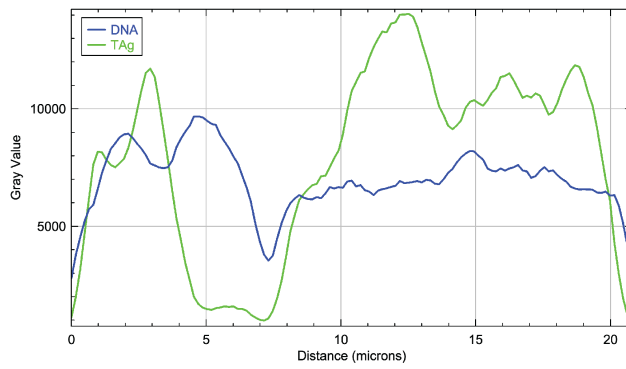
A. UMAP clusters from the global population were assigned a color and labeled with letters. B. Mock-infected cells are highlighted in dark grey. C. BKPvV-infected cells are highlighted in dark grey. D. TAG-positive cells are highlighted in dark grey. E. Cells are colored according to day post infection. Blue = 1 DPI, red = 2 DPI, green 3 DPI. F. Montages of representative nuclei from the five global clusters of cells. Channels and images are normalized to allow for comparison. Channel colors: TAG = green, PML = red, DNA = blue. Tile size is 150 pixels (px), or ~24.5 μm .

2.4.4 TAg⁺ cells form clusters defined by features other than timepoint and TAg level

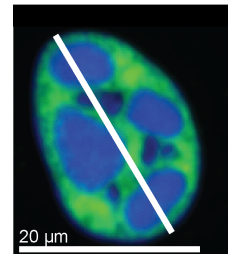
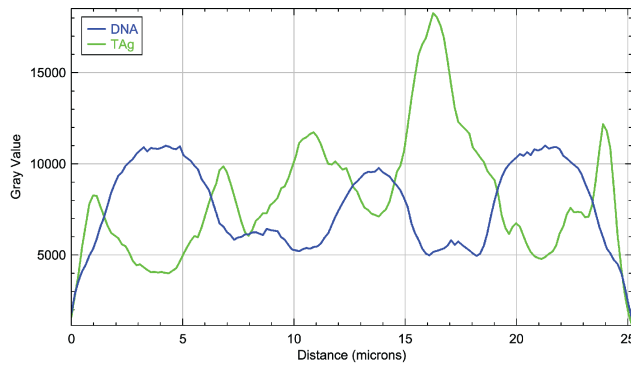
We next explored the characteristics of the TAg⁺ cells by sub-setting to TAg⁺ cells, re-embedding that population in a new UMAP, and labeling the individual clusters numerically (Figure 2.5A). Image montages for each of these eight clusters are shown in Supplemental Figure 2.9A. As with the global UMAP embedding, we wanted to see if the time post infection was influencing the structure of the clusters by coloring the individual cells to represent the days post infection (Figure 2.5B and Supplemental Figure 2.9B). We observed that most individual clusters of TAg⁺ cells were not defined by timepoint. Two exceptions were Cluster 6, which was almost entirely made up of cells at 3 days post-infection, and Cluster 4, with nearly as high a percentage of 3 dpi cells as Cluster 6 (Supplemental Table 2.7). Instead of entire clusters being comprised of a single timepoint or a random distribution of timepoints, there appeared to be distinct substructures within each cluster for each day. Table 2.2 lists the top image features that define each cluster.

To determine if the TAg⁺ cell clustering in the global UMAP was reflected in the TAg⁺-UMAP, we plotted the TAg⁺ cell data using the TAg⁺-UMAP coordinates but color-coded them using a given cell's cluster assignment from the global UMAP (Figure 2.5C). The TAg positive cells in cluster B on the global UMAP were also a single cluster on the positive-only UMAP, Cluster 1. The same was true for the TAg positive cells from global Cluster E, which were found on the TAg positive UMAP in Cluster 2. Global Cluster D was largely represented in positive Cluster 4, but was also dispersed through Clusters 3, 5, and 6. Global Cluster C, which contained only TAg⁺ cells, was present, but had become sub-divided into two new clusters, 7 and 8.

1 day post-infection



2 days post-infection



3 days post-infection

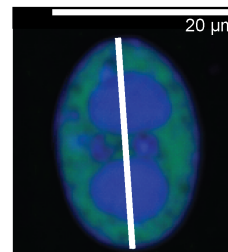
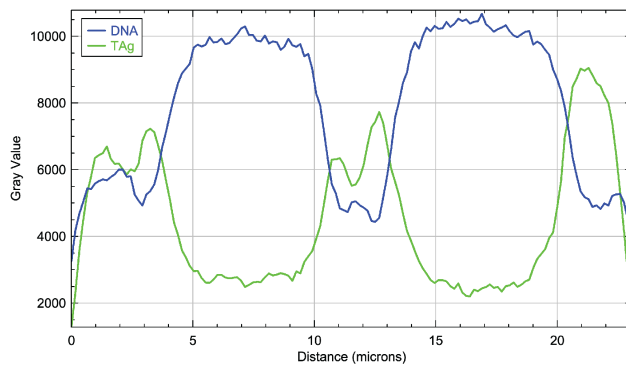


Figure 2.4 Appearance of cells with areas of mutual DNA/TAg exclusion.

A representative nucleus was selected from each timepoint and ImageJ software was used to plot the DNA and TAg intensities. Merged images of the two channels are shown. DNA = blue, TAg = green.

2.4.5 Unique subpopulations exist among cells within the same TAg expression levels

To begin to understand what was defining the clusters in the TAg⁺-only UMAP, we generated violin plots of the TAg, DNA, and PML signal intensities for each cluster (Figure 2.6A-C). We identified clusters that grouped together due to similar TAg intensities. Clusters 1, 2, and 3 all had low TAg levels, while clusters 7 and 8 had high TAg levels (Figure 2.6A). Images of the TAg signal in randomly selected cells from the low TAg group (Clusters 1–3) and the high TAg group (Clusters 7 and 8) show the differences in TAg levels between the two groups (Figure 2.6D). We next looked to see if clusters with the same TAg levels share DNA intensity levels, but none of them did (Figure 2.6B). Clusters 1, 2, and 3 were all significantly different from each other, with 2 having the highest DNA content of that group. In the high TAg group, we observed that Cluster 7 had a greater DNA intensity than Cluster 8. The PML intensity plot (Figure 2.6C) showed that the low TAg clusters, 1-3, also shared PML expression levels. Clusters 7 and 8 were significantly different, however, with Cluster 8 displaying the highest PML intensity of all the clusters.

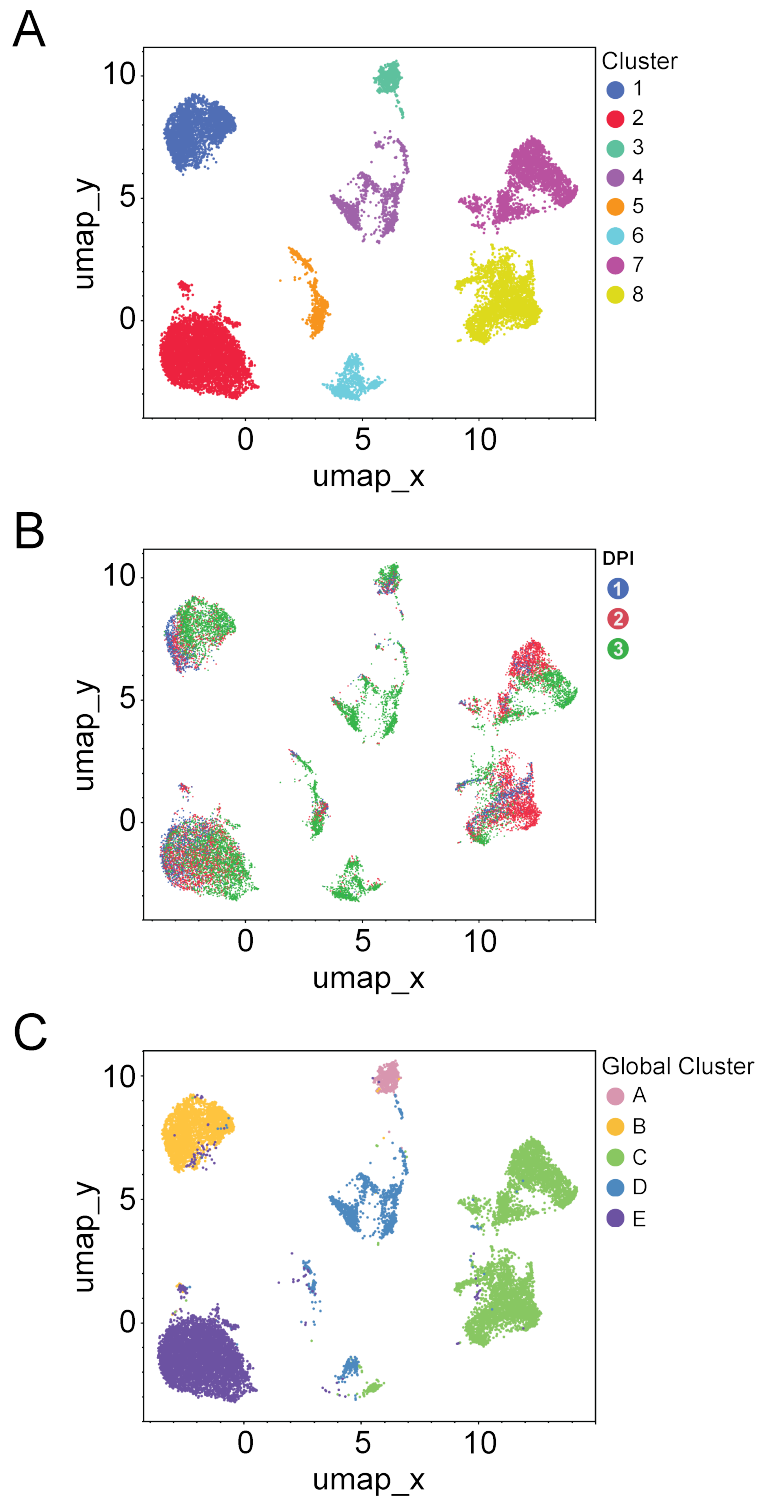


Figure 2.5 TAG⁺ cells form clusters defined by features other than timepoint and TAG level.

A. Clusters were identified from TAG-positive cells and labeled with numbers and different colors than the global clusters. B. Individual cells' timepoints were overlaid onto the positive UMAP. Blue = 1 DPI, red = 2 DPI, green 3 DPI. C. Positive UMAP with cluster colors from global UMAP overlaid.

To better quantify the relationships between TAg expression and either DNA content or PML expression in the infected population, we performed bivariate analyses (Figure 2.6E-F) to determine the correlation between either macromolecule and TAg in a given cell. As we had observed different distributions of TAg expression at different timepoints, we performed this analysis on a per-timepoint basis and for the TAg⁺ population as a whole. We observed that there were statistically significant correlations ($p < 0.0001$) between DNA and TAg intensity levels at all times post infection, but that the strengths of these correlations were not the same at each timepoint: the correlation was greatest at 2 DPI, with an R^2 value of 0.617 (Figure 2.6E). When we examined the relationship between PML and TAg levels, we found much lesser correlations (Figure 2.6F), suggesting very little of the variance seen in the PML signal was related to TAg.

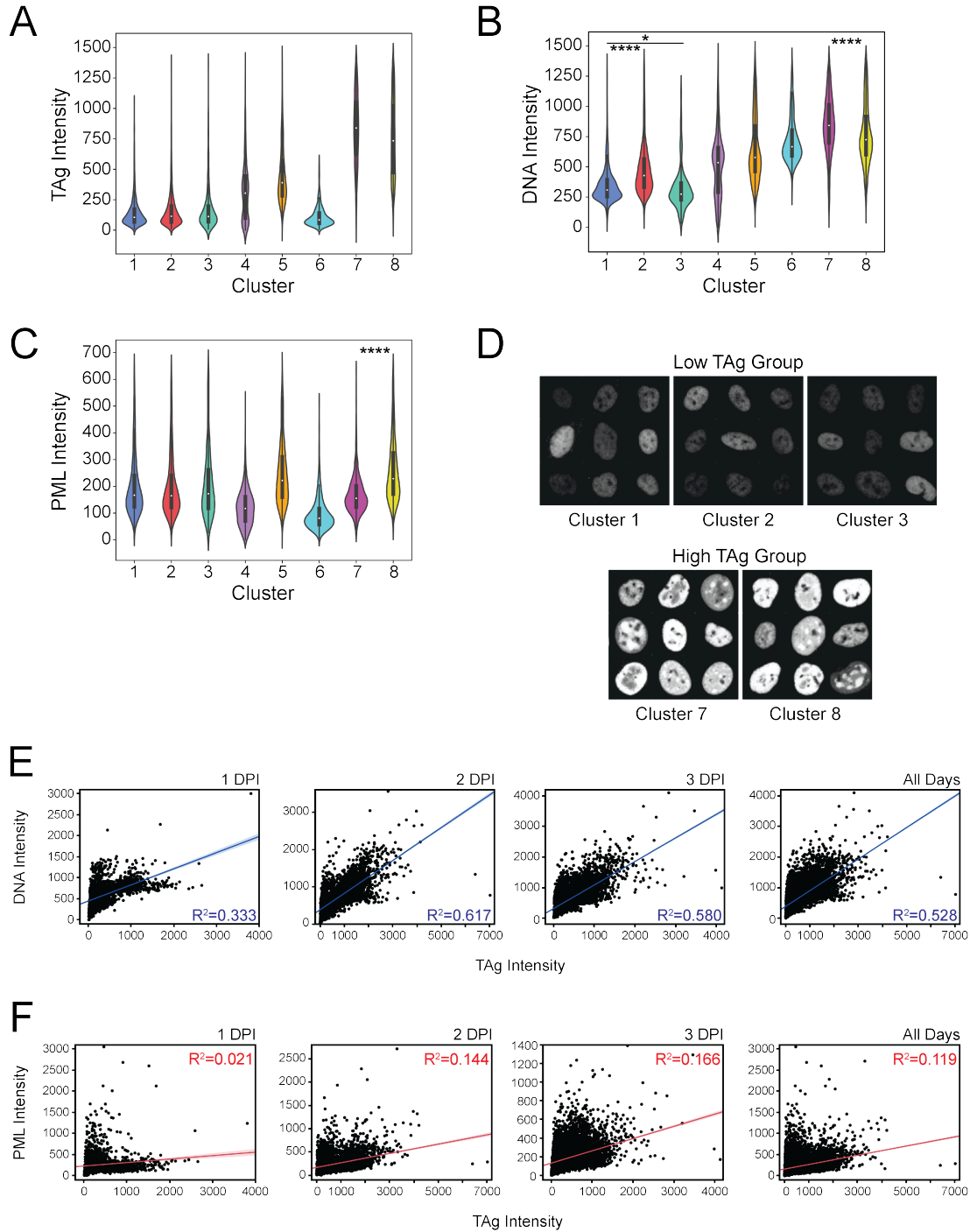


Figure 2.6 Unique subpopulations exist among cells within the same TAG expression levels.

A – C. Violin plots of per-nucleus level values for TAG-positive cells. A. TAG levels. B. DNA content. C. PML levels. D. Montages of randomly selected TAG-stained nuclei from the low TAG (Clusters 1 -3) and high TAG (Clusters 7 – 8) groups of cells. Images normalized to allow for comparison across groups. Tile size is 150 px (~ 24.5 μ m). E -F. Bivariate analyses to assess correlations at each day post infection and in the total population. R^2 values are shown. E. Correlation of DNA and TAG levels. F. Correlation of PML and TAG levels. **** $p < 0.0001$, * $p < 0.05$, Kruskal-Wallis multiple comparisons test

2.4.6 PML-NB characteristics account for differences between TAg expression groups

To explore other differences that may be present between the low and high TAg groups, we quantified nuclear area and the number of PML-NBs in each nucleus. In the low TAg group, we found that the clusters had similarly sized nuclei, but Clusters 1 and 2 were statistically different from each other, with Cluster 2 having slightly larger nuclei. Nuclei in the high TAg group were the largest, with Cluster 7 having significantly larger nuclei than Cluster 8 (Figure 2.7A and Supplemental Figure 2.10A). We quantified the number of PML-NB per nucleus and found that this number did not vary between clusters in the low TAg group, but that there was a difference in the number of PML-NBs per nucleus within the high TAg group, with fewer PML-NBs per nucleus in Cluster 7 than in Cluster 8 (Figure 2.7B and Supplemental Figure 2.10B). The difference in PML-NB numbers in the high TAg group is particularly notable in images of the DNA and PML channels (Fig. 7C). While Cluster 7 displayed a very homogeneous phenotype, Cluster 8 was more of a mixture, mostly containing cells with many PML-NBs, but with occasional low counts. This was readily apparent when graphed as a histogram (Supplemental Figure 2.10B). We also examined other features of the nuclear bodies within the individual clusters (Figure 2.7D-E). We found that the low TAg clusters had the lowest average area of PML-NB per nucleus, and that Cluster 3 had larger PML-NBs than Clusters 1 and 2. In the high TAg group, Cluster 7 had larger PML-NBs than those in Cluster 8 (Figure 2.7D). In what is likely a correlate of PML-NB size, the intensity of the PML signal of the average PML-NB per nucleus was also lowest in the low TAg group and was higher in Cluster 7 than Cluster 8 (Figure 2.7E).

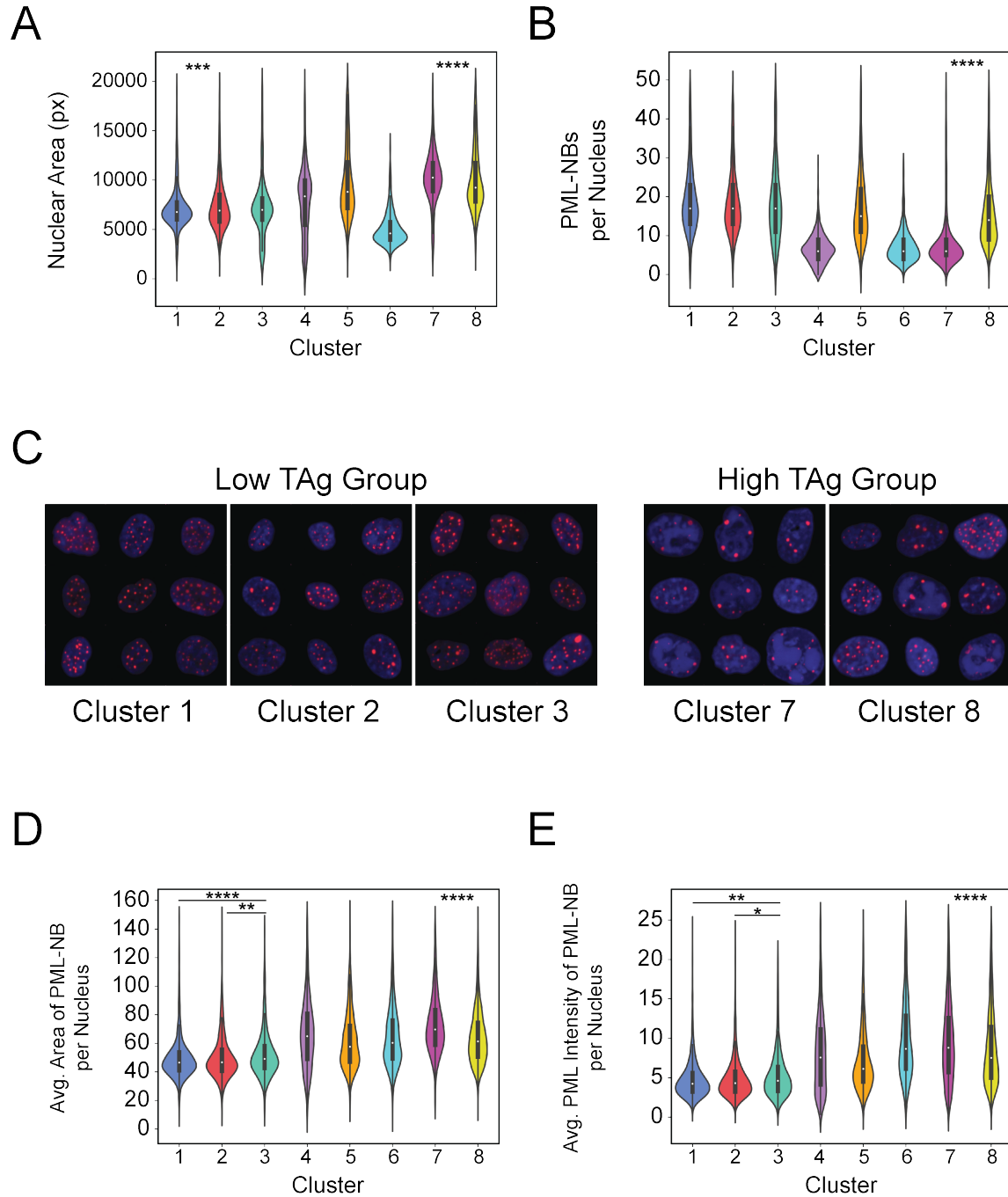


Figure 2.7 PML-NB characteristics account for differences between TAg expression groups.

A – B. Violin plots of per-nucleus values for TAg-positive cells. A. Nuclear area. B. Number of PML-NBs per nucleus. C. Montages of randomly selected nuclei from the low TAg (Clusters 1 -3) and high TAg (Clusters 7 – 8) groups of cells. Channels and images normalized to allow for comparison across groups. Channel colors: PML = red, DNA = blue. D- E. Violin plots of PML-NB level values per nucleus for TAg-positive cells. D. Average PML-NB area per nucleus. E. Average PML intensity per NB. Tile size is 150 px (~ 24.5 μm). **** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, Kruskal-Wallis multiple comparisons test.

2.5 Discussion

In establishing a single-cell high content microscopy assay of BKPyV infection, we have created a tool for studying a non-homogeneous infection. The overall characteristics of infection were consistent with our understanding of BKPyV biology. We observed an increased population of cells with >2N DNA content and fewer PML-NBs per nucleus in the TAg⁺ cells compared to the mock-infected cells. While the percentage of TAg⁺ cells increased over time, the TAg level within individual cells did not increase in a linear fashion. We found that the distribution of TAg expression levels was much more similar at 1- and 3-days post infection, with 2-days post infection being the outlier with a greater range. Overall, while there was a notable correlation between DNA and TAg levels, TAg and PML protein levels appeared to have no impact on each other.

By quantifying the percentage of TAg⁺ cells at each timepoint in the experiment we were able to see the contrast between looking at a bulk population and single cells. TAg-positivity did follow the same pattern as the percentage of cells with >2N DNA content – increasing over time. The correlation between DNA and TAg levels was not as strong as bulk studies might suggest, but the timepoint with the maximum DNA-TAg correlation was also the timepoint with maximum TAg levels. At 2-days post infection, the range of distribution of TAg levels was greater than at other timepoints, and a greater percentage of 2 DPI cells had higher TAg levels. This was dramatically different enough to appear as a second peak that was not apparent at one- or three-days post-infection. This finding was in some contrast to the Pipas lab study, which identified high- and low-viral protein-expressing populations of cells that were present by 2 dpi and remained until cell death ⁴³. In our experiment, we would have expected to see the second

peak on the histogram maintained through to the 3-day histogram (Figure 2.2D). As the two labs used different MOIs and final endpoints, direct comparisons are difficult to make.

The machine learning approach together with UMAP dimensionality reduction and clustering that we used here has provided new insights into BKPyV infection. The presence or absence of TAg did not seem to influence the structure or identity of the clusters except for Cluster C, for which TAg expression was the primary distinguishing factor. When focusing on the UMAP generated using only TAg⁺ cells, we observed that clusters can share TAg levels but have distinctions in DNA, PML, and PML-NB characteristics (Figure 2.6). What is clear from the clustering, however, is that the response of cells to BKPyV infection varies from cell to cell.

The UMAP analyses revealed that in the global population only the highest TAg levels defined a unique cluster of cells, Cluster C. The other four clusters all contained some combination of TAg⁺, mock infected, and TAg-negative cells. While time post infection does not define most individual clusters, it does form distinct substructures within the clusters. Cells from a single day post infection may not all cluster together, but within their given sub-populations they are more similar to each other than they are to other cells in that cluster from different timepoints. This influence isn't necessarily strong enough to overcome those features defining the clusters themselves, but it is shaping the architecture within them.

Our prior work on BKPyV and PML-NBs demonstrated that infection causes a reorganization of PML-NBs, with an increase in size and decrease in number of NBs per nucleus¹⁶. The prior study had also established that, in a bulk population, BKPyV infection did not change the level of PML protein, suggesting that TAg level did not correlate with PML level. Here, we found that despite having the same TAg levels, cells in the clusters had different PML

protein levels and PML-NB arrangements, indicating that TAg levels and PML parameters are not directly related. This is supported by the very low correlation between TAg and PML levels at all times post infection (Figure 2.6F). While TAg itself may not account for the changes in PML-NBs, viruses influence PML and PML-NBs through a myriad of mechanisms that could potentially be at play in our system. The simplest explanation is that PML is impacted directly by BKPyV genome replication, which is consistent with our previous findings¹⁶. Another possible explanation is that BKPyV infection may be inducing changes in PML protein and/or PML-NBs indirectly as the NBs are interacting with the DDR in response to double strand breaks and contributing to their repair. BKPyV infection induces and requires both arms of the DDR, mediated by ataxia telangiectasia mutated (ATM) and ATM and Rad3-related (ATR), for efficient replication and progeny production^{14,15}. PML-NBs co-localize with Mre11 and Nbs1, members of the MRN (Mre11, Rad50, and Nbs1) complex, which is involved in sensing of double-stranded breaks and activation of the ATM pathway⁴⁵⁻⁴⁷. Additionally, PML and PML-NBs support efficient homologous recombination^{47,48}.

In the present study, within the group of high TAg expressing clusters, there were two distinct PML-NB phenotypes: Cluster 7 displayed the previously described ‘reorganized’ PML-NB arrangement¹⁶ while Cluster 8 had a more variable phenotype, with cells more likely to have more PML-NBs. It is possible that these two phenotypes represent distinct viral interactions with the DDR. Perhaps one phenotype is associated with sensing of DNA damage and recruitment of repair proteins, while the other is associated with active repair. Future studies are needed to test this possibility.

Many cells with high TAg levels displayed a TAg and DNA exclusion phenotype, wherein there are very distinct regions of intense DNA and TAg signals that occupy non-overlapping regions of the nucleus. Cells with higher TAg and DNA content were more likely to display this morphology. This intense DNA signal could represent densely packed heterochromatin, which have been reported to be regions of heterochromatin that are functioning as phase-separated domains ^{49,50}. These membraneless structures within the nucleus have been demonstrated to regulate access to the DNA in ways beyond just volume exclusion. If the DNA is inaccessible due to phase separation, TAg may not be able to access it. This could mean that viral DNA replication, the initiation of which is mediated by TAg, is spatially separated from the host chromatin. The existence of this potential phase separation is supported by the observation that PML-NBs are generally found decorating the perimeter of these regions, rather than within them, as other proteins may not occupy the same phase as the intensely stained DNA.

There is precedent examining DNA and PML in a single-cell microscopic analysis of HPV infection, which showed great heterogeneity in infection kinetics among individual cells ⁵¹. Additionally, one of our labs has previously published results using a similar assay with SARS-CoV2 infection for a drug-repurposing screen ⁵², and other groups have published SARS-CoV2 findings as well, including another drug repurposing screen ⁵³. The single-cell assay used in our study comes with both benefits and limitations. This approach facilitates examination of BKPyV infection on a level that has not yet been widely explored. These tools have allowed for us to increase the quantity of measurements and objects measured. Using machine learning has allowed us to take an unbiased approach to scoring our cells and performing other analyses. The high-resolution confocal imaging has facilitated discovery of DNA staining patterns in the nucleus of infected cells that had not yet been reported in the BKPyV literature. In the future,

one will need to consider that observed increases in TAg levels in bulk populations may not represent equivalent increases in all the cells. At the transcriptional level, it has been demonstrated that only a small percentage of individual BKPyV positive cells express high levels of viral genes⁴³. Individual cells could be identified as high or low viral transcript producers by 2 days post infection and they did not progress beyond those states. In conjunction with our findings, the transcript data reinforce that there are likely gaps in our understanding of polyomavirus infection that can only be filled by studying single-cell infection.

There are limitations as well. For this initial study, we chose to examine only a small number of macromolecules. Therefore, we cannot draw direct conclusions about any factors beyond TAg, DNA, or PML. Additionally, while we have gained meaningful insights about differences between clusters of cells with high TAg levels, we have not yet been able to draw clear conclusions about some of the other clusters. We were able to identify that Clusters 1 – 3 all had the same quite low TAg levels, but beyond that, their differences were subtle. However, when looking at what defined them as individual clusters, all three had DNA features whose biological meaning is currently unclear at the top of their lists. Cluster 6 was even more influenced by DNA features, with its top 4 out of 5 defining characteristics being various measurements of DNA content. Cluster 4 had the greatest variety of defining characteristics, which may be reflected in its unique shape. Along with Cluster 5, it was the most challenging to interpret. Examining other parameters beyond our current three will help clarify what defines and shapes these cells in the absence of an obvious interpretable defining feature like high TAg.

Our results are only an introduction of what can be done using high-content automated microscopy and machine learning analysis. In the future we would like to expand both the size and scope of the assay to include more time points and more macromolecules, including a more

in-depth examination of the interaction between PML-NBs and the DDR. One could additionally explore the areas of intense DNA staining, paying particular attention to markers of chromatin. The addition of the single cell microscopy assay to the current arsenal of tools will facilitate a more thorough and clearer understanding of viral infections.

2.6 Materials and Methods

2.6.1 Cell culture

RPTE cells (Lonza #CC-2553) were grown in renal epithelial cell basal medium (REBM™, Lonza #CC-3191) supplemented with SingleQuots™ (Lonza #CC-4127) to make renal epithelial cell growth medium (REGM™) and passaged up to six times as previously described ⁵⁴. RPTE-hTERT cells ⁵⁵ were also maintained in REBM/REGM.

293TT cells ⁵⁶ were maintained in DMEM with 10% fetal bovine serum and 100 U/ml penicillin, 100 µg/ml streptomycin as described in Broekema, et al. ⁵⁷. All cells were grown in a humidified incubator at 37°C and 5% CO₂.

2.6.2 Virus growth and titration

The BKPyV Dunlop variant was propagated as described previously ⁵⁷, with some modifications. Briefly, the linearized Dunlop genome was digested from the pGEM backbone at flanking BamHI sites. Gel purification was performed followed by circularization of the genome. The remainder of the protocol proceeded as described.

Viral titers were determined using a modified version of the previously described fluorescent focus assay (FFA) ⁵⁴. Sub-confluent RPTE-hTERT cells were seeded on tissue-culture treated glass chamber slides. The next day, the cells were infected with either a media only control or ten-fold serial dilutions of purified virus, with a one-hour adsorption at 4°C, and media replacement following that. Forty-eight hours post infection the cells were fixed with 4% paraformaldehyde and probed for TAg (PAb416, 1:500 dilution). After the cells were incubated

with the secondary antibody, DNA was stained with 1 µg/ml Hoechst 33342 and coverslips were mounted using Prolong Glass Antifade mountant (*Invitrogen #P36982*). Images were acquired at 20X using an Olympus BX60 camera. ImageJ software ⁵⁸ was used to quantify the number of TAg positive cells per well. Cells were only considered positive if the TAg signal was brighter than any found in the mock control and if it was present in the nucleus. Viral titers were calculated as the mean of three wells.

2.6.3 Infection and Immunofluorescent staining

Using 96 well-plates with optically clear bottoms (Greiner, #655090), 8,000 RPTE-hTERT cells per well were seeded, avoiding the perimeter wells. The outside wells were filled with PBS to prevent evaporation of media from those containing cells. The day after seeding, the cells were incubated at 4°C for 15 min, then either mock-infected or infected at an MOI of 0.5 in 40 µl per well of REGM. The cells were returned to 4°C for a one-hour adsorption. After the adsorption, the inoculum was aspirated from the wells and replaced with 100 µl of pre-warmed REGM per well. The cells were returned to 37°C for one, two, or three days. At the appropriate timepoint, media was aspirated from cells and replaced with room temperature 4% paraformaldehyde for a 20-minute incubation (40 µl per well). Wells were washed thoroughly with PBS after fixation, especially if the plate was being returned to the incubator for later timepoints. Following the fixation and wash, the cells were permeabilized and blocked for four hours in 40 µl per well of Blocking Buffer at room temperature. The buffer was aspirated from the cells and replaced with 40 µl per well of primary antibodies in antibody dilution buffer. The cells were incubated in primary antibody overnight at 4°C. The following day, cells were washed

three times in PBS, then incubated with secondary antibodies and 10 µg/ml Hoechst 33342 for two hours at room temperature. After this final incubation step the cells were washed three more times with PBS, then the wells were filled with 100 µl PBS each for imaging. Details for all antibodies used in this study can be found in Table 2.1.

For all immunofluorescence, blocking steps were performed using Blocking Buffer (1X PBS/5% Goat Serum/0.3% Triton™ X-100) and antibody dilution steps were performed in Antibody Dilution Buffer (1X PBS/1% BSA/0.3% Triton™ X-100). All washes were done using 1X PBS without calcium or magnesium.

2.6.4 High-Content Imaging

All imaging was performed using a confocal Yokogawa Cell Voyager 8000 (CV8000) high content imaging system with a water immersion UPLAPO 40X/1.0NA objective. The optical configuration and fluorophores for the timecourse experiment were as follows: nuclei - Hoechst-33342 (405nm laser, 445/45 emission filter), TAg - Alexa-488 (488nm laser, 525/50 emission filter), PML - Alexa-594 (561nm laser, 600/37 emission filter). Laser power and exposure times were adjusted to optimize signal-to-background. Due to the high NA objectives and thin depth of focus, maximum projection images were collected over a 10 µm Z-range at 1 µm intervals to ensure that all cells were in focus. Thirty-two fields per well were collected to ensure an adequate number of cell observations. Post-acquisition background and geometry correction was performed for each channel.

2.6.5 Image Identification, Feature Extraction, Infection Classification, and Curve Fitting

Individual nuclei in the Hoechst-33342 images were identified and nuclear masks were generated using open-source CellPose 2.0 software; with a mean cell diameter of 110, CellPose's native edge removal to prevent generating masks of partial cells, and the built-in Cyto-2 pretrained model ^{59,60}. Nuclear masks were imported into CellProfiler 4.2.1 along with the fluorescent images for feature extraction ⁶¹. PML-NBs were identified within the nuclei using Otsu-adaptive two-class thresholding with a size range of 3-20 pixels. Intensity features were tabulated for each fluorescent channel along with size/shape measurements for nuclei and PML-NBs, resulting in 668 total features per nucleus.

All cells were classified as TAg positive or negative with CellProfiler Analyst 3.0.4 ⁶² using a semi-supervised machine learning approach. Approximately 100 each of TAg positive and negative cells were hand selected, and a Random Forest (RF) classifier model was trained using 5-fold cross-validation. The classification model resulted in high accuracy (>99.6%) on the test set. The RF feature importance was tabulated and, as expected, the top features included those that measured TAg signal intensity.

For the signal intensity plots, single channel grayscale images from either the DNA or TAg channels were opened using ImageJ software. A bisecting line was drawn down the longest axis of the selected cell at the same coordinates in each image. The PlotProfile function was run for both images and a plot was generated.

2.6.6 Data processing and UMAP embedding

To systematically explore the morphologies of BKPyV infected cells, UMAP dimensionality reduction was performed similar to Mirabelli, et. al.⁵². The embed_UMAP application of MPLearn (v0.1.0; <https://github.com/maomlab/MPLearn>) was used to generate UMAP embeddings. Briefly, for a set of cells, each feature was standardized using the scikit-learn StandardScaler. Then, features were reduced to two dimensions using the umap-learn python package originally described in McInnes, et al. with the defaults listed in the documentation⁶³. Embeddings were visualized using JMP 16 software, and Cellprofiler Analyst was used to generate single-nucleus tiled images of selected clusters. The subset of TAg⁺ cells was selected and re-embedded to further characterize the morphologic and intensity characteristics of the infected cell population.

2.6.7 Statistical Testing

Wilcoxon one-way and Kruskal-Wallis multiple comparison analyses were performed using either JMP or Prism.

2.7 Acknowledgements

We would like to thank Wei Zou and Jesse Wotring for their scientific and moral support, and Andrew Tai and Christiane Wobus for their critical reviews of this manuscript. This work was funded in part by R01 AI060584 to M.J.I. and a University of Michigan Medical School Office of Research Pandemic Research Recovery Award to M.C.P.

2.8 Tables

Table 2.1 Antibodies used in this study

Primary Antibodies

Name	Host	Dilution	Use	Source
PAb416 (anti-TAg)	mouse	1:250-1:1000	IFA and FFA	Harlow, et al. ⁶⁴
PML Ab (PG-M3)	mouse	1:250-1:500	IFA	Santa Cruz, sc-966

Secondary Antibodies

Name	Host	Dilution	Use	Source
Anti-rabbit-IgG-Alexa Fluor 594	goat	1:250-1:1000	IFA	Invitrogen, A11012
Anti-mouse-IgG-Alexa Fluor 488	goat	1:250-1:1000	IFA	Invitrogen, A11001
Anti-mouse-IgG-Alexa Fluor 594	goat	1:250-1:1000	FFA	Invitrogen, A11005

Table 2.2 Major defining features of TAg⁺ clusters

Cluster 1	lower DNA intensity (x3), lower TAg intensity
Cluster 2	lower DNA intensity, lower DNA granularity, lower TAg intensity (x2), more variation in PML signal
Cluster 3	lower DNA intensity (x3), higher DNA granularity, smaller nuclei
Cluster 4	fewer PML-NBs/nucleus, slightly higher TAg granularity, lower PML intensity, less variation in PML signal, lower DNA intensity
Cluster 5	slightly higher TAg intensity (x3), more PML-NBs per nucleus
Cluster 6	higher DNA intensity (x4), lower PML intensity
Cluster 7	high TAg intensity (x2), PML signal infrequently overlaps with DNA or TAg, fewer PML-NBs per nucleus
Cluster 8	high TAg intensity (x2), lower TAg granularity, higher PML intensity, slightly more PML-NBs/nucleus

Supplemental Table 2.3 Nuclear measurements obtained by Cell Profiler software*

AreaShape	AreaShape	AreaShape	Correlation
<i>Area</i>	<i>MajorAxisLength</i>	<i>Zernike_0_0</i>	Correlation_DNA_PML
<i>BoundingBoxArea</i>	<i>MaxFeretDiameter</i>	<i>Zernike_1_1</i>	Correlation_DNA_TAg
<i>BoundingBoxMaximum_X</i>	<i>MaximumRadius</i>	<i>Zernike_2_0</i>	Correlation_PML_TAg
<i>BoundingBoxMaximum_Y</i>	<i>MeanRadius</i>	<i>Zernike_2_2</i>	Costes_DNA_PML
<i>BoundingBoxMinimum_X</i>	<i>MedianRadius</i>	<i>Zernike_3_1</i>	Costes_DNA_TAg
<i>BoundingBoxMinimum_Y</i>	<i>MinFeretDiameter</i>	<i>Zernike_3_3</i>	Costes_PML_DNA
<i>CentralMoment_0_0</i>	<i>MinorAxisLength</i>	<i>Zernike_4_0</i>	Costes_PML_TAg
<i>CentralMoment_0_1</i>	<i>NormalizedMoment_0_0</i>	<i>Zernike_4_2</i>	Costes_TAg_DNA
<i>CentralMoment_0_2</i>	<i>NormalizedMoment_0_1</i>	<i>Zernike_4_4</i>	Costes_TAg_PML
<i>CentralMoment_0_3</i>	<i>NormalizedMoment_0_2</i>	<i>Zernike_5_1</i>	K_DNA_PML
<i>CentralMoment_1_0</i>	<i>NormalizedMoment_0_3</i>	<i>Zernike_5_3</i>	K_DNA_TAg
<i>CentralMoment_1_1</i>	<i>NormalizedMoment_1_0</i>	<i>Zernike_5_5</i>	K_PML_DNA
<i>CentralMoment_1_2</i>	<i>NormalizedMoment_1_1</i>	<i>Zernike_6_0</i>	K_PML_TAg
<i>CentralMoment_1_3</i>	<i>NormalizedMoment_1_2</i>	<i>Zernike_6_2</i>	K_TAg_DNA
<i>CentralMoment_2_0</i>	<i>NormalizedMoment_1_3</i>	<i>Zernike_6_4</i>	K_TAg_PML
<i>CentralMoment_2_1</i>	<i>NormalizedMoment_2_0</i>	<i>Zernike_6_6</i>	Manders_DNA_PML
<i>CentralMoment_2_2</i>	<i>NormalizedMoment_2_1</i>	<i>Zernike_7_1</i>	Manders_DNA_TAg
<i>CentralMoment_2_3</i>	<i>NormalizedMoment_2_2</i>	<i>Zernike_7_3</i>	Manders_PML_DNA
<i>Compactness</i>	<i>NormalizedMoment_2_3</i>	<i>Zernike_7_5</i>	Manders_PML_TAg
<i>ConvexArea</i>	<i>NormalizedMoment_3_0</i>	<i>Zernike_7_7</i>	Manders_TAg_DNA
<i>Eccentricity</i>	<i>NormalizedMoment_3_1</i>	<i>Zernike_8_0</i>	Manders_TAg_PML
<i>EquivalentDiameter</i>	<i>NormalizedMoment_3_2</i>	<i>Zernike_8_2</i>	Overlap_DNA_PML
<i>EulerNumber</i>	<i>NormalizedMoment_3_3</i>	<i>Zernike_8_4</i>	Overlap_DNA_TAg
<i>Extent</i>	<i>Orientation</i>	<i>Zernike_8_6</i>	Overlap_PML_TAg
<i>FormFactor</i>	<i>Perimeter</i>	<i>Zernike_8_8</i>	RWC_DNA_PML
<i>HuMoment_0</i>	<i>Solidity</i>	<i>Zernike_9_1</i>	RWC_DNA_TAg
<i>HuMoment_1</i>	<i>SpatialMoment_0_0</i>	<i>Zernike_9_3</i>	RWC_PML_DNA
<i>HuMoment_2</i>	<i>SpatialMoment_0_1</i>	<i>Zernike_9_5</i>	RWC_PML_TAg
<i>HuMoment_3</i>	<i>SpatialMoment_0_2</i>	<i>Zernike_9_7</i>	RWC_TAg_DNA
<i>HuMoment_4</i>	<i>SpatialMoment_0_3</i>	<i>Zernike_9_9</i>	RWC_TAg_PML
<i>HuMoment_5</i>	<i>SpatialMoment_1_0</i>		
<i>HuMoment_6</i>	<i>SpatialMoment_1_1</i>		
<i>InertiaTensorEigenvalues_0</i>	<i>SpatialMoment_1_2</i>		
<i>InertiaTensorEigenvalues_1</i>	<i>SpatialMoment_1_3</i>		
<i>InertiaTensor_0_0</i>	<i>SpatialMoment_2_0</i>		
<i>InertiaTensor_0_1</i>	<i>SpatialMoment_2_1</i>		
<i>InertiaTensor_1_0</i>	<i>SpatialMoment_2_2</i>		
<i>InertiaTensor_1_1</i>	<i>SpatialMoment_2_3</i>		

*Each measurement was obtained once for each object. Those listed in italics were also obtained for each PML-NB.

Supplemental Table 2.4 Nuclear measurements obtained by Cell Profiler software[#]

Granularity	Intensity	RadialDistribution	Texture*
Granularity_1	<i>IntegratedIntensityEdge</i>	FracAtD_1of4	AngularSecondMoment
Granularity_2	<i>IntegratedIntensity</i>	FracAtD_2of4	Contrast
Granularity_3	<i>LowerQuartileIntensity</i>	FracAtD_3of4	Correlation
Granularity_4	<i>MADIntensity</i>	FracAtD_4of4	DifferenceEntropy
Granularity_5	<i>MassDisplacement</i>	MeanFrac_1of4	DifferenceVariance
Granularity_6	<i>MaxIntensityEdge</i>	MeanFrac_2of4	Entropy
Granularity_7	<i>MaxIntensity</i>	MeanFrac_3of4	InfoMeas1
Granularity_8	<i>MeanIntensityEdge</i>	MeanFrac_4of4	InfoMeas2
Granularity_9	<i>MeanIntensity</i>	RadialCV_1of4	InverseDifferenceMoment
Granularity_10	<i>MedianIntensity</i>	RadialCV_2of4	SumAverage
	<i>MinIntensityEdge</i>	RadialCV_3of4	SumEntropy
	<i>MinIntensity</i>	RadialCV_4of4	SumVariance
	<i>StdIntensityEdge</i>	ZernikeMagnitude_0_0	Variance
	<i>StdIntensity</i>	ZernikeMagnitude_1_1	
	<i>UpperQuartileIntensity</i>	ZernikeMagnitude_2_0	
		ZernikeMagnitude_2_2	
		ZernikeMagnitude_3_1	
		ZernikeMagnitude_3_3	
		ZernikeMagnitude_4_0	
		ZernikeMagnitude_4_2	
		ZernikeMagnitude_4_4	
		ZernikeMagnitude_5_1	
		ZernikeMagnitude_5_3	
		ZernikeMagnitude_5_5	
		ZernikeMagnitude_6_0	
		ZernikeMagnitude_6_2	
		ZernikeMagnitude_6_4	
		ZernikeMagnitude_6_6	

[#]Each measurement was obtained once for each object in each channel. Those listed in italics were also obtained for each PML-NB.

* All textures measured at a scale of 4 and 12 pixels and in all four directions

Supplemental Table 2.5 PML-NB measurements obtained by Cell Profiler software

Location	Neighbors	RadialDistribution
CenterMassIntensity_X	AngleBetweenNeighbors_Expanded	FracAtD_PML_1of3
CenterMassIntensity_Y	FirstClosestDistance_Expanded	FracAtD_PML_2of3
MaxIntensity_X	FirstClosestObjectNumber_Expanded	FracAtD_PML_3of3
MaxIntensity_Y	NumberOfNeighbors_Expanded	MeanFrac_PML_1of3
	PercentTouching_Expanded	MeanFrac_PML_2of3
	SecondClosestDistance_Expanded	MeanFrac_PML_3of3
	SecondClosestObjectNumber_Expanded	RadialCV_PML_1of3
		RadialCV_PML_2of3
		RadialCV_PML_3of3

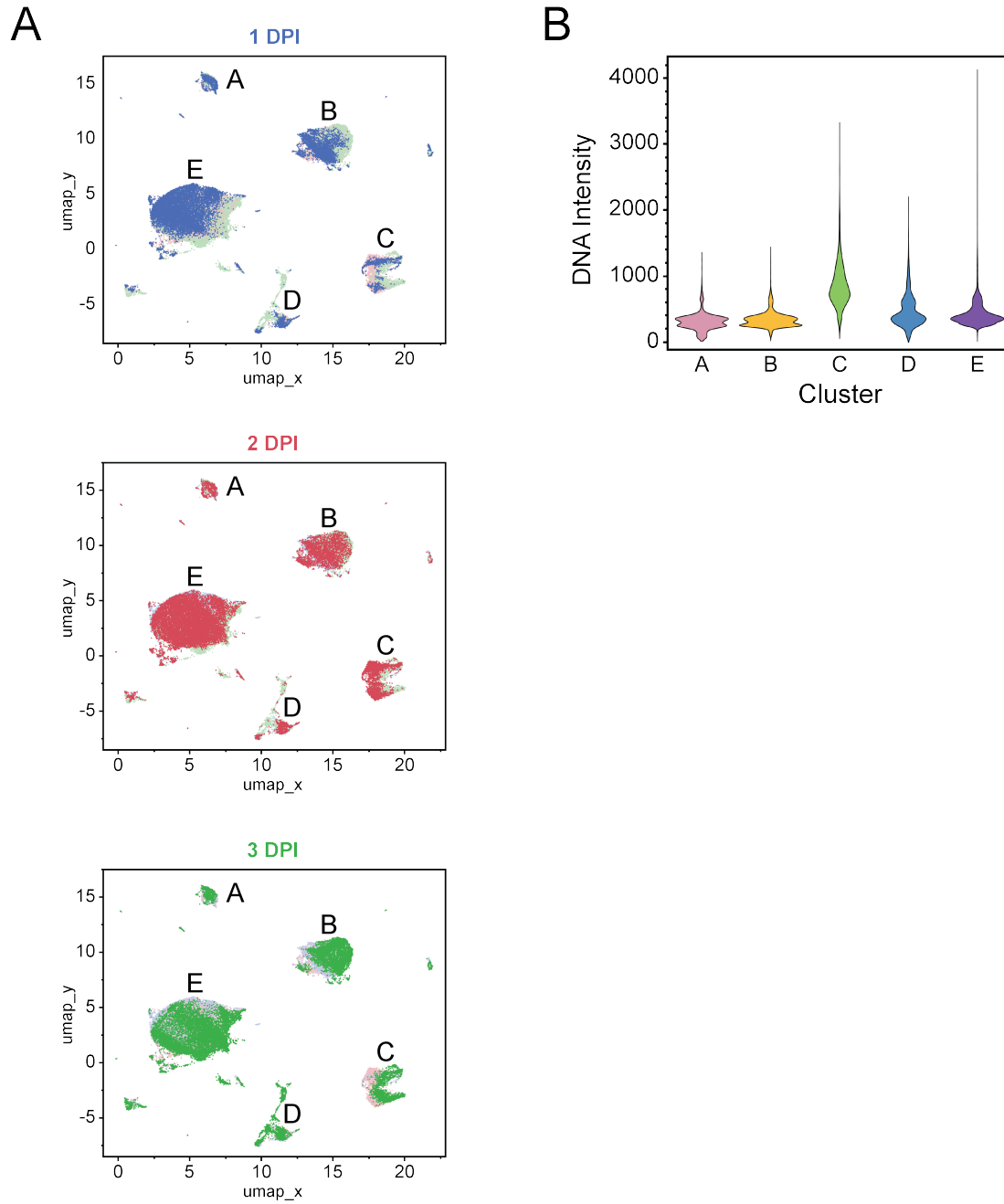
Supplemental Table 2.6 Percentages of cells at the different time points in each global cluster

Cluster	A	B	C	D	E
1 DPI	39%	32%	13%	23%	32%
2 DPI	20%	24%	48%	18%	36%
3 DPI	42%	44%	38%	59%	32%

Supplemental Table 2.7 Percentages of cells at the different time points in each TAG⁺ cluster

Cluster	1	2	3	4	5	6	7	8
1 DPI	19%	17%	25%	4%	14%	1%	8%	20%
2 DPI	21%	32%	17%	6%	22%	4%	41%	57%
3 DPI	61%	51%	59%	90%	65%	95%	51%	24%

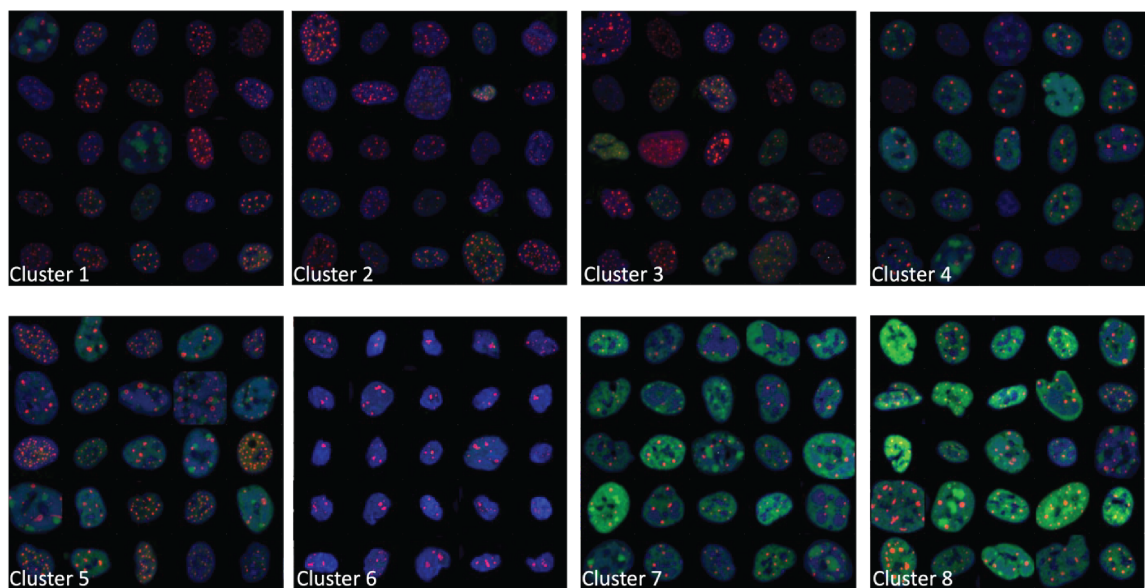
2.9 Supplemental Figures



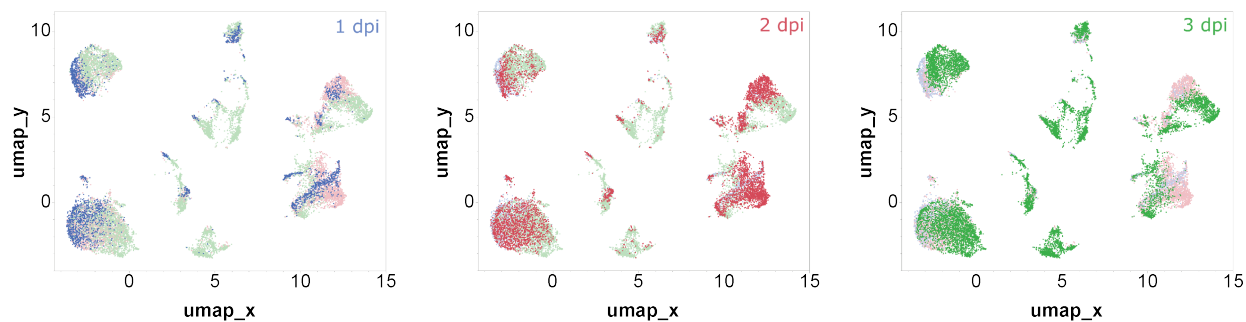
Supplemental Figure 2.8 Cluster C contains more cells at 2 dpi.

A. Global UMAP with cells colored by day post infection as in figure 3E. Each day has been highlighted separately to better demonstrate the distribution of timepoints within clusters. B. Violin plot of per-nucleus DNA content in each cluster.

A

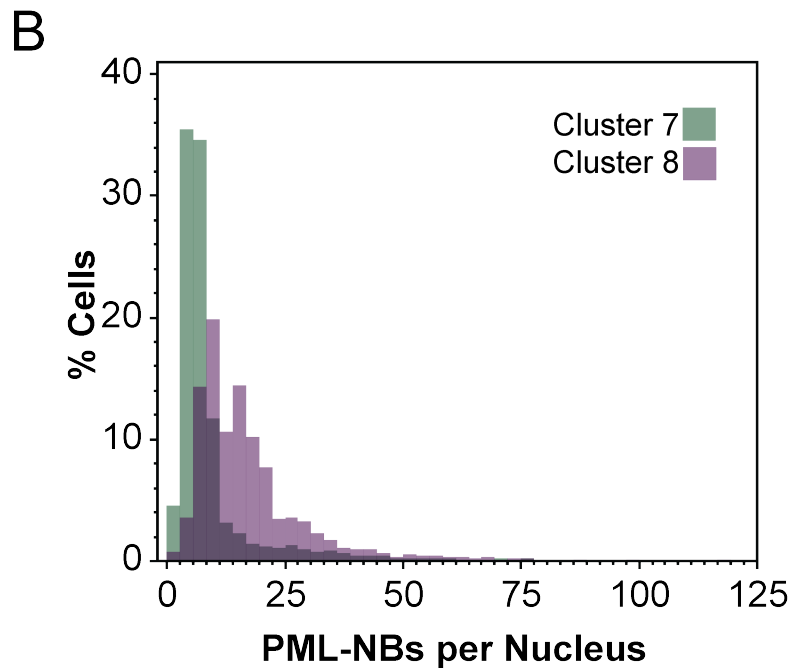
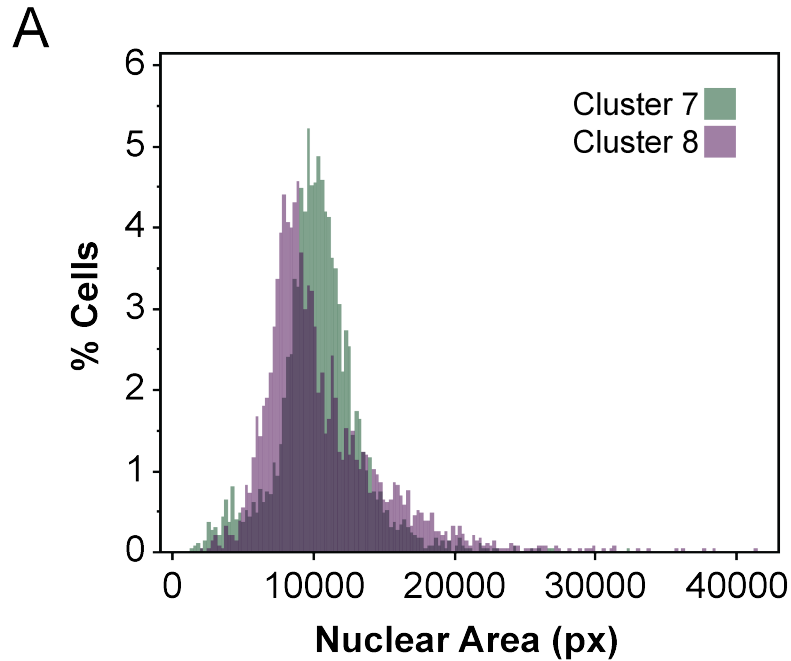


B



Supplemental Figure 2.9 Distribution of TAG-positive cells in the clusters.

A. Montages of representative nuclei from the eight TAG-positive clusters of cells. Channels and images are normalized to allow for comparison. Channel colors: TAG = green, PML = red, DNA = blue. Tile size is 150 px (~24.5 μ m). B. TAG-positive UMAPs with cells colored by day post infection as in figure 5B. Each day has been highlighted separately to better demonstrate the distribution of timepoints within clusters.



Supplemental Figure 2.10 Clusters in the high TAG group differ in multiple ways including nuclear area and PML-NB quantity.

A. Histogram depicting the area of the nuclei demonstrated in the violin plot in figure 7A, but only from Clusters 7 and 8. B Histogram depicting the quantity of PML-NBs per nucleus as was graphed via violin plot in figure 7B, only for Clusters 7 and 8. In both histograms Cluster 7 is represented in green, while Cluster 8 is represented in purple.

2.10 Bibliography

1. Kamminga, S., Meijden, E. van der, Feltkamp, M. C. W. & Zaaijer, H. L. Seroprevalence of fourteen human polyomaviruses determined in blood donors. *Plos One* **13**, e0206273 (2018).
2. Laine, H. K. *et al.* Seroprevalence of polyomaviruses BK and JC in Finnish women and their spouses followed-up for three years. *Sci Rep-uk* **13**, 879 (2023).
3. Prezioso, C. & Pietropaolo, V. BK Virus and Transplantation. *Viruses* **13**, 733 (2021).
4. Hirsch, H. H. *et al.* Prospective Study of Polyomavirus Type BK Replication and Nephropathy in Renal-Transplant Recipients. *New Engl J Medicine* **347**, 488–496 (2002).
5. Hirsch, H. H., Randhawa, P. S. & Practice, A. I. D. C. of. BK polyomavirus in solid organ transplantation—Guidelines from the American Society of Transplantation Infectious Diseases Community of Practice. *Clin Transplant* **33**, e13528 (2019).
6. Knoll, G. A. *et al.* Levofloxacin for BK Virus Prophylaxis Following Kidney Transplantation: A Randomized Clinical Trial. *Jama* **312**, 2106–2114 (2014).
7. Mallat, S. G. *et al.* CMV and BKPyV Infections in Renal Transplant Recipients Receiving an mTOR Inhibitor–Based Regimen Versus a CNI-Based Regimen: A Systematic Review and Meta-Analysis of Randomized, Controlled Trials. *Clin J Am Soc Nephro* **12**, 1321–1336 (2017).
8. Bischof, N. *et al.* Reducing calcineurin inhibitor first for treating BK polyomavirus replication after kidney transplantation: long-term outcomes. *Nephrol Dial Transpl* **34**, 1240–1250 (2018).
9. Drachenberg, C. B. *et al.* Histological Evolution of BK Virus–Associated Nephropathy: Importance of Integrating Clinical and Pathological Findings. *Am J Transplant* **17**, 2078–2091 (2017).
10. Mohammad, D., Kim, D. Y., Baracco, R., Kapur, G. & Jain, A. Treatment of BK virus with a stepwise immunosuppression reduction and intravenous immunoglobulin in pediatric kidney transplant. *Pediatr Transplant* **26**, e14241 (2022).
11. Abend, J. R., Joseph, A. E., Das, D., Campbell-Cecen, D. B. & Imperiale, M. J. A truncated T antigen expressed from an alternatively spliced BK virus early mRNA. *J Gen Virol* **90**, 1238–1245 (2009).
12. Nomburg, J. *et al.* Long-read sequencing reveals complex patterns of wraparound transcription in polyomaviruses. *Plos Pathog* **18**, e1010401 (2022).

13. DeCaprio, J. A., Imperiale, M. J. & Hirsch, H. H. Chapter 1 Polyomaviridae. in *Fields Virology: DNA Viruses* (eds. Howley, P. M. & Knipe, D. M.) 1–44 (Lippincott Williams & Wilkins, 2022).
14. Jiang, M., Zhao, L., Gamez, M. & Imperiale, M. J. Roles of ATM and ATR-Mediated DNA Damage Responses during Lytic BK Polyomavirus Infection. *Plos Pathog* **8**, e1002898 (2012).
15. Verhalen, B., Justice, J. L., Imperiale, M. J. & Jiang, M. Viral DNA Replication-Dependent DNA Damage Response Activation during BK Polyomavirus Infection. *J Virol* **89**, 5032–5039 (2015).
16. Jiang, M., Entezami, P., Gamez, M., Stamminger, T. & Imperiale, M. J. Functional Reorganization of Promyelocytic Leukemia Nuclear Bodies during BK Virus Infection. *Mbio* **2**, e00281-10 (2011).
17. Ishov, A. M. *et al.* Pml Is Critical for Nd10 Formation and Recruits the Pml-Interacting Protein Daxx to This Nuclear Structure When Modified by Sumo-1. *J Cell Biology* **147**, 221–234 (1999).
18. Negorev, D. & Maul, G. G. Cellular proteins localized at and interacting within ND10/PML nuclear bodies/PODs suggest functions of a nuclear depot. *Oncogene* **20**, 7234–7242 (2001).
19. Eskiw, C. H., Dellaire, G. & Bazett-Jones, D. P. Chromatin Contributes to Structural Integrity of Promyelocytic Leukemia Bodies through a SUMO-1-independent Mechanism*. *J Biol Chem* **279**, 9577–9585 (2004).
20. Bazett-Jones, D. P., Li, R., Fussner, E., Nisman, R. & Dehghani, H. Elucidating chromatin and nuclear domain architecture with electron spectroscopic imaging. *Chromosome Res* **16**, 397 (2008).
21. Zhong, S. *et al.* A role for PML and the nuclear body in genomic stability. *Oncogene* **18**, 7941–7947 (1999).
22. Everett, R. D., Lomonte, P., Sternsdorf, T., Driel, R. van & Orr, A. Cell cycle regulation of PML modification and ND10 composition. *J Cell Sci* **112**, 4581–4588 (1999).
23. Zhong, S. *et al.* Role of SUMO-1–modified PML in nuclear body formation. *Blood* **95**, 2748–2752 (2000).
24. Brazina, J. *et al.* DNA damage-induced regulatory interplay between DAXX, p53, ATM kinase and Wip1 phosphatase. *Cell Cycle* **14**, 375–387 (2015).
25. Masi, A. di *et al.* PML nuclear body disruption impairs DNA double-strand break sensing and repair in APL. *Cell Death Dis* **7**, e2308–e2308 (2016).

26. Scherer, M. *et al.* Dual signaling via interferon and DNA damage response elicits entrapment by giant PML nuclear bodies. *Elife* **11**, e73006 (2022).
27. Neerukonda, S. N. Interplay between RNA Viruses and Promyelocytic Leukemia Nuclear Bodies. *Vet Sci* **8**, 57 (2021).
28. Guion, L. G. & Sapp, M. The Role of Promyelocytic Leukemia Nuclear Bodies During HPV Infection. *Front Cell Infect Mi* **10**, 35 (2020).
29. Hofmann, S., Stubbe, M., Mai, J. & Schreiner, S. Double-edged role of PML nuclear bodies during human adenovirus infection. *Virus Res* **295**, 198280 (2021).
30. Scherer, M. & Stamminger, T. Emerging Role of PML Nuclear Bodies in Innate Immune Signaling. *J Virol* **90**, 5850–5854 (2016).
31. Low, J., Humes, H. D., Szczypka, M. & Imperiale, M. BKV and SV40 infection of human kidney tubular epithelial cells in vitro. *Virology* **323**, 182–188 (2004).
32. Zhao, L. & Imperiale, M. J. A Cell Culture Model of BK Polyomavirus Persistence, Genome Recombination, and Reactivation. *Mbio* **12**, e02356-21 (2021).
33. Jiang, M., Abend, J. R., Tsai, B. & Imperiale, M. J. Early Events during BK Virus Entry and Disassembly. *J Virol* **83**, 1350–1358 (2009).
34. Bennett, S. M., Jiang, M. & Imperiale, M. J. Role of Cell-Type-Specific Endoplasmic Reticulum-Associated Degradation in Polyomavirus Trafficking. *J Virol* **87**, 8843–8852 (2013).
35. Zhao, L. & Imperiale, M. J. Identification of Rab18 as an Essential Host Factor for BK Polyomavirus Infection Using a Whole-Genome RNA Interference Screen. *Msphere* **2**, e00291-17 (2017).
36. Zhao, L., Marciano, A. T., Rivet, C. R. & Imperiale, M. J. Caveolin- and clathrin-independent entry of BKPyV into primary human proximal tubule epithelial cells. *Virology* **492**, 66–72 (2016).
37. Gosert, R. *et al.* Polyomavirus BK with rearranged noncoding control region emerge in vivo in renal transplant patients and increase viral replication and cytopathology. *J Exp Medicine* **205**, 841–852 (2008).
38. Bethge, T. *et al.* Sp1 Sites in the Noncoding Control Region of BK Polyomavirus Are Key Regulators of Bidirectional Viral Early and Late Gene Expression. *J Virol* **89**, 3396–3411 (2015).
39. Manzetti, J. *et al.* BK Polyomavirus Evades Innate Immune Sensing by Disrupting the Mitochondrial Network and Promotes Mitophagy. *Isience* **23**, 101257 (2020).

40. Panou, M.-M. *et al.* Agnoprotein Is an Essential Egress Factor during BK Polyomavirus Infection. *Int J Mol Sci* **19**, 902 (2018).
41. Caller, L. G. *et al.* Temporal Proteomic Analysis of BK Polyomavirus Infection Reveals Virus-Induced G2 Arrest and Highly Effective Evasion of Innate Immune Sensing. *J Virol* **93**, e00595-19 (2019).
42. An, P., Robles, M. T. S., Duray, A. M., Cantalupo, P. G. & Pipas, J. M. Human polyomavirus BKV infection of endothelial cells results in interferon pathway induction and persistence. *Plos Pathog* **15**, e1007505 (2019).
43. An, P. *et al.* Single-Cell Transcriptomics Reveals a Heterogeneous Cellular Response to BK Virus Infection. *J Virol* **95**, (2021).
44. Justice, J. L., Needham, J. M. & Thompson, S. R. BK Polyomavirus Activates the DNA Damage Response to Prolong S Phase. *J Virol* **93**, (2019).
45. Dellaire, G. *et al.* Promyelocytic leukemia nuclear bodies behave as DNA damage sensors whose response to DNA double-strand breaks is regulated by NBS1 and the kinases ATM, Chk2, and ATR. *J Cell Biology* **175**, 55–66 (2006).
46. Boichuk, S., Hu, L., Hein, J. & Gjoerup, O. V. Multiple DNA Damage Signaling and Repair Pathways Deregulated by Simian Virus 40 Large T Antigen. *J Virol* **84**, 8007–8020 (2010).
47. Boichuk, S., Hu, L., Makielski, K., Pandolfi, P. P. & Gjoerup, O. V. Functional Connection between Rad51 and PML in Homology-Directed Repair. *Plos One* **6**, e25814 (2011).
48. Yeung, P. L. *et al.* Promyelocytic leukemia nuclear bodies support a late step in DNA double-strand break repair by homologous recombination. *J. Cell. Biochem.* **113**, 1787–1799 (2012).
49. Roemer, A., Mohammed, L., Strickfaden, H., Underhill, D. A. & Hendzel, M. J. Mechanisms governing the accessibility of DNA damage proteins to constitutive heterochromatin. *Frontiers Genetics* **13**, 876862 (2022).
50. Strom, A. R. *et al.* Phase separation drives heterochromatin domain formation. *Nature* **547**, 241–245 (2017).
51. Molenberghs, F. *et al.* Cells infected with human papilloma pseudovirus display nuclear reorganization and heterogenous infection kinetics. *Cytom Part A* **101**, 1035–1048 (2022).
52. Mirabelli, C. *et al.* Morphological cell profiling of SARS-CoV-2 infection identifies drug repurposing candidates for COVID-19. *Proc National Acad Sci* **118**, e2105815118 (2021).
53. Dittmar, M. *et al.* Drug repurposing screens reveal cell-type-specific entry pathways and FDA-approved drugs active against SARS-Cov-2. *Cell Reports* **35**, 108959 (2021).

54. Abend, J. R., Low, J. A. & Imperiale, M. J. Inhibitory Effect of Gamma Interferon on BK Virus Gene Expression and Replication. *J Virol* **81**, 272–279 (2007).
55. Zhao, L. & Imperiale, M. J. Establishing Renal Proximal Tubule Epithelial-Derived Cell Lines Expressing Human Telomerase Reverse Transcriptase for Studying BK Polyomavirus. *Microbiol Resour Announc* **8**, e01129-19 (2019).
56. Buck, C. B., Pastrana, D. V., Lowy, D. R. & Schiller, J. T. Efficient Intracellular Assembly of Papillomaviral Vectors. *J Virol* **78**, 751–757 (2004).
57. Broekema, N. M. & Imperiale, M. J. Efficient propagation of archetype BK and JC polyomaviruses. *Virology* **422**, 235–241 (2012).
58. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676–682 (2012).
59. Pachitariu, M. & Stringer, C. Cellpose 2.0: how to train your own model. *Nat Methods* **19**, 1634–1641 (2022).
60. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat Methods* **18**, 100–106 (2021).
61. McQuin, C. *et al.* CellProfiler 3.0: Next-generation image processing for biology. *Plos Biol* **16**, e2005970 (2018).
62. Stirling, D. R., Carpenter, A. E. & Cimini, B. A. CellProfiler Analyst 3.0: Accessible data exploration and machine learning for image analysis. *Bioinformatics* **37**, btab634- (2021).
63. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *Arxiv* (2018) doi:10.48550/arxiv.1802.03426.
64. Harlow, E., Crawford, L. V., Pim, D. C. & Williamson, N. M. Monoclonal antibodies specific for simian virus 40 tumor antigens. *J Virol* **39**, 861–869 (1981).

Chapter 3

Discussion

This dissertation has aimed to start describing BKPyV infection on a cellular level using high-content microscopy. Single-cell analysis of viral infections has multiple advantages over bulk population studies. If not all of the cells in a given population are infected, measuring parameters like gene expression or protein levels for the whole population can mask potentially meaningful information. Because these experiments need to combine many cells per sample, bulk population studies can't account for cellular heterogeneity that may influence viral heterogeneity. Factors like proximity to other cells, the infection status of those other cells, and paracrine and autocrine signaling in response to other cells could all influence a cell's response to viral infection ¹. None of these individual variables could be accounted for in a bulk culture experiment. This dissertation has demonstrated the value of single cell high content microscopy for studying BKPyV infection by identifying multiple distinct subpopulations of TAg⁺ cells that arise during the course of infection. Through imaging of just three nuclear factors, two host and one viral, we identified significant heterogeneity in both the presence and absence of virus, supporting the need for a single-cell approach. Only a subset of the cells in the experiment expressed the highest TAg levels, and this assay allowed us to determine the precise distribution of timepoints represented in those groups with the highest TAg expression. Unlike data in western blots that suggest TAg levels continually increase over the course of infection ², the highest TAg expressing cells were often found at 2 dpi. Additionally, studying PML-NBs on a

single-cell level painted a clearer picture of their behavior during BKPyV infection. The scale of the assay allowed us to validate past findings, namely that BKPyV infection induces reorganization of PML-NBs, and to describe the TAg and DNA characteristics of individual cells with those reorganized NBs. Our findings demonstrate the value of this assay for identifying and beginning to describe unique cell subpopulations during BKPyV infection. The findings to date and the assay provide a starting point for future single-cell microscopy studies of BKPyV.

The experiments and results included in this dissertation have utilized a new methodology and broadened the available data in the field of BKPyV. As we were establishing the single-cell high content microscopy assay, it was important to validate key markers of BKPyV infection. Consistent with known BKPyV biology, there was an increased population of cells with $>2N$ DNA content in the TAg⁺ cells compared to the mock-infected cells. However, while the percentage of TAg⁺ cells increased over time, the TAg level within individual cells did not increase in a linear fashion. This absence of a linear increase in TAg was surprising when compared to bulk population studies, but not when compared to a BKPyV scRNA-seq analysis, which had identified different populations of cells that did not appear to progress to high viral gene expression over time ³. In our assay, we found that the distribution of TAg expression levels within the TAg⁺ cells was similar at 1- and 3-days post infection, with 2-days post infection being the outlier with a greater range. Overall, while there was a notable correlation between DNA and TAg levels, TAg and PML protein levels appeared to have no impact on each other. However, these results recapitulated the prior finding that BKPyV infection leads to fewer PML-NBs per cell.

These previous findings regarding PML-NBs and BKPyV included data obtained from confocal imaging (cite). However, the current single-cell high content microscopy approach

facilitated a more comprehensive analysis of PML-NB morphology than had been previously possible. The benefits of this approach lie in the quantity of data that can be collected and the high resolution of the images acquired. This microscopy assay collected data from 60,000 cells in total. The high resolution of the images facilitated fine measurements including the number, size, and intensity of PML-NBs. Using PML signal intensity as a readout for PML protein expression levels, we could also determine the amount of PML protein present in each nuclear body and nucleus. The prior study on BKPyV and PML-NB relied on counting and measuring PML-NBs manually and used western blotting of bulk populations to assess PML protein levels⁴. In focusing on individual cells, and many more than before, this dissertation provides a more detailed description of PML-NB organization during BKPyV.

Prior to the current study, the work by Jiang, et al. was the most comprehensive study of BKPyV and PML-NBs. Since then, the relationship between BKPyV infection and PML-NBs is an area that has remained under-investigated. PML is remarkably challenging to work with, particularly from a classical genetics approach. The protein is highly post-translationally modified and exists as seven isoforms due to aberrant splicing. Even when an experimental question doesn't include the role of different isoforms, it becomes necessary to include these variants eventually if any knockdown/knockouts are used. Completely knocking out PML gene function by CRISPR, for example, is a straightforward approach to study infection in the absence of PML⁵. When knockouts aren't feasible, for reasons like primary cells, RNAi technology becomes necessary. Attempting to deplete PML on a transcriptional level introduces the potential for differential isoform expression. Even when siRNA or shRNA is designed to target a sequence common to all of the variants, it may not deplete all RNA equally. Researchers can attempt to control for this by using reagents in excess, but the potential for uneven depletion of PML

isoforms can complicate or obscure results. Once an appropriate PML-deficient system has been developed and characterized, any possible PML-dependent results should be confirmed. This confirmation is generally achieved by restoring PML protein to the cells ⁶. Global PML knockdowns with siRNA or shRNA and are often performed in human cell lines, like HEK293 or HeLa, rather than primary cells ⁷. Commonly used expression vectors used to restore or overexpress PML protein isoforms are tagged with FLAG or YFP for easy identification ⁶⁻¹⁰. The choice of promoter for PML isoform expression is important in these experiments and one should keep in mind the growth characteristics of the cell line, as some promoters are stronger in some cell types than others ¹¹. One must also be mindful that the virus they are studying does not affect the promoter or any other component of the expression vector. Some PML isoforms have been expressed using a CMV promoter ⁸. Another system intentionally utilizes the weak gD promoter from HSV-1 to ensure physiologic levels of PML isoforms ¹⁰. PML isoforms can form abnormal structures in the nucleus, especially when expressed in addition to endogenous PML. This phenomenon can make it challenging to interpret if results are due to PML or changes induced by protein overexpression. A visual representation of these abnormal PML-NBs can be found in an article by Li, et al., although the overexpression did not appear to interfere with the results in this case ¹². In one study on the effects of different PML isoforms on HAdV infection in two carcinoma cell lines, researchers observed dramatic differences in exogenous PML expression ¹³. The constructs used in this study contained the gD promoter, which may explain some of the very low expression, but not the variability.

In light of the consequence of the many challenges inherent to working with PML, single-cell high-content microscopy is an ideal medium for studying PML-NBs. This method allows one to examine the nuclear bodies in context of the bigger picture. There are many important questions

to ask about the roles of individual PML isoforms, but this dissertation focused on understanding more about PML-NBs in general. All PML-NBs contain PML protein, and any individual PML isoform is sufficient to create the backbone of a PML-NB ¹⁰. Microscopy is preferred because it maintains the conformation and structure of PML-NBs and allows for the study of spatial relationships between the bodies and other nuclear components.

This dissertation is not the first example of work using single-cell high content microscopy that also studied PML-NBs. A recent paper from Molenberghs, et al. described a single-cell microscopic analysis of infection with HPV pseudovirus (PsV) expressing GFP. HPV is very challenging to grow in the lab, so HPV PsV was generated by packaging a reporter plasmid with the HPV16 capsid proteins L1 and L2 ¹⁴. The results showed great heterogeneity in infection kinetics among individual cells across two cell lines, HeLa cells and a spontaneously transformed human keratinocyte cell line. These differences could be linked to the proliferation rate of the respective host cells. However, kinetics among cells could vary so much that even two daughter cells from a single parental cell could be different. The authors proposed that this was due to uneven distribution of PsV genomes during mitosis, because they had found that viral load was a predictor of infection rate. The study also demonstrated HPV's requirement for cell cycle progression for successful infection. Single cell high throughput microscopy was only employed for part of this study, but that portion of the study allowed for direct comparison of GFP positive and negative cells within the infected condition. The authors discovered differences between these groups that would have been obscured in a bulk population of cells, where it would not be apparent which cells were positive and which weren't. These differences included decreased homogeneity and increased contrast in nuclear images from infected cells. These findings suggested an altered chromatin condensation state in the infected cells ¹⁴. In our own assay we

used a similar strategy, by making our comparisons between mock-infected cells and infected cells that were TAg-positive. We chose mock-infected cells instead of the TAg-negative cells that had received BKPyV inoculum because we wanted to avoid any comparisons to cells that may have undetectable TAg expression, yet still are infected. The data pertaining to PML-NBs is a point of contrast between our work and this HPV PsV study. The authors did not utilize their high-throughput microscopy assay to measure PML-NBs, but instead calculated something they called “PML spot occupancy”. The PML spot occupancy was determined based on the percentage of nuclear area that stained positive for PML. This value was greater in GFP-positive (i.e., infected) cells than GFP-negative cells. The authors followed up on this finding by expressing a fluorescently-tagged PML (unspecified) isoform and performing live cell imaging, with the conclusion that infection increased the number and size of PML-NBs. While this study is a strong example of using single-cell high-throughput microscopy to study small DNA viruses, the PML-NB findings were not as robust. Both the quantity and resolution of images in our work made detailed measurements of individual PML-NBs possible. We have been able to demonstrate the value of these measurements in our interpretation of the subpopulations of cells formed during BKPyV infection, particularly when differentiating between those with the highest TAg expression.

Single cell experiments have been employed in the study of many viruses outside the context of PML-NBs as well, of course. Both RNA and DNA viruses have been investigated on a cellular level, most commonly using sequencing-based approaches^{15–21}. However, other techniques and combinations of techniques are also in use. A single cell study of HAdV infection employed a padlock probe (PLP)-based rolling circling amplification (RCA) method that allowed simultaneous detection of the viral genomic DNA and mRNA expression via

immunofluorescent microscopy²². The analysis revealed both the value of single cell analysis and simultaneous genomic/transcriptional readouts. Results from this study demonstrated that the extent of heterogeneity in the HAdV-infected cell population decreased over time. The authors highlighted that this would not have been identifiable in bulk samples. Prior to the work described in this dissertation only one other single-cell analysis of BKPyV infection had been reported. That study, out of the Pipas lab, provided the first valuable insight into the heterogeneity of BKPyV infection of RPTE cells by analyzing transcriptomics data³. While the work from the Pipas lab included some immunofluorescence microscopy, it was not for the purpose of large-scale analysis, but rather to validate the transcriptional findings.

In our study we used a non-biased approach to analyze features extracted from the high-content imaging on a cellular level. Resulting dimension reductionality analyses, including UMAPs, revealed distinct subpopulations of TAg⁺ cells with unique defining features. These subpopulations (or clusters) represented an array of phenotypes. Out of eight clusters of cells, four of them had quite low TAg levels (Clusters 1 -3 and 6), two had very high TAg levels (Clusters 7 and 8), and two fell somewhere in the middle (Clusters 4 and 5). Within a given level of TAg, multiple parameters differ between these clusters, like DNA content, nuclear area, PML protein expression, number and size of PML-NBs, and distribution of timepoints within the cluster. In this dissertation we focused particular attention on two groups of clusters that had indistinguishable levels of TAg, Clusters 1, 2, and 3 in the low TAg group and Clusters 7 and 8 in the high TAg group. We did not include Cluster 6 in the low TAg group because, while the cluster does have quite low TAg levels, it was unique from Clusters 1 - 3 in multiple ways, including timepoint distribution, as it is comprised of almost exclusively from cells at 3 dpi. Cluster 6 also has the smallest nuclei out of all the clusters, but not the least DNA content,

leading to bright nuclei. Cluster 6 was by far the most homogenous of the eight clusters. The small, bright nuclei of this cluster suggested that they may be in early or intermediate apoptosis

23.

In the low TAg group Clusters 1 – 3 all expressed similarly low levels of TAg, but Cluster 3 had the lowest DNA content of the three, followed by Cluster 1. While the three clusters shared PML protein levels and number of PML-NBs, Cluster 3 had more PML protein per nuclear body. In regard to nuclear size, all three low TAg-expressing clusters were similar, with Cluster 1 being the smallest out of the group. These three clusters all have similar ratios of cells at the different timepoints (see Chapter 2 Supplemental Table 2.7).

When the global data were subset to only the TAg⁺ cells and re-embedded for UMAP analysis, the low and medium TAg-expressing cells remained clustered in roughly the same populations as they had when mock-infected and TAg-negative cells were also a part of the analysis. This lack of change indicates that TAg features are not likely to be major contributors to the defining characteristics of the clusters. Table 2.2 describes the major defining features of the TAg⁺ clusters and supports this conclusion. The major defining feature of the clusters in the low TAg group is low DNA signal intensity, which represents low DNA content. The other defining features listed for these cluster differentiate them not just from each other, but from the entire TAg⁺ population. Comparing Clusters 1 – 3 only with the others in the low TAg is helpful in this circumstance. The low TAg group of clusters also shares low DNA content, but the different amounts of DNA are what likely make them three separate clusters. Cluster 1 has the least DNA, followed by Cluster 2, then Cluster 3. There are very subtle differences in nuclear area and PML-NB size and intensity, but the clearest difference is in DNA content. Perhaps the cells in each of these three clusters represent different points of cell progression. This difference

could be both a result of and cause of low TAg levels, or they could be due to abortive infection. The three different levels of DNA content in these three clusters could each represent different mechanisms by which the cell isn't transitioning into or from S phase. This would be relatively straightforward to test by incorporating cell cycle markers, like phospho-Cdk1, as well as EdU labeling, to monitor *de novo* DNA synthesis, into our assay. This would facilitate comparisons of cells at the same place in the cell cycle across different conditions, like the studies done in An, et al ³. One could also control the cell cycles pharmacologically, using targeted inhibitors to synchronize cells in a single phase ²⁴. Controlling for the cell cycle allowed the authors in the aforementioned paper to remove potential confounding variables from their assay.

The high TAg group, Clusters 7 and 8, which had emerged as subsets of global Cluster C, clearly had equivalently high levels of TAg expression. TAg level, however, was the only commonality shared between these clusters. Cells from Cluster 7 had a greater nuclear area, in correlation, at least to some extent, with having more DNA content than cells in Cluster 8. The differences in PML were more pronounced. Nuclear PML protein expression was greater for Cluster 8 than Cluster 7, but PML protein expression per PML-NB was greater in Cluster 7 than in 8. This finding is consistent with Cluster 7 containing cells with fewer and larger PML-NBs than those in Cluster 8. Cluster 7's PML-NBs are likely larger because they contain more PML protein, so their individual signal intensity was greater. However, because the NBs are still few in number in Cluster 7, the overall PML protein level isn't as great as in Cluster 8. The final difference between the two high TAg-expressing clusters is the distribution of timepoints within them. Cluster 8 has double the number of cells at 1 dpi than were found in Cluster 7. Cluster 7's timepoint distribution is 8%, 41%, and 51% of cells from days 1, 2, and 3 post-infection, respectively. This is in contrast to Cluster 8, with 20%, 57%, and 24% at the same timepoints.

It is unclear if cells could or do transition from one cluster to another with time. Results from the Pipas lab suggest that the cells are unlikely to change viral gene expression much ³. Or are the clusters more like branches from the same point? I propose that Clusters 7 and 8 are diverging from a central locus and their differences are reflected in their PML-NB phenotype. It is unclear whether or not the PML-NB changes are a cause or an effect of what makes these clusters different, but the NBs are a key component of our study that differentiates them. In this regard, cause and effect can only be assigned to concepts with known mechanisms, including high TAg levels indicating BKPyV infection and being the cause of increased DNA content, as in Clusters 7 and 8. Causality cannot be assigned to PML-NBs and any part of BKPyV infection at this point in time because the role of PML protein during the infection is still undefined.

In order to better understand the influences shaping cluster formation, particularly from the host side, we will need to expand the parameters describing the clusters. I would do this in a transcriptomics-linked approach. I would seek to collaborate with a lab that is experienced in scRNA-seq, ideally the Pipas lab, to do a tandem study in which both labs perform parallel experiments with matching conditions up until microscopy and sequencing and then link our results. For example, using the same cell lines and viral stocks both labs would perform infections at an MOI of 0.1 infectious units per cell and fix and stain cells or collect RNA every 24 hours for three days. After the RNAseq and microscopy data were analyzed, we would look for parallels between gene expression and signal intensity resulting from staining of that gene's product. We can use this information, combined with expertise in machine learning and data analysis from collaborators here at the University of Michigan, to draw parallels between clusters formed by the RNAseq data and the image analysis. Combining these two single-cell assays into

a single tool would enhance our ability to interpret the morphologic data from the images and the put the transcriptomic data in the context of a visual phenotype.

Focusing on the clusters in the high TAg group revealed many differences. Some were more subtle, like the PML protein expression per NB and nuclear area, but others were quite dramatic, like the number of PML-NBs per nucleus. Cluster 7 was largely defined by the same reorganized PML-NB phenotype described by Jiang, et al ⁴. Cluster 7's PML-NBs were fewer in number and larger in size than mock-infected cells and most of the other TAg⁺ cells. Cluster 8 displayed more and smaller PML-NBs than 7. Cluster 8 had the largest population of cells from 2 dpi of any TAg⁺ cluster, and had twice as many cells from 1 dpi compared to Cluster 7.

Due to the size of our dataset, resulting from both the number of individual cells imaged and measurements collected from each image, we employed techniques to make it more digestible. The general aim of these techniques was to take a large amount of information and extract the most valuable components necessary to understand the data. The primary method that we utilized was UMAP, which was very helpful in grouping the cells into clusters based on the many measured features. While the UMAP approach provides many benefits, it has potential pitfalls as well. One potential concern when using UMAP is due to an assumption made by the algorithm that there is manifold structure within the data ²⁵. Using a sufficiently large sample size to reduce noise in the dataset helps alleviate this concern. While we took care to use a large dataset, it is important to keep this characteristic of UMAP in mind when considering how much weight to assign to the results of the algorithm. Our plans to follow up with tandem single-cell image and transcriptional analysis, as described earlier in this chapter, would be an additional opportunity to validate these results.

As discussed earlier in this chapter, BKPyV infection alters PML-NBs and it does so to a greater extent in some cells than others. Prior work has demonstrated that BKPyV infection alters PML-NB structure and function ⁴. The aforementioned study and the results from this dissertation indicate that this alteration of PML-NBs is not directly induced by TAg alone, nor are TAg and PML protein levels correlated, respectively. Rather, it is possible that BKPyV infection could be inducing changes in PML-NBs indirectly via the DNA damage response (DDR). Ample evidence exists supporting the importance of the DDR in BKPyV infection ^{26–29}. There is also substantial amount of work in the field of PML-NBs defining them as components of the DDR ³⁰. While it is not clear what exact component of the DDR may be causing the reorganization of PML-NBs and how, a few possibilities come to mind. PML contains multiple SUMOylation sites and a SUMO interacting motif (SIM). The protein utilizes SUMO-SIM interactions in the formation of nuclear bodies. During activation of the DDR, ATM/ATR-dependent phosphorylation of the E3 Ub ligase, SIAH stabilizes and activates HIPK2 ³¹. HIPK2 phosphorylates PML at serines 8 and 38, which triggers SUMOylation of PML ³². One option for the indirect influence of BKPyV infection on PML-NBs via the DDR is that induction of ATM and ATR pathways by BKPyV is triggering phosphorylation, and in turn, SUMOylation of PML. SUMOylation, or even hyper-SUMOylation of PML may alter the aggregation of the nuclear bodies or perhaps alter which proteins are associated with them. Another option is that PML-NBs are involved in ss and dsDNA repair mechanisms being utilized during viral genome replication. PML-NBs are required for fully efficient homologous recombination repair of cellular DNA ^{33–35}. That may cause reorganization of the nuclear bodies. Yet another option is related to PML-NBs' major role in inducing apoptosis either in a p53-dependent or -independent manner ³⁶. An unknown event could be occurring at the interface of TAg's regulation of p53 and

PML's attempts to induce apoptosis that causes a serious disruption in the system. DNA damage-induced polyubiquitination of p53 is blocked by PML via sequestration of the E3 Ub ligase, MDM2, to nucleoli³⁷. This prevents degradation of p53, which can also associate with PML-NBs in response to DNA damage. MDM2 levels are decreased and p53 levels are increased during BKPyV infection³⁸. If BKPyV infection triggers DNA damage-induced MDM2 sequestration via PML this may be complicated by TAG's sequestration of p53, especially if said p53 is residing in PML-NBs at the time. To my knowledge, no one has examined PML-NBs during ATM or ATR inhibition. Doing so would be an important first step in the process of determining a potential connection between the kinases and PML. I would test my hypothesis that BKPyV infection is indirectly altering PML-NBs via the DDR by inhibiting ATM, ATR, or both kinases together during BKPyV infection. Because viral replication is required to induce the DDR. I will include both a WT virus and a replication deficient mutant²⁸ as a control. If my hypothesis is correct, I anticipate that infection with WT BKPyV in the absence of ATM and/or ATR inhibition will be the only condition with reorganized PML-NBs because it is the only condition with both arms of the DDR intact and a replication-sufficient BKPyV to induce the DDR.

Although a meaningful body of work exists regarding BKPyV and the DDR, until recently there have not been any published high-resolution images of DNA staining in the nuclei of BKPyV-infected cells. Of the few studies that have published high resolution immunofluorescence images of BKPyV-infected cells, those that exist have had low resolution DNA staining or have not published the DNA channel individually. It has been challenging overall to find quality images in the literature of the nuclei of BKPyV-infected cells. The lack of published high-resolution microscopy images of BKPyV-infected cells has been a genuine gap in

the field, especially images that capture the variety of intensities and textures of DNA staining in the nucleus. High-resolution DNA images provide information about its distribution, density, location, and other properties in the nucleus. This information is valuable when studying BKPyV infection, as the virus replicates in the nucleus and utilizes host DNA replication machinery for its own genome replication. During PyV replication, DNA from both the host cell and the virus increase in quantity. High resolution images provide a means for gathering information about how that increased DNA content is arranged in the nucleus. Prior studies have looked for co-localization between areas of DNA intensity and specific nuclear markers, but these confocal images facilitate analysis of DNA alone or in the context of other nuclear markers/viral protein(s). We used the high-resolution images of single cells that we had obtained to detect a variety of different patterns of DNA staining. We examined the infection on a cellular level and used an automated analysis pipeline to find populations of cells with interesting DNA patterns that may not have been obvious using manual examination due to their frequency in the total population of cell. We identified cells that displayed a ‘TAg and DNA exclusion’ phenotype, wherein there are very distinct regions of intense DNA and TAg signals that occupy non-overlapping regions of the nucleus. Heterochromatin has been proposed to be organized in phase-separated domains ³⁹. These domains restrict access to the chromatin, but in more complex ways than were previously understood. Rather than just being size-excluded, as was expected, data indicate that a molecules’ ability to interact with the heterochromatin in these phase-separated domains is dependent on the molecules’ solubility ⁴⁰. The heterochromatin in these domains is still highly methylated at histone H3K9, a marker of heterochromatin, and is bound to heterochromatin protein 1 (HP1). I propose that the cells with the exclusion phenotype contain examples of these phase separated domains, and that agnoprotein may be partly responsible for

their development. JCPyV's agnoprotein has been shown to bind HP1alpha⁴¹. Given the homology shared by the viruses, BKPyV's likely does as well. This phenotype was only detected in cells with high TAg and DNA content at 2 days or later, indicating that the late proteins should be present by this point. Agnoprotein may be binding to HP1 and contributing to this unusual phenotype. While agnoprotein is normal cytoplasmic, it is occasionally found in the nucleus, and this may be one of those occasions. There are a few places to start to answer some of these questions. One would be a pull-down assay to see if BKPyV's agnoprotein really does bind to HP1alpha. Another would be to see if the intensely stained areas of DNA contained markers of heterochromatin. A straightforward first experiment would be to repeat my single-cell microscopy BKPyV infection assay, but to focus on heterochromatin by using H3K9me3 and HP1 antibodies. I would still stain for TAg and DNA as before and look for prior phenotypes, but this time in the context of chromatin markers. I anticipate that the areas of the nucleus that display the 'TAg and DNA exclusion' phenotype will be sites of co-localized H3K9me3 and HP1 signal, indicating that they are indeed heterochromatin. I would also stain the cells with an antibody against BKPyV agnoprotein to determine agnoprotein's location in the cell as it relates to TAg and DNA.

BKPyV infection remains a cause of morbidity and mortality for patients following transplants, and we are still without effective antivirals or vaccines. Substantial work has been done over the last fifty years to bring the field of BKPyV research to where it is today. Nearly every basic science study of BKPyV has relied on bulk populations of cells. While these studies have been essential, it is now important to gain insight into BKPyV infection on a single-cell level. The first BKPyV single-cell study was performed by An, et al. focusing on transcriptomics on cells during viral infection³. In this dissertation I have described the results of the second

single-cell study of BKPyV infection. This includes results that are consistent with the prior single-cell study's findings of significant infection heterogeneity, including highly variable TAg expression levels. Additionally, I found that TAg expression was the defining characteristic for clusters of cells only when it was at high levels and that this population of cells with high TAg levels could be sub-set into two clusters. These two clusters were distinguished by their PML-NB characteristics, reinforcing the absence of a direct TAg-PML interaction.

In summary, this dissertation has presented a better understanding of BKPyV infection and a method to help expand our knowledge. Bulk population studies in tissue culture have been the cornerstone of work in this field until recently and have been necessary tools for understanding PyV biology. Using single-cell technology, we have narrowed the focus of our experimental questions from 'what does BKPyV do in cells?' to 'what does BKPyV do in *THIS* cell?'. Using single cell high content microscopy, we asked that question about tens of thousands of cells and learned that the answers were varied. Our analysis of DNA, TAg, and PML-NBs during a 3-day infection revealed a more heterogeneous landscape of cells than previous bulk studies would have predicted. This study will be a starting point for future discovery among BKPyV researchers.

3.1 Bibliography

1. Ramos, I. *et al.* Innate Immune Response to Influenza Virus at Single-Cell Resolution in Human Epithelial Cells Revealed Paracrine Induction of Interferon Lambda 1. *J Virol* **93**, e00559-19 (2019).
2. Low, J., Humes, H. D., Szczypka, M. & Imperiale, M. BKV and SV40 infection of human kidney tubular epithelial cells in vitro. *Virology* **323**, 182–188 (2004).
3. An, P. *et al.* Single-Cell Transcriptomics Reveals a Heterogeneous Cellular Response to BK Virus Infection. *J Virol* **95**, (2021).
4. Jiang, M., Entezami, P., Gamez, M., Stamminger, T. & Imperiale, M. J. Functional Reorganization of Promyelocytic Leukemia Nuclear Bodies during BK Virus Infection. *Mbio* **2**, e00281-10 (2011).
5. Masroori, N. *et al.* Gene Knockout Shows That PML (TRIM19) Does Not Restrict the Early Stages of HIV-1 Infection in Human Cell Lines. *Msphere* **2**, e00233-17 (2017).
6. Guccione, E., Lethbridge, K. J., Killick, N., Leppard, K. N. & Banks, L. HPV E6 proteins interact with specific PML isoforms and allow distinctions to be made between different POD structures. *Oncogene* **23**, 4662–4672 (2004).
7. Chen, Y., Wright, J., Meng, X. & Leppard, K. N. Promyelocytic Leukemia Protein Isoform II Promotes Transcription Factor Recruitment to Activate Interferon Beta and Interferon-Responsive Gene Expression. *Mol Cell Biol* **35**, 1660–1672 (2015).
8. Beech, S. J., Lethbridge, K. J., Killick, N., McGlinchey, N. & Leppard, K. N. Isoforms of the promyelocytic leukemia protein differ in their effects on ND10 organization. *Exp Cell Res* **307**, 109–117 (2005).
9. Wimmer, P. *et al.* SUMO modification of E1B-55K oncoprotein regulates isoform-specific binding to the tumour suppressor protein PML. *Oncogene* **29**, 5511–5522 (2010).
10. Cuchet, D. *et al.* PML isoforms I and II participate in PML-dependent restriction of HSV-1 replication. *J Cell Sci* **124**, 280–291 (2010).
11. Smith, R. L. *et al.* Characterization of Promoter Function and Cell-Type-Specific Expression from Viral Vectors in the Nervous System. *J Virol* **74**, 11254–11261 (2000).
12. Li, Z. *et al.* 3C protease of enterovirus 71 cleaves promyelocytic leukemia protein and impairs PML-NBs production. *Virol J* **18**, 255 (2021).
13. Mai, J. *et al.* PML Alternative Splice Products Differentially Regulate HAdV Productive Infection. *Microbiol Spectr* **10**, e00785-22 (2022).
14. Molenberghs, F. *et al.* Cells infected with human papilloma pseudovirus display nuclear reorganization and heterogenous infection kinetics. *Cytom Part A* **101**, 1035–1048 (2022).

15. Liu, W., He, H. & Zheng, S.-Y. Microfluidics in Single-Cell Virology: Technologies and Applications. *Trends Biotechnol* **38**, 1360–1372 (2020).
16. Suomalainen, M. & Greber, U. F. Virus Infection Variability by Single-Cell Profiling. *Viruses* **13**, 1568 (2021).
17. Guo, F. *et al.* Single-Cell Virology: On-Chip Investigation of Viral Infection Dynamics. *Cell Reports* **21**, 1692–1704 (2017).
18. Rato, S., Golumbeanu, M., Telenti, A. & Ciuffi, A. Exploring viral infection using single-cell sequencing. *Virus Res* **239**, 55–68 (2017).
19. Brandt, L., Cristinelli, S. & Ciuffi, A. Single-Cell Analysis Reveals Heterogeneity of Virus Infection, Pathogenicity, and Host Responses: HIV as a Pioneering Example. *Ann Rev Virol* **7**, 333–350 (2020).
20. Ciuffi, A., Rato, S. & Telenti, A. Single-Cell Genomics for Virology. *Viruses* **8**, 123 (2016).
21. Combe, M., Garijo, R., Geller, R., Cuevas, J. M. & Sanjuán, R. Single-Cell Analysis of RNA Virus Infection Identifies Multiple Genetically Diverse Viral Genomes within Single Infectious Units. *Cell Host Microbe* **18**, 424–432 (2015).
22. Krzywkowski, T., Ciftci, S., Assadian, F., Nilsson, M. & Punga, T. Simultaneous Single-Cell In Situ Analysis of Human Adenovirus Type 5 DNA and mRNA Expression Patterns in Lytic and Persistent Infection. *J Virol* **91**, (2017).
23. Filippi-Chiela, E. C. *et al.* Nuclear Morphometric Analysis (NMA): Screening of Senescence, Apoptosis and Nuclear Irregularities. *PLoS ONE* **7**, e42522 (2012).
24. Seidman, M. A. *et al.* Variation in Adenovirus Receptor Expression and Adenovirus Vector-Mediated Transgene Expression at Defined Stages of the Cell Cycle. *Mol Ther* **4**, 13–21 (2001).
25. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *Arxiv* (2018) doi:10.48550/arxiv.1802.03426.
26. Justice, J. L., Needham, J. M., Verhalen, B., Jiang, M. & Thompson, S. R. BK Polyomavirus Requires the Mismatch Repair Pathway for DNA Damage Response Activation. *J Virol* **96**, e02028-21 (2022).
27. Jiang, M., Zhao, L., Gamez, M. & Imperiale, M. J. Roles of ATM and ATR-Mediated DNA Damage Responses during Lytic BK Polyomavirus Infection. *Plos Pathog* **8**, e1002898 (2012).
28. Verhalen, B., Justice, J. L., Imperiale, M. J. & Jiang, M. Viral DNA Replication-Dependent DNA Damage Response Activation during BK Polyomavirus Infection. *J Virol* **89**, 5032–5039 (2015).
29. Justice, J. L., Needham, J. M. & Thompson, S. R. BK Polyomavirus Activates the DNA Damage Response to Prolong S Phase. *J Virol* **93**, (2019).
30. Dellaire, G. & Bazett-Jones, D. P. PML nuclear bodies: dynamic sensors of DNA damage and cellular stress. *Bioessays* **26**, 963–977 (2004).

31. Winter, M. *et al.* Control of HIPK2 stability by ubiquitin ligase Siah-1 and checkpoint kinases ATM and ATR. *Nat Cell Biol* **10**, 812–824 (2008).
32. Gresko, E. *et al.* PML tumor suppressor is regulated by HIPK2-mediated phosphorylation in response to DNA damage. *Oncogene* **28**, 698–708 (2009).
33. Yeung, P. L. *et al.* Promyelocytic leukemia nuclear bodies support a late step in DNA double-strand break repair by homologous recombination. *J. Cell. Biochem.* **113**, 1787–1799 (2012).
34. Attwood, K. M. *et al.* PML isoform expression and DNA break location relative to PML nuclear bodies impacts the efficiency of homologous recombination. *Biochem Cell Biol* **98**, 314–326 (2020).
35. Boichuk, S., Hu, L., Makielski, K., Pandolfi, P. P. & Gjoerup, O. V. Functional Connection between Rad51 and PML in Homology-Directed Repair. *Plos One* **6**, e25814 (2011).
36. Yang, S., Kuo, C., Bisi, J. E. & Kim, M. K. PML-dependent apoptosis after DNA damage is regulated by the checkpoint kinase hCds1/Chk2. *Nat Cell Biol* **4**, 865–870 (2002).
37. Bernardi, R. *et al.* PML regulates p53 stability by sequestering Mdm2 to the nucleolus. *Nat Cell Biol* **6**, 665–672 (2004).
38. Caller, L. G. *et al.* Temporal Proteomic Analysis of BK Polyomavirus Infection Reveals Virus-Induced G2 Arrest and Highly Effective Evasion of Innate Immune Sensing. *J Virol* **93**, e00595-19 (2019).
39. Roemer, A., Mohammed, L., Strickfaden, H., Underhill, D. A. & Hendzel, M. J. Mechanisms governing the accessibility of DNA damage proteins to constitutive heterochromatin. *Frontiers Genetics* **13**, 876862 (2022).
40. Strom, A. R. *et al.* Phase separation drives heterochromatin domain formation. *Nature* **547**, 241–245 (2017).
41. Okada, Y. *et al.* Dissociation of heterochromatin protein 1 from lamin B receptor induced by human polyomavirus agnoprotein: role in nuclear egress of viral particles. *Embo Rep* **6**, 452–457 (2005).