

# **Control of archetype BK polyomavirus miRNA expression**

Wei Zou<sup>1</sup>, Gau Shoua Vue<sup>1</sup>, Benedetta Assetta<sup>2</sup>, Heather Manza<sup>1</sup>, Walter J. Atwood<sup>2</sup>, and Michael J. Imperiale<sup>1\*</sup>

<sup>1</sup>Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI 48109

<sup>2</sup>Department of Molecular Biology, Cell Biology, and Biochemistry, Brown University, Providence, RI 02912

\*Corresponding author

# Abstract

BK polyomavirus (BKPv) is a ubiquitous human pathogen, with over 80% of adults worldwide persistently infected. BKPv infection is usually asymptomatic in healthy people; however, it causes polyomavirus-associated nephropathy in renal transplant patients and hemorrhagic cystitis in bone marrow transplant patients. BKPv has a circular, double-stranded DNA genome that is divided genetically into three parts: an early region, a late region, and a non-coding control region (NCCR). The NCCR contains the viral DNA replication origin and cis-acting elements regulating viral early and late gene expression. It was previously shown that a BKPv miRNA expressed from the late strand regulates viral large T antigen expression and limits the replication capacity of archetype BKPv. A major unanswered question in the field is how expression of the viral miRNA is regulated. Typically, miRNA is expressed from introns in cellular genes but there is no intron readily apparent in the BKPv from which the miRNA could derive. Here we provide evidence for primary RNA transcripts that circle the genome more than once and include the NCCR. We identified splice junctions resulting from splicing of primary transcripts circling the genome more than once, and Sanger sequencing of RT-PCR products indicates that there are viral transcripts that circle the genome up to four times. Our data suggest that the miRNA is expressed from the intron of these greater-than-genome size primary transcripts.

## 34 Introduction

35 BK polyomavirus (BKPyV), a member of the polyomavirus family, is a ubiquitous human pathogen with  
 36 over 80% of adults worldwide persistently infected (1, 2). Infection with BKPyV is generally thought to  
 37 be asymptomatic in healthy individuals despite periodic episodes of reactivation and viral shedding in  
 38 the urine. In renal transplant patients, reactivation of BKPyV in the donor-derived kidney causes  
 39 polyomavirus-associated nephropathy (PVAN), and in the bladder of bone marrow transplant patients  
 40 causes hemorrhagic cystitis (3-5). There is increasing evidence that BKPyV plays a role in the  
 41 development of urothelial and bladder cancer in transplant recipients who have had BKPyV viremia or  
 42 PVAN (6-11). There are currently no FDA-approved drugs to treat BKPyV-associated diseases. The  
 43 primary treatment for interstitial cystitis is palliative, and for PVAN is lowering the dosage of  
 44 immunosuppressants to allow the immune system to clear the virus, which inevitably increases the risk  
 45 of acute rejection and graft failure (12).

46 BKPyV has a double-stranded circular DNA genome which is divided genetically into the early region,  
 47 late region, and the noncoding control region (NCCR). The early region encodes the large T and small  
 48 T antigens (TAg and tAg), which are important for BKPyV replication, and a truncated TAg that induces  
 49 expression of cellular APOBEC3B (13, 14). The late region is expressed from the complementary  
 50 strand of the genome and encodes the viral capsid proteins VP1, VP2, and VP3, and an auxiliary  
 51 protein called agnoprotein. Agnoprotein has been reported to facilitate BKPyV replication by disrupting  
 52 the mitochondrial network and mitochondrial membrane potential, allowing evasion of innate immunity  
 53 (15), as well as playing a role as a viroporin in viral egress (16, 17). The NCCR contains the cis-acting  
 54 elements that are required for regulating DNA replication, including the origin of replication, and  
 55 transcription of the early and late regions. Based on the NCCR structure, there are two forms of the  
 56 BKPyV genome, archetype virus and rearranged variants. The NCCR structure of archetype virus is  
 57 arbitrarily divided into five sequence blocks termed O, P, Q, R, and S, while the NCCRs of rearranged  
 58 variants are generally characterized by duplications and/or deletions of the P, Q, R, and S blocks (18,  
 59 19). The O block is essential and cannot be deleted because it contains the origin of DNA replication.

60 The archetype virus is thought to be the transmissible form of the virus because it is found in both  
61 healthy people and in diseased patients, while rearranged variants are usually isolated from patients  
62 with BKPyV disease (20, 21).

63 BKPyV also encodes a miRNA on the late gene strand that targets early mRNAs for degradation, as is  
64 the case for other polyomavirus miRNAs (22-26). Our previous study showed that the BKPyV miRNA  
65 plays an important role in limiting archetype virus replication but not rearranged variant replication: a  
66 miRNA mutant archetype virus showed a much higher replication capability than the wild type virus. We  
67 proposed a model in which the miRNA's function is mediated through the balance of regulatory  
68 elements in the NCCR that control relative expression of the miRNA and the early genes (23).

69 Although it is clear that the BKPyV miRNA plays an important role in controlling early gene expression,  
70 how the expression of the miRNA is regulated is not known. Previous studies have shown that most  
71 cellular miRNAs are expressed from introns of protein coding genes. However, the BKPyV miRNA is  
72 not encoded in an obvious intron. In the present study, we provide evidence that the miRNA is  
73 expressed from a putative intron that is generated by transcription that circles the viral genome more  
74 than once.

75

## 76 **Materials and Methods**

77 **Cell culture.** Primary human Renal Proximal Tubule Epithelial (RPTE) cells (Lonza) were maintained  
78 and passaged in REGM medium (REGM/REBM, Lonza) as described previously. RPTE cells obtained  
79 from ATCC (American Type Culture Collection) were cultured in Renal Epithelial Cell Basal Medium  
80 (RECBM) supplemented with Renal Epithelial Cell Growth Kit. 293 cells and 293TT cells (14) were  
81 cultured in Dulbecco's modified Eagle medium (DMEM) with 10% FBS and 100 units/mL penicillin, 100  
82 µg/mL streptomycin. All cells were cultured in a humidified incubator at 37°C with 5% CO<sub>2</sub>.

83 **Virus infection.** Archetype and rearranged BKPyV (Dik and Dunlop, respectively) were purified and  
84 titrated as described previously (27). RPTE cells at passage 5 were plated in 12-well plates for  
85 infection. Cells were prechilled for 15 min at 4°C and infected with Dik or Dunlop at a MOI of 2  
86 fluorescence-forming unit (FFU)/cell. A 400 µl volume of the diluted virus was added to each well and  
87 incubated at 4°C for 1 h with gentle shaking every 15 min. The virus was aspirated off and the cells  
88 were incubated with fresh REGM medium.

89 **Transfection.** For the experiments comparing circular and linearized plasmids, pGEM-Dik-LT/O, that  
90 contains the NCCR, late coding region, and miRNA coding region, and the GFP expression vector  
91 pEGFP-N1 were linearized with DraIII and treated with alkaline phosphatase before purifying with a  
92 PCR purification kit (Qiagen). Then 500 ng each of circular or linearized pGEM-Dik-LT/O and pEGFP-  
93 N1 were co-transfected into 293 cells with Lipofectamine 2000 reagent according to the manufacturer's  
94 instruction and the medium was replaced with fresh DMEM four hours after transfection. Two days  
95 post-transfection, total RNA was extracted using the Direct-zol RNA MiniPrep kit (ZYMO Research,  
96 USA). GFP expression was visualized using an Olympus microscope, and the number of GFP-positive  
97 cells were determined using ImageJ (NIH).

98 **RT-PCR amplification of the NCCR.** RPTE cells were infected with Dik and Dunlop as described  
99 above. Total RNA was extracted using the Direct-zol RNA MiniPrep kit (ZYMO Research, USA) and  
100 treated with DNaseI (Promega). Reverse transcription (RT) was performed using an NCCR late strand-

specific primer NCCR-r (5'-GGTACCGGCCTTTGTCCAGTTTAACT-3') (Figure 1A), followed by PCR to amplify the NCCR using primers NCCR-f (5'-CTCGAGTTGCAAAAATTGCAAAAGAATAGG-3') and NCCR-r. pGEM-Dunlop plasmid, which contains the entire Dunlop genome, and water were used as positive and negative controls, respectively. The PCR products were electrophoresed in a 2% agarose gel and visualized using a Bio-Rad gel system (Bio-Rad).

**Quantitative stem-loop qPCR.** miRNA stem-loop qPCR was performed and the miRNA expression level was normalized to human miRNA hsa-let-7a-1 using the  $2^{-\Delta\Delta C(t)}$  method as previously described (23, 28).

**Next generation sequencing (RNA-seq) analysis.** RPTE cells were infected with BKPvV Dunlop at an MOI of 0.03 FFU/cell as follows. Virus was diluted in culture media and cells were infected for 2 hours. The inoculum was aspirated and cells were incubated with fresh medium. Cells were fed every 2 days and medium replaced at 6 days. Total RNA was harvested from infected and mock-infected cells at 3 days post infection (dpi), 6 dpi, and 9 dpi with a RNeasy Mini Kit (Qiagen) with DNase treatment. Quantification was performed using a Nanodrop 2000c and the quality of the RNA was assessed with an Agilent 2100 Bioanalyzer. Libraries were prepared with a TruSeq Stranded Total RNA Sample with Ribo-Zero Prep Kit and the RNA sequencing read type was 2x50bp. Library preparation and RNA sequencing were performed by Beckman Coulter Genomics (Genewiz). The number of Quality Control (QC)-passed reads varied between ~60 and 89 million across all samples (29). Hisat2 2.1.0 was used to map RNA-seq reads against the Dunlop full-length genome using default parameters. The BKPvV reads were extracted using an in-house-developed script. The alignment data were used to compute coverage by igvtools count (<https://software.broadinstitute.org/software/igv/igvtools>). Transcript abundance estimation of NCCR, early genes, and late genes was conducted with StringTie 1.3.5 software and is reported in Fragments Per Kilobase of exon per Million fragments mapped (FPKM).

**TA cloning and Sanger sequencing.** Specific primer pairs for confirming the most abundant splice junctions identified from the RNA-seq data were used to amplify cDNA transcribed from RNA isolated

126 from Dik-infected RPTE cells using SuperScript<sup>TM</sup> III Reverse Transcriptase kit (Invitrogen). The  
 127 sequences of the primers are primer 1 (P1) 5'- GACTCTGTAAAAGACTCCTAG-3' and primer 2 (P2) 5'-  
 128 CTAAAATAAAAATAAAAATCCTCTG-3' (Figure 3A). The amplified band was purified using a Qiagen  
 129 gel purification kit and TA cloning was performed using a TOPO<sup>TM</sup> TA cloning<sup>TM</sup> kit (Invitrogen). 12  
 130 clones were randomly picked for Sanger sequence analysis by the Advanced Genomics Core at the  
 131 University of Michigan.

132

## Results

**A circular template expresses higher levels of the miRNA than a linear template.** It is known that miRNAs are usually expressed from introns of protein-coding genes (30). The BKPv miRNA is not encoded in an obvious intron since it maps downstream of the late polyadenylation site (Figure 1A). However, it is possible that the miRNA could be expressed from an intron of a primary RNA transcript that results from transcription initiating in the NCCR and circling the genome more than once (Figure 1A). Such greater-than-genome size primary transcripts are known to be produced from the related murine polyomavirus (MuPyV) (31, 32). If this is the case, the prediction is that linearizing the genome would eliminate such transcripts and reduce miRNA expression. To test this hypothesis, we transfected 293 cells with pGEM-Dik-LT/O, which is a genomic clone of the BKPv NCCR, late region, and sequences downstream of the late poly(A) site that encode the miRNA; or the same plasmid linearized by DraIII digestion (whose recognition site is located in the vector backbone and does not cut within the viral sequences), to mimic circular and linear viral genomes, respectively. A circular or linearized GFP-expressing plasmid was co-transfected with the circular and linear viral plasmids, respectively, to monitor transfection efficiency. The results showed that miRNA expression is significantly lower from the linear viral plasmid than from the circular plasmid (Figure 1B).

**The NCCR region is transcribed.** To demonstrate that RNA polymerase circles the viral genome, we infected RPTE cells with archetype BKPv and extracted RNA 3 days post infection. Reverse transcription was performed using a primer that is complementary to the late strand of the NCCR (Figure 1A), followed by PCR to amplify the NCCR. We detected a specific band in archetype-infected RPTE cells but not mock-infected cells or when the RT step was omitted, indicating that late strand viral RNA transcripts containing the NCCR are produced (Figure 2A). Viral RNA containing the NCCR was also observed in rearranged BKPv-infected RPTE cells (Figure 2B). To confirm that the NCCR was transcribed, we analyzed RNA-seq transcriptome data from rearranged virus-infected RPTE cells. The number of reads mapping to the viral genome increased over time, from approximately 5% (3.5 million reads) at 3 dpi to almost 30% (26 million) of the total reads at 9 dpi. A significant number of reads



spanning the NCCR were detected (Figure 2C, Table 1). The overall NCCR late strand steady state expression level is approximately 10% of that of the late viral coding sequences (Table 2).

**Identification of viral transcripts resulting from RNA polymerase transcribing the genome more than once.** The detection of viral transcripts containing the NCCR by both RT-PCR and RNA-seq supported the possibility of production of a primary RNA transcript resulting from transcription circling the genome more than once (Figure 1A). Such a primary transcript could be spliced in a manner similar to what has been reported for MuPyV, in which a genome-sized intron is removed. To test this, we further analyzed the RNA-seq data of Dunlop-infected RPTE cells for the presence of splice junctions that could only be produced by splicing as indicated in Figure 1A. To facilitate this analysis, we aligned the reads to a template consisting of two tandem copies of the viral genome. We identified many reads containing splice junctions that support the existence of primary transcripts which circle the genome more than once (Table 3). To produce such junctions, an intron containing the miRNA coding sequence would have had to have been excised. The most abundant splice junctions are ones in which the donor site is in the first copy of the genome while the acceptor site is located in the second copy of the genome, which only could occur if the primary transcripts circled the genome more than once and splicing occurred from a donor site such as D1 to an acceptor site such as A2 (Figure 3A, Table 3). Overall, the splice junctions arising from transcripts circling the genome more than once occur at different sites on the viral genome and are enriched within the agnoprotein and VP1 genes. We also observed an increase in the number of transcripts containing these splice junctions over time, although their relative abundance compared to the authentic VP1 splice junction (D3041/A3974) remained constant (Table 3).

To confirm the presence of the most abundant splice junction, we designed a PCR primer pair that would specifically amplify RNAs that are spliced from one copy of the tandem primary transcript to the second copy (Figure 3A). We used this pair of primers to amplify cDNA transcribed from RNA of archetype-infected RPTE cells and cloned the PCR products into the pCR2.1-TOPO vector. We identified specific inserts from individual colonies that matched the splice junctions detected in the

185 RNA-seq analysis (Figure 3B, lane 1-8, 11-12). Sanger sequencing of the DNA isolated from individual  
 186 colonies confirmed the PCR products were perfectly matched with the predicted splicing patterns. Most  
 187 interestingly, we identified an RT-PCR product whose sequence indicates that the RNA polymerase  
 188 circled the genome four times (Figure 3B, lane 9).

189

## 190 Discussion

191 Most polyomaviruses encode a miRNA from the late strand of their circular genome. Although the  
 192 genetic location of the miRNA and the target sequence differ among the polyomaviruses, all the  
 193 miRNAs share in common the ability to regulate expression of the major viral regulatory protein, TAg  
 194 (23-26, 33). We previously proposed a model we called the promoter balance model to explain how the  
 195 NCCR regulates viral early gene, late gene, and miRNA expression (23). In archetype virus, the NCCR  
 196 drives low early promoter activity and high late promoter activity, resulting in a relatively high level of  
 197 miRNA, which further inhibits viral early gene expression and virus replication. In rearranged variants,  
 198 the NCCR has weak late promoter activity and high early promoter activity, resulting in a high level of  
 199 early gene expression and robust virus replication. However, how the viral miRNA was regulated or if  
 200 there any other promoter sequence regulates miRNA expression is not fully understood. A previous  
 201 study with Merkel cell polyomavirus (MCPyV) showed that an approximately 100 base pair sequence  
 202 upstream of the miRNA coding sequence contains a promoter activity that can initiate viral miRNA  
 203 expression independently of NCCR-initiated transcription (34). Unlike these authors, we did not detect  
 204 any obvious spike in miRNA reads in our RNA-seq analysis of BKPyV, although our RNA-seq  
 205 preparations were not optimized for small RNAs. We have also begun to dissect the viral cis-acting  
 206 sequences that regulate miRNA expression. Our preliminary results suggest that there are sequences  
 207 both within and outside the NCCR that serve this function (W.Z., unpublished data).

208 Previous studies have shown that most cellular miRNAs are expressed from intron sequences.  
 209 However, the BKPyV miRNA is not encoded in an obvious intron. The apparent answer to this dilemma  
 210 was first demonstrated by the Carmichael lab in their studies of MuPyV. They determined that RNA  
 211 polymerase II can generate a primary RNA transcript that circles the genome more than once, due to  
 212 the presence of a weak late strand polyadenylation site (31, 35). Their results showed that genome-  
 213 length presumptive introns could be spliced out of these primary transcripts that represented tandem  
 214 copies of the viral genome. Viral late transcripts that circle the genome more than once have also  
 215 observed more recently in MCPyV-infected PSFK-1 cells (34). Based on these results, we took three

approaches to address the question of whether the BKPyV miRNA can be expressed without being present in an apparent intron. First, we showed that a circular DNA template expresses higher levels of the miRNA than a linear template. Second, we detected the NCCR in viral RNA transcripts by both RT-PCR amplification and RNA-seq analysis. Third, we identified splice junctions that could only be generated if primary transcripts circled the viral genome more than once. Interestingly, we found evidence of primary RNAs that circle the genome up to four times, albeit at a low frequency. In the MuPyV study, a sequence they named the leader sequence was found in the mature late gene mRNA (31, 32). This sequence contained multiple tandem copies of an untranslated 57 nucleotide segment located at the 5' end, which indicated that the primary viral RNA circled the genome more than once. It was also shown that the length of the leader sequence is important for virus viability, which includes regulating both pre-mRNA splicing and stability. In these experiments, precise donor and acceptor sites were used for splicing of the leader sequence and an exact genome length intron was excised. In our study, we found using more sensitive assays that the splice junctions occur at different locations, albeit with enrichment in the agnoprotein and VP1 genes, producing different tandem exons and excised introns. Although back splicing or splicing of circular RNAs also could generate the splice junctions that we detect, we think this is a less likely explanation.

Why the virus uses this complicated way to regulate viral late gene expression is not clear. However, since the coding capacity of small viruses like polyomavirus is limited, the virus could take advantage of this unique method of expressing both proteins and a miRNA despite their genome size constraint. The MuPyV studies showed that the leader-to-leader splicing is also essential for leader-VP1 splicing, which is required for mature VP1 mRNA production (32). Since these splice junctions occur at different locations on the BKPyV primary transcripts, the function of these transcripts is unknown and warrants further study.

Taken together, these results indicate the complexity of the regulation of viral RNA expression. We confirmed that the miRNA is likely expressed from an intron that would be present in viral primary transcripts. The RNA-seq data showed that the expression level of the NCCR is about 10% of the late

gene expression level. While this may indicate that only a portion of the RNA transcripts circle the genome more than once, since RNA-seq only measures steady state RNA, the NCCR may be transcribed at a higher rate but the products are not detectable because the putative intron would be degraded after splicing and processing of the miRNA.

Although it is well known that the miRNA from different polyomaviruses share the common ability to regulate expression of viral TAg, which has been proposed to modulate persistent infection (23, 34, 36, 37), the miRNAs may also harbor distinct functions. The SV40 miRNA has been shown to downregulate TAg expression as a means of immune evasion, reducing cytotoxic T lymphocyte susceptibility of infected cells (26). However, a mutant MuPyV virus that does not express the miRNA exhibited the same infection and transformation phenotypes as wild type MuPyV, and the mutant induced an indistinguishable immune response as the wild type virus in mice (24). On the other hand, the MuPyV miRNA seems to control viruria in acutely-infected animals (36). It has also been reported that the BKPyV and JCPyV miRNAs target the cellular stress-induced ligand ULBP3 to allow the virus to escape detection by the immune system (38). In addition, the MCPyV miRNA regulates cellular mRNA expression (39). Illustrating how the expression of polyomavirus miRNAs are regulated will shed light on their function in pathogenesis and oncogenesis.

**Acknowledgements.** The authors would like to thank Andrew Tai for critical reading of the manuscript. This work was supported by NIH/NIAID grant AI060584 awarded to M.J.I. and NIH/NINDS grants R01NS043097 and P01NS065719 to W.J.A. Research reported in this publication was also supported by the NIH/NCI under Award Number P30CA046592 by the use of the following Cancer Center Shared Resource: DNA Sequencing Core.

## References

1. Kean JM, Rao S, Wang M, Garcea RL. 2009. Seroepidemiology of human polyomaviruses. PLoS Pathog 5:e1000363.
2. Egli A, Infanti L, Dumoulin A, Buser A, Samaridis J, Stebler C, Gosert R, Hirsch Hans H. 2009. Prevalence of Polyomavirus BK and JC Infection and Replication in 400 Healthy Blood Donors. The Journal of Infectious Diseases 199:837-846.
3. Bennett SM, Broekema NM, Imperiale MJ. 2012. BK polyomavirus: emerging pathogen. Microbes Infect 14:672-83.
4. Jiang M, Abend JR, Johnson SF, Imperiale MJ. 2009. The role of polyomaviruses in human disease. Virology 384:266-73.
5. Hirsch HH. 2005. BK virus: opportunity makes a pathogen. Clin Infect Dis 41:354-60.
6. Cantalupo PG, Katz JP, Pipas JM. 2018. Viral sequences in human cancer. Virology 513:208-216.
7. Robertson AG, Kim J, Al-Ahmadie H, Bellmunt J, Guo G, Cherniack AD, Hinoue T, Laird PW, Hoadley KA, Akbani R, Castro MAA, Gibb EA, Kanchi RS, Gordenin DA, Shukla SA, Sanchez-Vega F, Hansel DE, Czerniak BA, Reuter VE, Su X, de Sa Carvalho B, Chagas VS, Mungall KL, Sadeghi S, Pedamallu CS, Lu Y, Klimczak LJ, Zhang J, Choo C, Ojesina AI, Bullman S, Leraas KM, Lichtenberg TM, Wu CJ, Schultz N, Getz G, Meyerson M, Mills GB, McConkey DJ, Network TR, Weinstein JN, Kwiatkowski DJ, Lerner SP. 2018. Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer. Cell 174:1033.
8. Kenan DJ, Mieczkowski PA, Burger-Calderon R, Singh HK, Nickeleit V. 2015. The oncogenic potential of BK-polyomavirus is linked to viral integration into the human genome. J Pathol 237:379-89.
9. Rogers R, Gohh R, Noska A. 2017. Urothelial cell carcinoma after BK polyomavirus infection in kidney transplant recipients: A cohort study of veterans. Transpl Infect Dis 19.

- 289 10. Gupta G, Kuppachi S, Kalil RS, Buck CB, Lynch CF, Engels EA. 2018. Treatment for presumed  
290 BK polyomavirus nephropathy and risk of urinary tract cancers among kidney transplant  
291 recipients in the United States. *Am J Transplant* 18:245-252.
- 292 11. Zeng Y, Sun J, Bao J, Zhu T. 2020. BK polyomavirus infection promotes growth and  
293 aggressiveness in bladder cancer. *Virol J* 17:139.
- 294 12. Kuypers DR. 2012. Management of polyomavirus-associated nephropathy in renal transplant  
295 recipients. *Nat Rev Nephrol* 8:390-402.
- 296 13. Verhalen B, Starrett GJ, Harris RS, Jiang M. 2016. Functional Upregulation of the DNA Cytosine  
297 Deaminase APOBEC3B by Polyomaviruses. *J Virol* 90:6379-6386.
- 298 14. Starrett GJ, Serebrenik AA, Roelofs PA, McCann JL, Verhalen B, Jarvis MC, Stewart TA, Law  
299 EK, Krupp A, Jiang M, Martens JWM, Cahir-McFarland E, Span PN, Harris RS. 2019.  
300 Polyomavirus T Antigen Induces APOBEC3B Expression Using an LXCXE-Dependent and  
301 TP53-Independent Mechanism. *MBio* 10.
- 302 15. Manzetti J, Weissbach FH, Graf FE, Unterstab G, Wernli M, Hopfer H, Drachenberg CB,  
303 Rinaldo CH, Hirsch HH. 2020. BK Polyomavirus Evades Innate Immune Sensing by Disrupting  
304 the Mitochondrial Network and Promotes Mitophagy. *iScience* 23:101257.
- 305 16. Panou MM, Prescott EL, Hurdiss DL, Swinscoe G, Hollinshead M, Caller LG, Morgan EL,  
306 Carlisle L, Muller M, Antoni M, Kealy D, Ranson NA, Crump CM, Macdonald A. 2018.  
307 Agnoprotein Is an Essential Egress Factor during BK Polyomavirus Infection. *Int J Mol Sci* 19.
- 308 17. Swinscoe GE. 2019. Identification of BK polyomavirus agnoprotein as a novel viroporin  
309 doi:<https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.805334>.
- 310 18. Moens U, Johansen T, Johnsen JI, Seternes OM, Traavik T. 1995. Noncoding control region of  
311 naturally occurring BK virus variants: sequence comparison and functional analysis. *Virus*  
312 *Genes* 10:261-75.
- 313 19. Moens U, Van Ghelue M. 2005. Polymorphism in the genome of non-passaged human  
314 polyomavirus BK: implications for cell tropism and the pathological role of the virus. *Virology*  
315 331:209-31.

- 316 20. Doerries K. 2006. Human polyomavirus JC and BK persistent infection. *Adv Exp Med Biol*  
317 577:102-16.
- 318 21. Gosert R, Rinaldo CH, Funk GA, Egli A, Ramos E, Drachenberg CB, Hirsch HH. 2008.  
319 Polyomavirus BK with rearranged noncoding control region emerge in vivo in renal transplant  
320 patients and increase viral replication and cytopathology. *J Exp Med* 205:841-52.
- 321 22. McNees AL, Harrigal LJ, Kelly A, Minard CG, Wong C, Butel JS. 2018. Viral microRNA effects  
322 on persistent infection of human lymphoid cells by polyomavirus SV40. *PLoS One* 13:e0192799.
- 323 23. Broekema NM, Imperiale MJ. 2013. miRNA regulation of BK polyomavirus replication during  
324 early infection. *Proc Natl Acad Sci U S A* 110:8200-5.
- 325 24. Sullivan CS, Sung CK, Pack CD, Grundhoff A, Lukacher AE, Benjamin TL, Ganem D. 2009.  
326 Murine Polyomavirus encodes a microRNA that cleaves early RNA transcripts but is not  
327 essential for experimental infection. *Virology* 387:157-67.
- 328 25. Seo GJ, Chen CJ, Sullivan CS. 2009. Merkel cell polyomavirus encodes a microRNA with the  
329 ability to autoregulate viral gene expression. *Virology* 383:183-7.
- 330 26. Sullivan CS, Grundhoff AT, Tevethia S, Pipas JM, Ganem D. 2005. SV40-encoded microRNAs  
331 regulate viral gene expression and reduce susceptibility to cytotoxic T cells. *Nature* 435:682-6.
- 332 27. Jiang M, Abend JR, Tsai B, Imperiale MJ. 2009. Early events during BK virus entry and  
333 disassembly. *J Virol* 83:1350-8.
- 334 28. Varkonyi-Gasic E, Wu R, Wood M, Walton EF, Hellens RP. 2007. Protocol: a highly sensitive  
335 RT-PCR method for detection and quantification of microRNAs. *Plant Methods* 3:12.
- 336 29. Assetta B, De Cecco M, O'Hara B, Atwood WJ. 2016. JC Polyomavirus Infection of Primary  
337 Human Renal Epithelial Cells Is Controlled by a Type I IFN-Induced Response. *MBio* 7.
- 338 30. Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A. 2004. Identification of mammalian  
339 microRNA host genes and transcription units. *Genome Res* 14:1902-10.
- 340 31. Batt DB, Carmichael GG. 1995. Characterization of the polyomavirus late polyadenylation  
341 signal. *Mol Cell Biol* 15:4783-90.



32. Adami GR, Marlor CW, Barrett NL, Carmichael GG. 1989. Leader-to-leader splicing is required for efficient production and accumulation of polyomavirus late mRNAs. *J Virol* 63:85-93.
33. Seo GJ, Fink LH, O'Hara B, Atwood WJ, Sullivan CS. 2008. Evolutionarily conserved function of a viral microRNA. *J Virol* 82:9823-8.
34. Theiss JM, Gunther T, Alawi M, Neumann F, Tessmer U, Fischer N, Grundhoff A. 2015. A Comprehensive Analysis of Replicating Merkel Cell Polyomavirus Genomes Delineates the Viral Transcription Program and Suggests a Role for mcv-miR-M1 in Episomal Persistence. *PLoS Pathog* 11:e1004974.
35. Acheson NH. 1978. Polyoma virus giant RNAs contain tandem repeats of the nucleotide sequence of the entire viral genome. *Proc Natl Acad Sci U S A* 75:4754-8.
36. Burke JM, Bass CR, Kincaid RP, Ulug ET, Sullivan CS. 2018. The Murine Polyomavirus MicroRNA Locus Is Required To Promote Viruria during the Acute Phase of Infection. *J Virol* 92.
37. Zhang S, Sroller V, Zanwar P, Chen CJ, Halvorson SJ, Ajami NJ, Hecksel CW, Swain JL, Wong C, Sullivan CS, Butel JS. 2014. Viral microRNA effects on pathogenesis of polyomavirus SV40 infections in syrian golden hamsters. *PLoS Pathog* 10:e1003912.
38. Bauman Y, Nachmani D, Vicenshtein A, Tsukerman P, Drayman N, Stern-Ginossar N, Lankry D, Gruda R, Mandelboim O. 2011. An identical miRNA of the human JC and BK polyoma viruses targets the stress-induced ligand ULBP3 to escape immune elimination. *Cell Host Microbe* 9:93-102.
39. Xie H, Lee L, Caramuta S, Hoog A, Browaldh N, Bjornhagen V, Larsson C, Lui WO. 2014. MicroRNA expression patterns related to merkel cell polyomavirus infection in human merkel cell carcinoma. *J Invest Dermatol* 134:507-517.

Table 1. Summary of reads mapped to viral genome from RNA-seq data

| Time point                | 3 dpi      |            |            | 6 dpi      |            |            | 9 dpi      |            |            |
|---------------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| Repeat                    | 1          | 2          | 3          | 1          | 2          | 3          | 1          | 2          | 3          |
| Total Reads               | 77,980,168 | 67,149,942 | 80,478,692 | 63,718,248 | 89,026,206 | 77,592,824 | 87,857,620 | 89,707,110 | 82,655,640 |
| Mapping rate <sup>a</sup> | 5.47%      | 5.29%      | 9.63%      | 21.39%     | 19.44%     | 26.83%     | 24.80%     | 29.93%     | 22.87%     |
| Mapped to BKPvV genome    | 4,265,515  | 3,552,232  | 7,750,098  | 13,629,333 | 17,306,694 | 20,818,155 | 21,788,690 | 26,849,338 | 18,903,345 |

<sup>a</sup> ratio of the reads mapped to viral genome to total reads

Table 2. NCCR expression levels

| Time point | Experiment | Late region (RPKM) | NCCR (RPKM) | (NCCR/late)% | mean $\pm$ SD (NCCR/late)% |
|------------|------------|--------------------|-------------|--------------|----------------------------|
| 3 dpi      | 1          | 17,304             | 1,986       | 11.48        | 11.48 $\pm$ 0.67           |
|            | 2          | 17,033             | 2,095       | 12.30        |                            |
|            | 3          | 31,035             | 3,305       | 10.65        |                            |
| 6 dpi      | 1          | 66,143             | 6,561       | 9.92         | 9.47 $\pm$ 0.39            |
|            | 2          | 60,987             | 5,815       | 9.53         |                            |
|            | 3          | 84,661             | 7,582       | 8.96         |                            |
| 9 dpi      | 1          | 75,326             | 7,055       | 9.37         | 9.16 $\pm$ 0.88            |
|            | 2          | 100,071            | 7,992       | 7.99         |                            |
|            | 3          | 73,899             | 7,480       | 10.12        |                            |

Table 3. Summary of splice junctions of viral RNA

|                        |                           | 3 dpi <sup>c</sup>   |         |         | 6 dpi   |         |         | 9 dpi   |         |         | location <sup>d</sup> |
|------------------------|---------------------------|----------------------|---------|---------|---------|---------|---------|---------|---------|---------|-----------------------|
|                        |                           | 1                    | 2       | 3       | 1       | 2       | 3       | 1       | 2       | 3       |                       |
| Donor (D) <sup>a</sup> | Acceptor (A) <sup>b</sup> | BK3DPI1 <sup>c</sup> | BK3DPI2 | BK3DPI3 | BK6DPI1 | BK6DPI2 | BK6DPI3 | BK9DPI1 | BK9DPI2 | BK9DPI3 |                       |
| 3,041                  | 8,093 (2,940)             | 117,947              | 102,833 | 230,080 | 311,641 | 426,286 | 491,182 | 433,764 | 606,023 | 359,356 | Agno                  |
| 3,014                  | 8,117 (2,964)             | 20,128               | 18,445  | 40,988  | 61,794  | 80,724  | 95,341  | 82,856  | 149,456 | 93,262  | Agno                  |
| 3,041                  | 8,115 (2,962)             | 16,855               | 10,218  | 29,058  | 45,737  | 64,424  | 69,149  | 59,551  | 106,723 | 75,976  | Agno                  |
| 5,109                  | 10,104 (4,951)            | 1                    | 16,795  | 3       | 70,650  | 96,021  | 168,128 | 164,276 | 205,883 | 2       | VP1                   |
| 5,109                  | 10,231 (5,078)            | 5                    | 9       | 37,384  | 1,007   | 55      | 52      | 44      | 135,253 | 97,293  | VP1                   |
| 5,000                  | 10,114 (4,961)            | 2                    | 1,945   | 8,738   | 20,236  | 37,697  | 8,229   | 13,706  | 96,622  | 61,331  | VP1                   |
| 3,041 <sup>e</sup>     | 3,974                     | 57,854               | 52,111  | 107,881 | 150,276 | 202,325 | 248,315 | 206,876 | 344,419 | 203,219 |                       |

<sup>a</sup> We set the T in the sequence TAAGCCAC, located between the 3' ends of the VP1 and TAg coding sequences of the Dunlop genome, as nucleotide 1, which places the NCCR in the middle of the genome and facilitates comparing the reads to a tandem genome as described in the text (i.e., nucleotide 5154 is the same T in the second copy of the genome).

<sup>b</sup> the location of the acceptor site of the reads containing a splice junction in the second copy of the virus genome. The number in parentheses indicates the location of the equivalent acceptor site in the first copy of the virus genome (note that these acceptor sites in the first copy are located upstream of the donor sites).

<sup>c</sup> The number of the reads for each representative splice pattern is shown.

<sup>d</sup> the coding location of the splice junction in the viral genome. Agno, agnoprotein.

<sup>e</sup> The splice junction D3041/A3974 produces authentic VP1 mRNA.



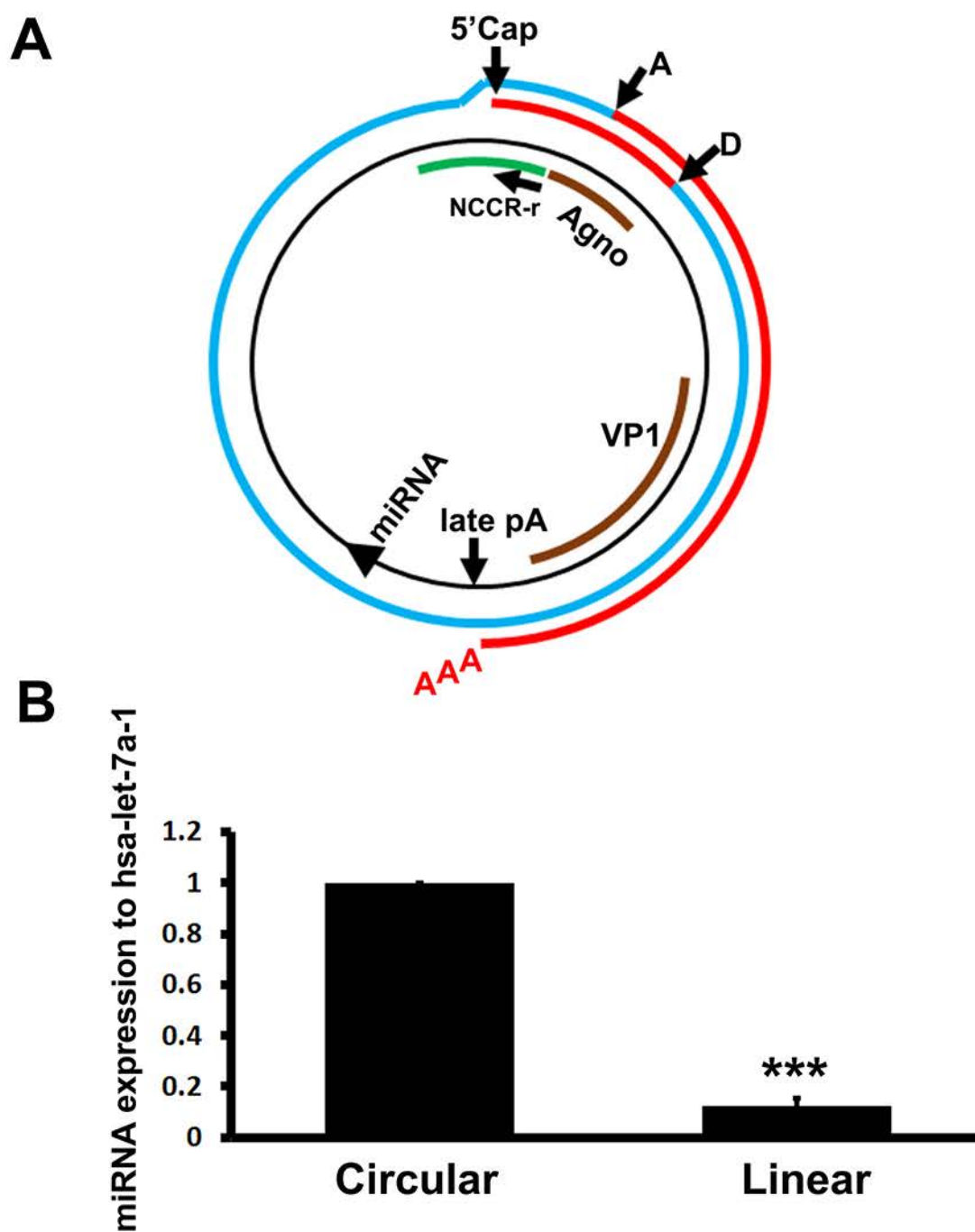
Figure 1. A circular viral template expresses higher levels of miRNA than a linear template. **(A)** Model of miRNA biogenesis, showing the pre-miRNA coming from the intron (blue line) of a primary transcript that is longer than the genome. The exon in this example (red), when spliced, forms a VP1 mRNA. D, splice donor; A, splice acceptor; Late pA, late polyadenylation signal site. The mature miRNA is processed from the putative intron. The location of the primer NCCR-r, which was used to perform reverse transcription to confirm transcription of the NCCR, is shown **(B)**. Total RNA was extracted from transfected 293 cells and assayed for BKPyV miRNA expression using stem-loop RT qPCR. The results are normalized to the cellular miRNA control, hsa-let-7a-1. The circular template value was arbitrarily set to 1. \*\*\*  $p < 0.005$ .

Figure 2. Viral RNA transcripts contain the NCCR region. RPTE cells were infected with archetype Dik virus **(A)** or rearranged variant Dunlop virus **(B)**, or mock infected. Total RNA was extracted 3 dpi. Reverse transcription (RT) and PCR amplification with specific NCCR primers (Methods and Figure 1A) were performed and the PCR products were electrophoresed on a 2% DNA agarose gel. Representative gels from three independent repeats with each virus are shown. **(C)**. RNA-seq read coverage of the Dunlop genome. One representative sample was shown. Read coverage on the late strand (red) and early strand (blue) are shown graphed along the length of the viral genome. The locations of the non-coding control region (NCCR), agnoprotein gene (Agno), early genes (EG), late genes (LG), and miRNA gene are indicated. The enlarged area showed the coverage for NCCR and Agno regions.

Figure 3. Evidence for viral primary mRNAs that circle the genome more than once. **(A)**. (1) linear representation of a primary tandem genomic transcript resulting from the RNA polymerase transcribing the genome more than once. The relevant genomic regions are shown in different colors (not to scale), and the locations of putative splice donors (D) and acceptors (A) based on the RNA-seq data are shown. A1 and D1 are in the first tandem copy, and A2 and D2 are in the second copy. Also shown are the two PCR primers (1 and 2) used to analyze the RNA products. cDNA from this unspliced RNA would not be amplified by primers 1 and 2 in our assay because the product would be  $> 5$  Kb. (2) Spliced product resulting from excision of a genome-sized intron. Primers P1 and P2 can now amplify cDNA from this RNA, yielding a product of 99 bp, or larger in size if the transcript circled the genome more than twice and was subsequently spliced. 3) Intron resulting from splicing to give the product in (2). **(B)**. TA cloning to confirm the splice junctions detected in the RNA-seq data. Specific primers shown in (A) were used to amplify the cDNA transcribed from RNA of archetype-infected RPTE cells. PCR

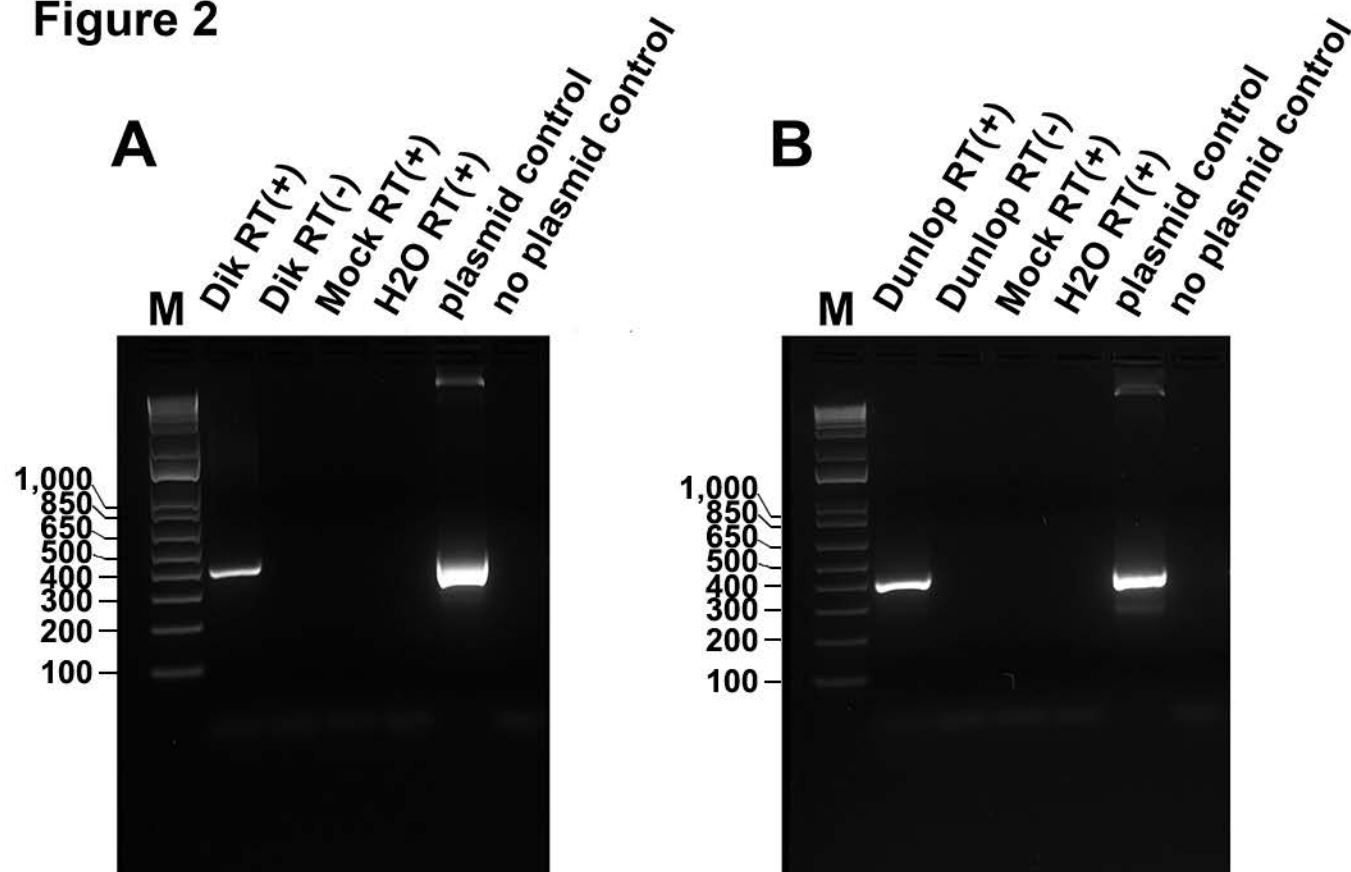
products were cloned into pCR2.1-TOPO vector, and 12 random clones were chosen for DNA minipreps. The DNA was digested with BamHI and XbaI and electrophoresed on a 2% DNA agarose gel. The sizes of the bands are 94 bp longer than the cDNA insert due to the location of the restriction sites used to excise the inserts. The diagrams to the left of the gel indicate the structure of the spliced products represented by the different sized bands. Lane 10 is empty pCR2.1-TOPO vector digested with BamHI and XbaI, which produces a 94 bp band because there is no insert.

# Figure 1

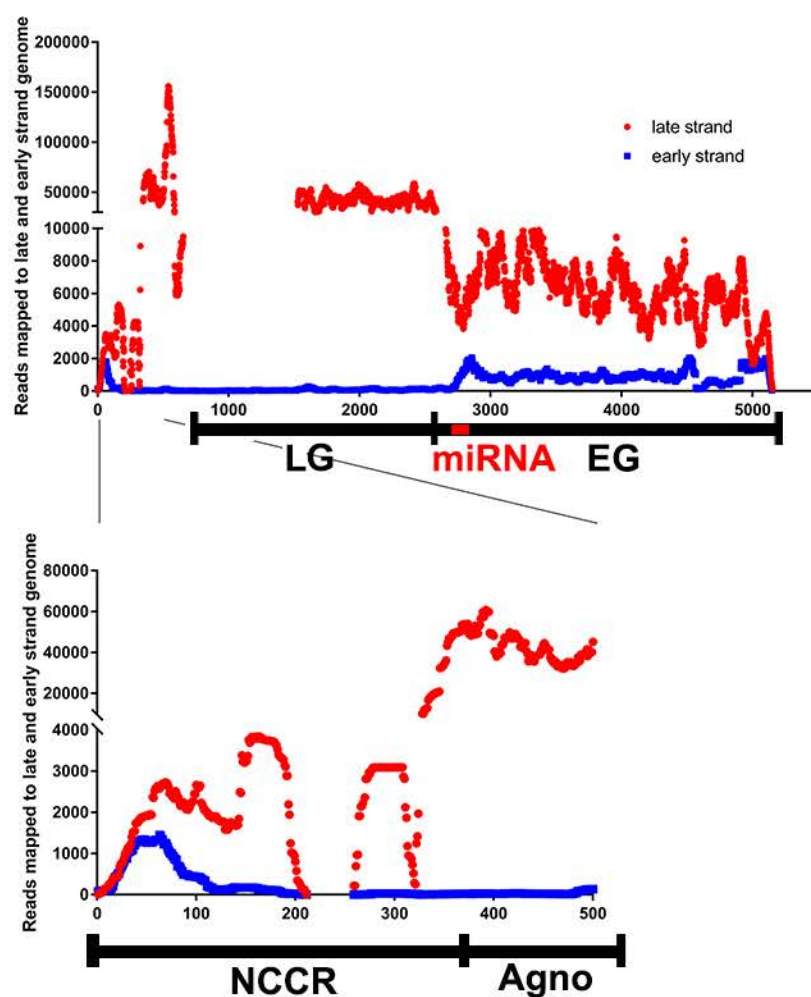




## Figure 2

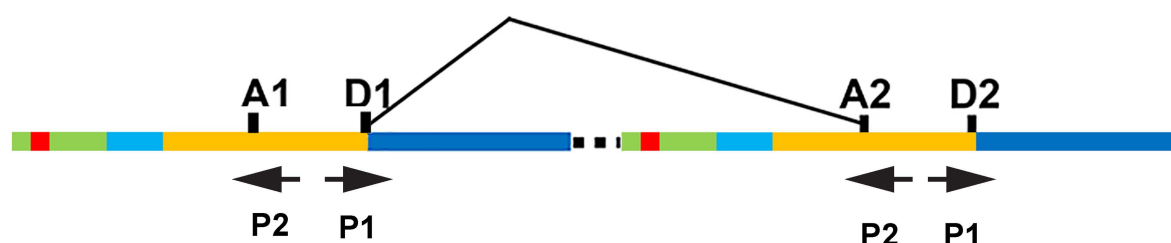


## C



**A**

1. unspliced primary tandem genome transcript



splicing

2. spliced exons



+

3. excised genome length intron that encodes miRNA



■ miRNA; ■ EG; ■ NCCR; ■ Agnoprotein; ■ LG

**B**

