Published in final edited form as:

Curr Opin Virol. 2013 June ; 3(3): 276–284. doi:10.1016/j.coviro.2013.04.006.

Complexities of gammaherpesvirus transcription revealed by microarrays and RNAseq

Laurie T. Krug

Department of Molecular Genetics and Microbiology, Stony Brook University, Stony Brook, NY 11794

Abstract

Technological advances in genome-wide transcript analysis, referred to as the transcriptome, using microarrays and deep RNA sequencing methodologies are rapidly extending our understanding of the genetic content of the gammaherpesviruses. These vast transcript analyses continue to uncover the complexity of coding transcripts due to alternative splicing, translation initiation and termination, as well as regulatory RNAs of the gammaherpesviruses. A full assessment of the transcriptome requires that our analysis be extended to the virion and exosomes of infected cells since viral and host mRNAs, miRNAs, and other non-coding RNAs seem purposefully incorporated to exert function upon delivery to naïve cells. Understanding the regulation, biogenesis and function of the recently discovered transcripts will extend beyond pathogenesis and oncogenic events to offer key insights for basic RNA processes of the cell.

Introduction

Infection with the gammaherpesviruses (γ HV) Epstein-Barr virus (EBV) or Kaposi's sarcoma-associated herpesvirus (KSHV) places an individual at risk for developing multiple cancers. This is especially prevalent within immune deficient patients, as well as populations in specific geographical regions. Without vaccines to prevent the establishment of latent infection and antiviral drugs to treat latent cells, novel strategies for targeted eradication are needed. An understanding of transcript biogenesis and function may provide new molecular targets for the development of therapeutic interventions.

Prior to genome-wide transcript analyses, original annotations of herpesvirus genomes portrayed a rather straightforward genomic organization. ORFs were predominantly non-overlapping and configured in a similar order and orientation within gene blocks across closely related viruses, with virus-specific unique genes interspersed within these blocks or at the genomic termini. However, the relative simple map of rightward and leftward genes with occasional overlap and splicing has recently become complicated by a vast assortment of novel RNA transcripts emanating from coding and 'non-coding' regions, discovered by the technological feats of genome wide-transcript analyses- namely tiled DNA microarrays and RNAseq.

© 2013 Elsevier B.V. All rights reserved.

*Corresponding author: Department of Molecular Genetics and Microbiology, Stony Brook University, Life Sciences Bldg. Room 130, Stony Brook, NY 11794, 631-632-9055, laurie.krug@stonybrook.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Primary herpesvirus infections are generally a short productive course of replication that serves to disseminate an expanded number of infectious particles. Resolution of primary infection leads to the establishment of a quiescent infection in a cell-type deemed a latency reservoir that is poised to reactivate and undergo productive infection upon intracellular signals or cues from the microenvironment. The γ HV have long been recognized for the distinct lytic and latent viral gene expression programs that are characteristic of these disparate types of infections. However, the pace of the discovery of alternative splicing profiles of viral genes, the usage of multiple promoters and transcription initiation sites, and the production of number non-coding RNAs has been accelerated by genome-wide analysis.

Here we focus on the discovery, biogenesis, and function of novel viral transcripts in γ HV infection. A substantial portion of the transcriptome is a direct consequence of the chromatin structure and milieu of transcription factors bound to the viral episome (*as reviewed elsewhere in this issue*). The control of viral gene expression per se, is not the focus of this review. Instead, we highlight recent advances in identifying novel γ HV transcripts and explore the consequences of novel regulatory and coding RNAs for virus infection and associated oncogenic events.

Comparison of technologies for transcriptome analysis

Genome-wide, high-resolution analysis of γ HV transcription began with the use of custom-designed microarrays comprised of overlapping, single-stranded oligonucleotides representing the entire genome. A tiled microarray approach provides strand-specific data, redundancy in coverage, and resolution of transcript boundaries that is directly proportional to the degree of oligonucleotide overlap. Benefits of the array include dual color hybridization with dual fluorescent labeling of samples and an economical, well-established platform of reagents, methodologies, hardware, and bioinformatics pipelines, and the inclusion of oligonucleotides for host gene profiling (Table 1). RNAseq is an unbiased global sequencing approach with strand-specificity that enables the identification of spliced exon junctions and isoforms (Table 1). As high-throughput sequencing costs diminish and the bioinformatics toolbox becomes more standardized, RNAseq will likely supplant microarrays. At the present, the technologies are complimentary and as reviewed here, both have revealed unanticipated complexities of γ HV transcriptomes.

Analysis of transcript kinetics and structure

The signal for overlapping oligonucleotides that tile a viral genome in a microarray or sequencing reads aligned to a viral genome in an RNAseq experiment provides a robust quantitative measure of gene expression (Figure 1A). Viral genes may be assigned to immediate early, early, or late gene classes based on transcript sensitivity to drugs that block protein synthesis (early and late genes) or DNA replication (late genes) (Figure 1B). This classical approach provides critical information regarding gene regulation and suggests an ordered cascade of expression. However, genes may also be clustered by their pattern of expression over the course of a naturally, progressive infection. This temporal-based hierarchical classification reveals staggered phases of viral gene transcript initiation and peak expression, wherein genes classified as early or late might be expressed concomitantly with other classes [*1].

Low-level spontaneous reactivation in latent cell culture systems makes the designation of a viral gene as latent or lytic difficult; however, there is a clear trend for the detection of latency-associated gene transcripts during productive γ HV infection [*1,*2,*3,**4]. Limiting dilution analysis followed by RT-PCR enables the analysis of latent transcripts in such a background [5], but genome-wide analysis of single cells by RNAseq is technically challenging. Interestingly, the cytoplasmic to nuclear ratio of EBV latency genes is lower

than lytic gene counterparts upon B cell receptor stimulation [**4] suggesting a mechanism that retains select transcripts to function in the nucleus.

Overlapping transcripts that share a polyA signal are difficult to distinguish by either RNAseq or tiled array. Time course analysis permits the identification of boundaries for ORFs with early expression kinetics before that ORF signal is masked by adjacent ORFs [*1] or long read-through transcription occurs. For example, the MHV68 ORF6 5' boundary is only distinguished from the upstream ORF4 transcript signal by tiled array in a transient period of 4–6 hpi [*1]. The sequencing of a cDNA library is an alternative approach to isolating single transcripts, but introduces bias for abundant and shorter spliced isoforms. Regardless, molecular validation of transcript structure by 5' and 3' RACE, Northern blot hybridization, and S1 nuclease assays in the context of latent and productive infections in multiple cell types is required to fully catalogue multiple transcript isoforms that will certainly inform the expression profile and function for a gene of interest.

Splicing—Extensive splicing of viral transcripts has long been a hallmark of EBV latency gene expression programs. Paired-end deep sequencing has identified over a hundred novel splice junctions for both latent and lytic transcripts of EBV [**4,**6]. The differential frequency of some transcript isoforms such as those of the lytic transactivator BZLF1 and the viral bcl-2 BHRF1 suggests that gene expression is regulated at the level of RNA splicing [**4,**6]. Additionally, the EBV latency associated genes LMP-1 and LMP2A have markedly different splice patterns in latent cells compared to reactivated cells upon BCR stimulation [**4]. Knowledge of alternative transcript structures also provides insight to gene regulation. MHV68 RTA has multiple 5' exons that lead to alternative promoter usage [7] and a novel, unspliced transcript for LANA was detected by microarray and validated by 5' RACE during MHV68 infection [*1]. Alternative splicing leads to significant functional outcomes; a protein derived from a spliced isoform of EBV BZLF1 functions as a dominant-negative to inhibit transactivation by full-length BZLF1 [**4]. The impact of splicing on the regulation of γ HV gene expression and the consequences of generating protein isoforms with different roles during infection is a ripe area of investigation.

Transcript Discovery

Expressed genomic regions—Genome-wide transcript analyses indicate that a substantial portion of the viral genome outside of the annotated ORFs is expressed over the course of a productive infection [*1,*2,*3,**4]. Johnson and colleagues [*3] designated MHV68 transcript signals beyond the putative UTRs of annotated ORFs or antisense to annotated ORFs as expressed genomic regions (EGRs). This nondescript designation is appropriate until the biogenesis and regulatory potential of these elements are characterized and their coding potential is determined as described below. Some EGRs align almost completely anti-parallel to annotated ORFs [*2,*3]. This might be attributed to either aberrant transcription of the non-coding strand of a highly transcribed and open region of the genome or a process of transcriptional interference wherein antisense transcripts that overlap with the sense-strand promoter may clear transcription initiation complexes to impair transcription or possibly lead to premature termination of RNA polymerase [8]. Lastly, due to the high level of transcription on both strands of gammaherpesvirus genomes, investigators should be cautious in attributing transcript levels in a genomic region to a particular ORF since most RT-PCR assays are not strand-specific and do not span splice junctions.

Small RNAs—Although many γ HV small RNAs including miRNAs, tRNA pri-miRNAs, and Sm-class small nuclear RNAs were identified prior to deep sequencing, RNAseq of

small RNA-sized fractions will continue to be instrumental in the comprehensive capture of viral and cellular small RNAs and a tool of target transcript discovery [9–12]. Interestingly, some γ HV use atypical processes for generating non-coding small RNAs to promote high-level expression that have significant impacts on infection. For example, EBV EBERs 1 and 2 are RNAPol III-driven transcripts with dsRNA structure that may promote inflammatory responses [13], but the recent finding of an interaction of EBER1 with the host AU-rich element binding factor AUF1 [14] suggests that EBER1 may modulate the stability of mRNA with AU-rich elements during infection. In addition, Herpesvirus saimiri encodes Sm-class small nuclear RNAs (HSURs) embedded with miRNAs that are found in and are processed by the nuclear Integrator complex [15]. HSUR1 binds to and drives the degradation of the host miR-2, that in turn, promotes expression of miR-27 targets such as the tumor suppressor FOXO1 [16]. Finally, MHV68 pre-microRNA is generated from the 3' ends of RNA polymerase III-driven tRNA-like transcripts that are initially processed by the host tRNA processing machinery independent of Dicer [17,18]. The role of many small RNAs in the context of pathogenesis remains to be determined.

RNAseq of transcripts associated with the RNA induced silencing complex (RISC) has also been applied to define the target mRNAs that are regulated by miRNAs in the context of infection [19–21]. A more comprehensive review of miRNA targets and functions are found elsewhere in this issue and reviewed in [22–24]. However the identification of γ HV small RNAs including miRNAs in the virion and exosomes of infected cells is described below.

non-coding RNA—Long non-coding RNAs (lncRNAs), some of which span greater than 10 Kb regions of the genomes have been reported for KSHV, EBV, and MHV68 [*2,*3]. Not surprisingly, many of these lncRNAs are antisense to genes critical for both latency and lytic replication processes[25]. While most lncRNAs have only been identified by genome-wide transcript analysis and validated by Northern, the investigation of the function and biogenesis of a single lncRNA, the KSHV polyadenylated nuclear RNA (PAN), has unveiled a rich and dynamic interaction of viral and host factors that modulate the transcriptome of γ HV infections [26]. Similar to host lncRNAs that modulate chromatin regulators to impact gene expression [27], KSHV PAN modulates the chromatin structure of the RTA promoter, thereby promoting lytic cycle gene expression [28]. The KSHV ncRNA PAN has cis elements that promote high transcript abundance and nuclear retention. The 3' end of PAN forms a major-groove triple helix that likely prevents deadenylation by host factors that promote RNA decay [29]. ORF57 recruits host Rev/Aly export factors to a binding site in the 5' UTR of PAN to promote nuclear stability of PAN [30]. Lastly, PAN is also bound by the polyA binding protein PABPC1 to promote late lytic gene expression [31]. Thus transitioning from a catalogue of mapped lncRNAs such as the recently reported EBV BCRT2 and BCLT1 transcripts [**4] to validation and functional studies will certainly unlock further novel virus-host interactions that regulate viral gene expression.

Transcription in the oriLyt regions of EBV and KSHV may promote access and recruitment of replication factors [32,33]. BHLF1 is a highly expressed oriLyt-associated transcript [**6], yet may not possess a coding function since some EBV strains contain stop codons in the BHLF1 transcript [34]. The repeated observation of an A to G change between the DNA and RNA of BHLF1 indicates RNA editing by the host adenosine deaminase acting on RNAs (ADAR) and an intriguing link between RNA editing and DNA replication [**6]. LF3 is another viral transcript containing stop codons in some EBV strains, again suggesting a non-coding, regulatory role during infection [34]. Taken together, recently discovered EGRs and antisense transcripts may function as long non-coding RNAs to influence multiple aspects of infection.

Evaluating the coding potential of novel transcripts

Novel transcripts with short ORFs (< 80 amino acids) were previously dismissed as non-coding. However, many EGRs including transcripts antisense to γ HV lytic and latency genes possess the coding potential for hundreds of predicted short ORFs [2,3,25,35]. Indeed, antisense transcripts identified by tiled array as having low coding potential based on the short ORFs they contained were isolated from the cytoplasmic polysome fraction during infection (Figure 2A), suggesting occupancy and translation by multiple ribosomes [2,25]. Indeed, two of the three predicted ORFs within an antisense transcript to KSHV RTA were detected by in-frame fusions with GFP. Jaber and Yan [36] recently reported that vSP-1, a 48 aa peptide encoded in the T3.0 transcript antisense to RTA, binds and stabilizes the RTA transactivator. In addition, annotated ORFs may harbor internal in-frame ORFs that produce truncated isoforms; Feng et al. [37] discovered an ORF for the mitochondrial anti-apoptotic protein, vMAP, internal to and out-of-frame with ORF57 of MHV68.

Validation of novel ORFs can be a painstaking process, but recent higher throughput, non-biased analyses have been reported. Dresang and colleagues [35] merged array-based transcriptomics data with mass spectrometry proteomics datasets to identify new EBV and KSHV transcripts with coding potential. While their mass spec analysis did not identify peptides for all annotated ORFs, peptides derived from short ORFs in novel transcripts were identified by tiled array. Thus, peptide spectra of infected cells may serve as a tool for novel protein and associated transcript discovery (Figure 2A and 2B).

There are also lessons to be learned from the recent approaches of Stern-Ginossar and colleagues [38] that unearthed a breathtaking amount of novel coding regions in the human cytomegalovirus (HCMV) genome (Figure 2C). The analysis of RNA bound by stalled ribosomes enabled them to identify hundreds of translation initiation sites, indicative of 5' starts of ORFs during HCMV infection. These novel ORFs were located internal to, between, and antisense to annotated ORFs, and ranged in size from less than 20 aa to greater than 80 aa. Longer ORFs were primarily alternate polypeptide forms of previously annotated ORFs generated by differential exon splicing. Whole cell proteomics of the infected cell lysates validated dozens of novel ORFs, while 24 previously annotated ORFs not detected by ribosome profiling might be non-coding. As a further validation, several novel ORFs were independently detected by epitope-tagging the ORFs in their native virus context [38]. As summarized in Figure 2, these genome-wide approaches to viral ORF discovery will identify novel γ HV molecules that may provide novel insight towards mechanisms of chronic infection and lymphomagenesis.

Short ORFs encoding peptides upstream of annotated ORFs (uORFs) provide an additional layer of post-transcriptional control of viral gene expression. For instance, two uORFs in the 5' UTR of the ORF35-37 polycistronic transcript of KSHV have differential effects on the downstream cistrons; ORF35 translation is repressed by both uORF1 and uORF2 while ORF36 translation is enhanced only by uORF2 [26]. Xu and colleagues have proposed that short ORFs may yield peptides that are presented on MHC complexes, expanding the immunogenic content of the viruses [25].

Epigenomics meets transcriptomics

Chromatin modification and transcription factor occupancy are critical elements of γ HV gene regulation. ChIPseq and RNAseq data for Epstein-Barr virus positive lymphoma cell lines are included in the HapMap and ENCODE (encyclopedia of DNA elements) consortiums [6,39]. Arvey et al. [6] integrated these vast datasets with previously generated epigenetic and viral protein binding data to begin the characterization of the

functional EBV genome. This comprehensive analysis of 143 LCLs affirmed previous observations that these cells comprise a spectrum ranging from tightly latent cells to cells with high levels of spontaneous reactivation. This analysis revealed a positive correlation of reactivation with cellular markers of proliferation coupled to EBNA1 occupancy at the promoters of these proliferative host genes. This study also uncovered hotspots for host transcription factor complexes; occupancy of the B cell transcription factor Pax5 at the terminal repeats was associated with maintenance of the circular episome and repression of latent gene expression. As public repositories of genome-wide data expand, sophisticated systems biology analyses will discover additional dynamic links between epigenetics and regulation of viral and host gene expression, building many hypotheses for mechanistic and functional investigations of the γ HV transcriptome.

RNA in the virion and exosomes

The γ HV transcriptome extends temporally beyond late gene expression and physically beyond the infected cell. Nuclease-treated EBV, KSHV, and MHV68 virions contain a striking assortment of RNAs of host and virus origin that are packaged with apparent specificity [40–43]. KSHV virion RNAs first identified by tiled array and validated by qRT-PCR included the non-coding PAN RNA and mature polyadenylated, spliced mRNA for lytic genes [40]. MicroRNAs and unusual small RNAs were recently uncovered in KSHV mature virion particles by deep sequencing. Differential enrichment of coding and non-coding RNAs in the virions compared to their abundance relative to other RNAs in host cells strongly suggests targeted incorporation [40,43], perhaps involving RNA-binding proteins of the tegument [44]. Virion-associated mRNA might comprise a portion of early lytic gene expression that precedes the canonical IE-E-L gene expression cascade observed upon *de novo* KSHV infection [45]. Thus the inclusion of actinomycin D [40] as a control treatment to inhibit RNA pol II-dependent transcription is appropriate for kinetics studies of gene expression to differentiate virion mRNAs from classical IE [*41,*42].

Importantly, the biological relevance of virion miRNAs and mRNAs as modulators of *de novo* infection events has been established using several elegant approaches. UV-treated virions harboring KSHV miRNAs can silence specific reporter gene targets upon infection of 293T cells [*42], suggesting the presence of functional miRNAs within the incoming particles. In the presence of actinomycin D, encapsidated ORF59 mRNA is translated in 293 cells infected with KSHV [40] and EBV BRLF1 is detected in newly infected B cells [*41]. Jochum et al. [*41] extended these studies by generating EBV genome-negative, yet RNA positive virion-like particles (VLPs) that led to specific CD8+T cell recognition of EBV BRLF1, BZLF1, and BMLF1 in HLA-matched B cells and enhanced EBV-infected primary B cell out-growth. VLPs produced from cells lacking EBERS failed to repress IFN responses in infected Raji cells while VLPs lacking the mRNA encoding the immunoevasin BNLF2 had increased recognition by EBV-specific CD8+ T cells [*41]. Lastly, VLPs produced from cells expressing an estrogen-regulatable BZLF1 led to transactivation of the early BMRF1 promoter in target reporter cells upon tamoxifen induction [*41]. Taken together, these studies demonstrate that mRNAs and non-coding RNAs incorporated into virions can have an immediate and biologically significant impact on *de novo* infected cells.

Virions are not the only potential source for transmission of viral transcripts to native cells. Exosomes are secreted intraluminal vesicles derived from late endosomal multivesicular bodies of the cells [46]. EBV exosomes deliver viral proteins such as LMP-1 that alter signaling and immune responses of neighboring cells [46,47] or prevent new B cell infection via CD21 receptor engagement by exosomes with EBV gp350 [48]. It is tempting to speculate that exosome-containing RNAs may function in a similar fashion to exosome proteins. Exosomes are differentially enriched with some EBV miRNAs compared to the miRNA composition of the producing cell [49,50]. Functional consequences include the

silencing of the T cell attractant chemokine CXCL11 gene expression in monocyte-derived dendritic cells [50] and the host NLRP3 to block the inflammasome of co-cultured macrophage cells [51]. Taken together, the exosome-associated transcriptome is an intriguing dimension of immune evasion and γ HV pathogenesis. Deep sequencing of γ HV exosomes and virion particles may continue to reveal that the RNA profile varies with the cell type and gene expression program of the producing cell [50], with differential consequences for the recipient cell, including cells not supportive of virus infection [52]. Importantly, since the process by which these miRNA are sorted and localized to sites for virus assembly or MVB formation likely involves RNA-binding proteins of viral and host origin [53], profiling the targetome of these binding proteins may provide mechanistic insight. Taken together, many γ HV transcripts will likely gain a new classification as virion and/or exosomal-associated RNA, spurring investigations of novel *de novo* infection functions.

RNA biogenesis and RNA editing

The strategies of RNA enrichment and processing for transcriptome profiling enable specific classes of transcripts to be analyzed (Table 2). Isolation of small RNA or poly-adenylated mRNA are common strategies to capture non-coding or coding RNA respectively, but other classes of transcripts might not be efficiently captured by current genome-wide transcript approaches. For example, the non-coding EBV EBERs [54] and MHV68 pri-miRNAs that are transcribed by RNA polymerase III [17,18] and the non-canonical processing of herpesvirus saimiri SM-class small nuclear RNAs and HSUR-miRNA fusions [55] result in transcripts that are larger than the typical small RNA cutoff and are not polyadenylated. In addition, whole cell RNA preps do not distinguish between RNA in the nucleus from RNA in the cytoplasm. This is critical as a transcript location provides important insight into its potential function. Nuclear transcripts may arise due to incomplete splicing, a block in export, a component of ribonucleoprotein complexes, or providing a regulatory function. Cytoplasmic transcripts are often associated with ribosomes, targeted to silencing complexes, or associated with other organelles.

Strikingly, considering the small size of the viral genome in relation to the host chromatin, a disproportionate level of transcript signal, ~1% of all LCL reads and ~7% of reads from induced type I latency cells map to EBV [**4,**6]. This bias is a testament to the multiple strategies by which γ HV evade host degradation pathways and redirect the RNA processing machinery to favor viral transcripts (reviewed in [55]). This dramatic take-over also poses challenges to fundamental issues such as normalization strategies when comparing the transcriptome between infected and uninfected cells [56]. Recent array and deep sequencing investigations have begun to parse the RNA composition of subcellular fractions [**4,25] and RNA-binding proteins [57] in addition to viral modulators of RNA biogenesis [58,59]. Any single transcriptome study merely provides a snapshot of transcripts that are in an active process of transcription, processing, export, and localization to organelles of translation or silencing and degradation. The resolution of these events and the true dynamic nature of the transcriptome will only be realized through a collective analysis of RNA associated with RNP-complexes, particular organelles, or in the context of altered virus-host interactions over a timecourse of infection (Table 2).

RNA editing has been reported for a few γ HV transcripts, indicated by a discrepancy between the genomic DNA sequence and the RNA transcript sequence. EBV BHLF1 and EBV pri-miR-BART6 RNAs [**4,**6] and KSHV Kaposin K12 [60] are likely targets of host ADAR RNA editing of A to G. EBV encoded BART6-miRNAs can suppress cellular Dicer and lytic transcript signals. RNA editing of the pri-miR-BART6 RNAs might represent a mechanism for suppressing lytic gene expression and maintaining latency [61]. Kaposin K12 editing increases with lytic replication and reduces the transforming potential

[60]. Thus RNA editing expands the complexity of the transcriptome beyond what would be predicted from the genomic sequence, with biological impacts on latency, lytic replication, and associated pathologies.

Concluding remarks

Arrays and RNAseq are powerful approaches to analyze changes in both virus and host gene expression in the context of wild-type or mutant infections or in gain-of-function studies wherein viral or host factors are examined outside of the context of infection. The simplistic designation of genes as either immediate early, early, or late genes or latent vs lytic will likely become less accurate as transcript isoforms and hidden ORFs continue to emerge in studies of cell culture systems, patient specimens, and animal models. Bona fide large non-coding RNAs and miRNAs will likely act in concert to impact both viral and host gene expression. Public deposition of data and the associated metadata descriptors will provide a rich resource for data mining for years to come. This will ultimately lead to the development of a gene signature for infected cells in culture, in the tissue of an infected host, or in the context of cancer. In addition, within novel transcripts, many novel peptides and polypeptides that impact the virus lifecycle await discovery. Systems biology approaches involving the integration of the transcriptome, proteome, and epigenome will enable the construction of informative network models enabling the dissection of signaling pathways that influence gene expression and ultimately reveal the molecular mechanisms that toggle the switch between latency and reactivation. The construction of a viral-host interactome will generate testable hypotheses and identify targets for effective pharmacologic and molecular interventions.

Acknowledgments

I thank Eain Murphy for discussion and editing of this manuscript. This work was supported in part by the National Institutes of Health/National Institute of Allergy and Infectious Diseases (AI097875) and the American Cancer Society Research Scholar Grant (RSG-11-160-01-MPC).

References and recommended reading

We apologize for any reference omissions due to the nature of this brief review. Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

- 1*. Cheng BY, Zhi J, Santana A, Khan S, Salinas E, Forrest JC, Zheng Y, Jaggi S, Leatherwood J, Krug LT. Tiled microarray identification of novel viral transcript structures and distinct transcriptional profiles during two modes of productive murine gammaherpesvirus 68 infection. *J Virol.* 2012; 86:4340–4357. This paper uses tiled arrays to define the kinetics of gene expression during a naturally progressing *de novo* infection or upon TPA stimulated reactivation from latency, and map transcript structures for dozens of MHV68 genes. [PubMed: 22318145]
- 2*. Chandriani S, Xu Y, Ganem D. The lytic transcriptome of Kaposi's sarcoma-associated herpesvirus reveals extensive transcription of noncoding regions, including regions antisense to important genes. *J Virol.* 2010; 84:7934–7942. This report of genome-wide transcript analysis describes extensive transcription antisense to coding regions upon reactivation of KSHV. Small ORFs are embedded in a long transcript antisense to ORF50. [PubMed: 20534856]
- 3*. Johnson LS, Willert EK, Virgin HW. Redefining the genetics of murine gammaherpesvirus 68 via transcriptome-based annotation. *Cell Host Microbe.* 2010; 7:516–526. This paper identifies transcripts outside of annotated ORFs as expressed genomic regions using tiled arrays and Northern blot hybridization. [PubMed: 20542255]

- 4**. Concha M, Wang X, Cao S, Baddoo M, Fewell C, Lin Z, Hulme W, Hedges D, McBride J, Flemington EK. Identification of new viral genes and transcript isoforms during Epstein-Barr virus reactivation using RNA-Seq. *J Virol.* 2012; 86:1458–1467. This study presents a robust RNAseq pipeline to identify novel transcripts and transcript isoforms in addition to nuclear versus cytoplasmic transcript abundances in EBV+ cell lines uninduced or induced to reactivate. [PubMed: 22090128]
5. Chandriani S, Ganem D. Array-based transcript profiling and limiting-dilution reverse transcription-PCR analysis identify additional latent genes in Kaposi's sarcoma-associated herpesvirus. *J Virol.* 2010; 84:5565–5573. [PubMed: 20219929]
- 6**. Arvey A, Tempera I, Tsai K, Chen H-S, Tikhmyanova N, Klichinsky M, Leslie C, Lieberman PM. An atlas of the Epstein-Barr virus transcriptome and epigenome reveals host-virus regulatory interactions. *Cell Host Microbe.* 2012; 12:233–245. This landmark study evaluates hundreds of RNAseq and ChIPseq profiles of EBV+ lymphoblastoid cell lines and reports novel transcripts, a correlation of viral reactivation with host proliferative gene expression, and hotspots for host DNA binding factors in the viral genome. [PubMed: 22901543]
7. Gray KS, Allen RD, Farrell ML, Forrest JC, Speck SH. Alternatively initiated gene 50/RTA transcripts expressed during murine and human gammaherpesvirus reactivation from latency. *J Virol.* 2009; 83:314–328. [PubMed: 18971285]
8. Chen HM, Rosebrock AP, Khan SR, Futcher B, Leatherwood JK. Repression of meiotic genes by antisense transcription and by Fkh2 transcription factor in *Schizosaccharomyces pombe*. *PLoS ONE.* 2012; 7:e29917. [PubMed: 22238674]
9. Zhu JY, Strehle M, Frohn A, Kremmer E, Hofig KP, Meister G, Adler H. Identification and analysis of expression of novel microRNAs of murine gammaherpesvirus 68. *J Virol.* 2010; 84:10266–10275. [PubMed: 20668074]
10. Reese TA, Xia J, Johnson LS, Zhou X, Zhang W, Virgin HW. Identification of novel microRNA-like molecules generated from herpesvirus and host tRNA transcripts. *J Virol.* 2010; 84:10344–10353. [PubMed: 20660200]
11. Xia J, Zhang W. Noncanonical microRNAs and endogenous siRNAs in lytic infection of murine gammaherpesvirus. *PLoS ONE.* 2012; 7:e47863. [PubMed: 23110115]
12. Haecker I, Gay LA, Yang Y, Hu J, Morse AM, McIntyre LM, Renne R. Ago HITS-CLIP expands understanding of Kaposi's sarcoma-associated herpesvirus miRNA function in primary effusion lymphomas. *PLoS Pathog.* 2012; 8:e1002884. [PubMed: 22927820]
13. Iwakiri D, Zhou L, Samanta M, Matsumoto M, Ebihara T, Seya T, Imai S, Fujieda M, Kawa K, Takada K. Epstein-Barr virus (EBV)-encoded small RNA is released from EBV-infected cells and activates signaling from Toll-like receptor 3. *J Exp Med.* 2009; 206:2091–2099. [PubMed: 19720839]
14. Lee N, Pimienta G, Steitz JA. AUF1/hnRNP D is a novel protein partner of the EBER1 noncoding RNA of Epstein-Barr virus. *RNA.* 2012; 18:2073–2082. [PubMed: 23012480]
15. Cazalla D, Xie M, Steitz JA. A primate herpesvirus uses the integrator complex to generate viral microRNAs. *Mol Cell.* 2011; 43:982–992. [PubMed: 21925386]
16. Cazalla D, Steitz JA. Down-regulation of a host microRNA by a viral noncoding RNA. *Cold Spring Harb Symp Quant Biol.* 2010; 75:321–324. [PubMed: 21139068]
17. Diebel KW, Smith AL, van Dyk LF. Mature and functional viral miRNAs transcribed from novel RNA polymerase III promoters. *RNA.* 2010; 16:170–185. [PubMed: 19948768]
18. Bogerd HP, Karnowski HW, Cai X, Shin J, Pohlers M, Cullen BR. A mammalian herpesvirus uses noncanonical expression and processing mechanisms to generate viral MicroRNAs. *Mol Cell.* 2010; 37:135–142. [PubMed: 20129062]
19. Gottwein E, Corcoran DL, Mukherjee N, Skalsky RL, Hafner M, Nusbaum JD, Shamulailatpam P, Love CL, Dave SS, Tuschl T, et al. Viral microRNA targetome of KSHV-infected primary effusion lymphoma cell lines. *Cell Host Microbe.* 2011; 10:515–526. [PubMed: 22100165]
20. Skalsky RL, Corcoran DL, Gottwein E, Frank CL, Kang D, Hafner M, Nusbaum JD, Feederle R, Delecluse HJ, Luftig MA, et al. The viral and cellular microRNA targetome in lymphoblastoid cell lines. *PLoS Pathog.* 2012; 8:e1002484. [PubMed: 22291592]

21. Riley KJ, Rabinowitz GS, Yario TA, Luna JM, Darnell RB, Steitz JA. EBV and human microRNAs co-target oncogenic and apoptotic viral and human genes during latency. *EMBO J*. 2012; 31:2207–2221. [PubMed: 22473208]
22. Gottwein E. Kaposi's Sarcoma-Associated Herpesvirus microRNAs. *Front Microbiol*. 2012; 3:165. [PubMed: 22563327]
23. Forte E, Luftig MA. The role of microRNAs in Epstein-Barr virus latency and lytic reactivation. *Microbes Infect*. 2011; 13:1156–1167. [PubMed: 21835261]
24. Ramalingam D, Kieffer-Kwon P, Ziegelbauer JM. Emerging themes from EBV and KSHV microRNA targets. *Viruses*. 2012; 4:1687–1710. [PubMed: 23170179]
25. Xu Y, Ganem D. Making sense of antisense: seemingly noncoding RNAs antisense to the master regulator of Kaposi's sarcoma-associated herpesvirus lytic replication do not regulate that transcript but serve as mRNAs encoding small peptides. *J Virol*. 2010; 84:5465–5475. [PubMed: 20357088]
26. Kronstad LM, Brulois KF, Jung JU, Glaunsinger BA. Dual short upstream open reading frames control translation of a herpesviral polycistronic mRNA. *PLoS Pathog*. 2013; 9:e1003156. [PubMed: 23382684]
27. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. *Annu Rev Biochem*. 2012; 81:145–166. [PubMed: 22663078]
28. Rossetto CC, Pari G. KSHV PAN RNA associates with demethylases UTX and JMJD3 to activate lytic replication through a physical interaction with the virus genome. *PLoS Pathog*. 2012; 8:e1002680. [PubMed: 22589717]
29. Mitton-Fry RM, DeGregorio SJ, Wang J, Steitz TA, Steitz JA. Poly(A) tail recognition by a viral RNA element through assembly of a triple helix. *Science*. 2010; 330:1244–1247. [PubMed: 21109672]
30. Stubbs SH, Hunter OV, Hoover A, Conrad NK. Viral factors reveal a role for REF/Aly in nuclear RNA stability. *Mol Cell Biol*. 2012; 32:1260–1270. [PubMed: 22290432]
31. Borah S, Darricarrere N, Darnell A, Myoung J, Steitz JA. A viral nuclear noncoding RNA binds re-localized poly(A) binding protein and is required for late KSHV gene expression. *PLoS Pathog*. 2011; 7:e1002300. [PubMed: 22022268]
32. Wang Y, Tang Q, Maul GG, Yuan Y. Kaposi's sarcoma-associated herpesvirus ori-Lyt-dependent DNA replication: dual role of replication and transcription activator. *J Virol*. 2006; 80:12171–12186. [PubMed: 17020951]
33. Wang Y, Li H, Tang Q, Maul GG, Yuan Y. Kaposi's sarcoma-associated herpesvirus ori-Lyt-dependent DNA replication: involvement of host cellular factors. *J Virol*. 2008; 82:2867–2882. [PubMed: 18199640]
34. Lin Z, Wang X, Strong MJ, Concha M, Baddoo M, Xu G, Baribault C, Fewell C, Hulme W, Hedges D, et al. Whole-genome sequencing of the Akata and Mutu Epstein-Barr virus strains. *J Virol*. 2013; 87:1172–1182. [PubMed: 23152513]
35. Dresang LR, Teuton JR, Feng H, Jacobs JM, Camp DG, Purvine SO, Gritsenko MA, Li Z, Smith RD, Sugden B, et al. Coupled transcriptome and proteome analysis of human lymphotropic tumor viruses: insights on the detection and discovery of viral genes. *BMC Genomics*. 2011; 12:625. [PubMed: 22185355]
- 36*. Jaber T, Yuan Y. A Virally Encoded Small Peptide Regulates RTA Stability and Facilitates Kaposi's Sarcoma-associated Herpesvirus Lytic Replication. *J Virol*. 2013 This paper is the first report of a biological function for a novel ORF encoded by an antisense transcript in the context of a gammaherpesvirus infection.
37. Feng P, Liang C, Shin YC, Xiaofei E, Zhang W, Gravel R, Wu TT, Sun R, Usherwood E, Jung JU. A novel inhibitory mechanism of mitochondrion-dependent apoptosis by a herpesviral protein. *PLoS Pathog*. 2007; 3:e174. [PubMed: 18069888]
- 38*. Stern-Ginossar N, Weisburd B, Michalski A, Le VT, Hein MY, Huang SX, Ma M, Shen B, Qian SB, Hengel H, et al. Decoding human cytomegalovirus. *Science*. 2012; 338:1088–1093. This ground-breaking report identified hundreds of novel coding regions of sense and antisense transcripts by capturing the RNA bound by stalled ribosomes at the 5' end of ORFs throughout the human cytomegalovirus genome. [PubMed: 23180859]

39. Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. 2007; 447:799–816. [PubMed: 17571346]
40. Bechtel JT, Winant RC, Ganem D. Host and viral proteins in the virion of Kaposi's sarcoma-associated herpesvirus. *J Virol*. 2005; 79:4952–4964. [PubMed: 15795281]
- 41*. Jochum S, Ruiss R, Moosmann A, Hammerschmidt W, Zeidler R. RNAs in Epstein-Barr virions control early steps of infection. *Proc Natl Acad Sci U S A*. 2012; 109:E1396–1404. EBV vRNAs delivered by virion like particles have multiple roles including viral gene transactivation, promotion of cytokine expression, immune evasion and transformation. [PubMed: 22543160]
- 42*. Lin X, Li X, Liang D, Lan K. MicroRNAs and unusual small RNAs discovered in Kaposi's sarcoma-associated herpesvirus virions. *J Virol*. 2012; 86:12717–12730. miRNAs of viral and host origin are detected in KSHV virions. [PubMed: 22973026]
43. Cliffe AR, Nash AA, Dutia BM. Selective uptake of small RNA molecules in the virion of murine gammaherpesvirus 68. *J Virol*. 2009; 83:2321–2326. [PubMed: 19109392]
44. Amen MA, Griffiths A. Packaging of Non-Coding RNAs into Herpesvirus Virions: Comparisons to Coding RNAs. *Front Genet*. 2011; 2:81. [PubMed: 22303375]
45. Krishnan HH, Naranatt PP, Smith MS, Zeng L, Bloomer C, Chandran B. Concurrent expression of latent and a limited number of lytic genes with immune modulation and antiapoptotic function by Kaposi's sarcoma-associated herpesvirus early during infection of primary endothelial and fibroblast cells and subsequent decline of lytic gene expression. *J Virol*. 2004; 78:3601–3620. [PubMed: 15016882]
46. Meckes DG Jr, Raab-Traub N. Microvesicles and viral infection. *J Virol*. 2011; 85:12844–12854. [PubMed: 21976651]
47. Verweij FJ, van Eijndhoven MA, Hopmans ES, Vondra T, Wurdinger T, Cahir-McFarland E, Kieff E, Geerts D, van der Kant R, Neefjes J, et al. LMP1 association with CD63 in endosomes and secretion via exosomes limits constitutive NF-kappaB activation. *EMBO J*. 2011; 30:2115–2129. [PubMed: 21527913]
48. Vallhov H, Gutzeit C, Johansson SM, Nagy N, Paul M, Li Q, Friend S, George TC, Klein E, Scheynius A, et al. Exosomes containing glycoprotein 350 released by EBV-transformed B cells selectively target B cells through CD21 and block EBV infection in vitro. *J Immunol*. 2011; 186:73–82. [PubMed: 21106852]
49. Meckes DG Jr, Shair KH, Marquitz AR, Kung CP, Edwards RH, Raab-Traub N. Human tumor virus utilizes exosomes for intercellular communication. *Proc Natl Acad Sci U S A*. 2010; 107:20370–20375. [PubMed: 21059916]
50. Pegtel DM, Cosmopoulos K, Thorley-Lawson DA, van Eijndhoven MA, Hopmans ES, Lindenberg JL, de Gruijl TD, Wurdinger T, Middeldorp JM. Functional delivery of viral miRNAs via exosomes. *Proc Natl Acad Sci U S A*. 2010; 107:6328–6333. [PubMed: 20304794]
51. Haneklaus M, Gerlic M, Kurowska-Stolarska M, Rainey AA, Pich D, McInnes IB, Hammerschmidt W, O'Neill LA, Masters SL. Cutting edge: miR-223 and EBV miR-BART15 regulate the NLRP3 inflammasome and IL-1beta production. *J Immunol*. 2012; 189:3795–3799. [PubMed: 22984081]
52. Bellingham SA, Coleman BM, Hill AF. Small RNA deep sequencing reveals a distinct miRNA signature released in exosomes from prion-infected neuronal cells. *Nucleic Acids Res*. 2012; 40:10937–10949. [PubMed: 22965126]
53. Pegtel DM, van de Garde MD, Middeldorp JM. Viral miRNAs exploiting the endosomal-exosomal pathway for intercellular cross-talk and immune evasion. *Biochim Biophys Acta*. 2011; 1809:715–721. [PubMed: 21855666]
54. Minarovits J, Hu LF, Marcsek Z, Minarovits-Kormuta S, Klein G, Ernberg I. RNA polymerase III-transcribed EBER 1 and 2 transcription units are expressed and hypomethylated in the major Epstein-Barr virus-carrying cell types. *J Gen Virol*. 1992; 73 (Pt 7):1687–1692. [PubMed: 1321209]

55. Kronstad LM, Glaunsinger BA. Diverse virus-host interactions influence RNA-based regulation during gamma-herpesvirus infection. *Curr Opin Microbiol.* 2012; 15:506–511. [PubMed: 22682977]
56. Loven J, Orlando DA, Sigova AA, Lin CY, Rahl PB, Burge CB, Levens DL, Lee TI, Young RA. Revisiting global gene expression analysis. *Cell.* 2012; 151:476–482. [PubMed: 23101621]
57. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M Jr, Jungkamp AC, Munschauer M, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell.* 2010; 141:129–141. [PubMed: 20371350]
58. Clyde K, Glaunsinger BA. Deep sequencing reveals direct targets of gammaherpesvirus-induced mRNA decay and suggests that multiple mechanisms govern cellular transcript escape. *PLoS ONE.* 2011; 6:e19655. [PubMed: 21573023]
59. Kang JG, Pripuzova N, Majerciak V, Kruhlak M, Le SY, Zheng ZM. Kaposi's sarcoma-associated herpesvirus ORF57 promotes escape of viral and human interleukin-6 from microRNA-mediated suppression. *J Virol.* 2011; 85:2620–2630. [PubMed: 21209110]
60. Gandy SZ, Linnstaedt SD, Muralidhar S, Cashman KA, Rosenthal LJ, Casey JL. RNA editing of the human herpesvirus 8 kaposin transcript eliminates its transforming activity and is induced during lytic replication. *J Virol.* 2007; 81:13544–13551. [PubMed: 17913828]
61. Iizasa H, Wulff BE, Alla NR, Maragkakis M, Megraw M, Hatzigeorgiou A, Iwakiri D, Takada K, Wiedmer A, Showe L, et al. Editing of Epstein-Barr virus-encoded BART6 microRNAs controls their dicer targeting and consequently affects viral latency. *J Biol Chem.* 2010; 285:33358–33370. [PubMed: 20716523]

Highlights

- Genome-wide transcript studies define novel splicing events and transcript boundaries for annotated genes.
- Recently identified expressed genomic regions (EGRs) are transcripts that are intergenic or antisense to annotated genes.
- Antisense viral transcription is extensive and has regulatory and coding potential.
- Viral RNAs are selectively packaged in the virions and exosomes to modulate target cell processes.
- Regulation, biogenesis, and RNA function studies require systems biology integration of transcriptome.

A. Tiled array of MHV68 genome **B. Impact of protein and DNA synthesis inhibitors**

Figure 1. Tiled microarray analysis of herpesvirus transcript structure and expression kinetics
A. RNA was isolated from fibroblast cells newly infected with murine gammaherpesvirus 68 (MHV68) at an MOI of 10 plaque forming units (PFU) per cell at the indicated timepoints and hybridized to a tiled microarray. The Integrated Genome Browser (IGB) was used to visualize the distribution of transcript signal along annotated ORFs (red, top bars) and intergenic expressed genomic regions (EGRs-gray, bottom bars). The exons of ORF50 are illustrated by a connecting bar. **B.** Transcript abundance in cycloheximide-treated samples to identify IE gene ORF50 and in phosphonoacetic acid-treated samples to identify early gene ORF54 and late gene M7.

Figure 2. Novel transcripts with coding potential discovered by genome-wide analysis

A. Novel spliced isoforms of previously annotated ORFs might be uncovered by microarray and RNAseq. Alternative splicing might alter the coding potential of an ORF as discovered by a proteomics analysis. **B.** Novel coding ORFs uncovered by microarray and RNAseq are validated by a proteomic analysis. **C.** The initiation of novel internal and antisense ORFs can be identified as transcript regions bound by stalled ribosomes.

Table 1

Comparison of genome-wide transcript analyses

	Microarray	RNAseq
Resolution	Strand-specific with linear amplification, overlapping transcripts and polycistrons difficult to distinguish	Strand-specific with dUTP incorporation in second-strand synthesis
Boundaries	Dependent on tiled probe overlap; 3' bias with polyA labeling	Dependent on sequencing depth of reads; 5' and 3' boundaries
Splicing	Inferred from distinct exons	Discovery optimal with paired-end sequencing
Virus coverage	Custom-designed tiled probes enable complete coverage	Entire genome Enrichment with primers possible but costly
Host coverage	Exome-specific	Genome-wide
Expression kinetics	Semi-quantitative, cost-effective for timecourse analysis	Semi-quantitative, Replicates costly
Bioinformatic tools	Established	Rapidly evolving
Sensitivity	Ideal for productive infections	Ideal for latent specimens and in vivo analysis

Table 2

RNA samples for targeted transcriptomics

Type of RNA	polyA; 5' capped; size-selected for small RNA, miRNA, or lncRNA
Source of RNA	wild-type or mutant virus infections; knock-out or silenced factors; ectopic expression of host or viral factors
Cellular localization	whole cell; nuclear; cytoplasmic; organelle; virion; exosome
Associated factors	stalled ribosome or polysomes; Ribonucleoprotein complexes mediating RNA biogenesis and export; RNA induced silencing complex