Simultaneous high-resolution analysis of vaccinia virus and host cell transcriptomes by deep RNA sequencing

Zhilong Yang[a], Daniel P. Bruno[b], Craig A. Martens[b], Stephen F. Porcella[b], and Bernard Moss[a,1]

[a]Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892-3210; and [b]Research Technologies Section, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, MT 59840

Contributed by Bernard Moss, May 11, 2010 (sent for review May 1, 2010)

Deep RNA sequencing was used to simultaneously analyze vaccinia virus (VACV) and HeLa cell transcriptomes at progressive times following infection. VACV, the prototypic member of the poxvirus family, replicates in the cytoplasm and contains a double-stranded DNA genome with ≈200 closely spaced open reading frames (ORFs). The acquisition of a total of nearly 500 million short cDNA sequences allowed construction of temporal strand-specific maps of the entire VACV transcriptome at single-base resolution and analysis of over 14,000 host mRNAs. Before viral DNA replication, transcripts from 118 VACV ORFs were detected; after replication, transcripts from 93 additional ORFs were characterized. The high resolution permitted determination of the precise boundaries of many mRNAs including read-through transcripts and location of mRNA start sites and adjacent promoters. Temporal analysis revealed two clusters of early mRNAs that were synthesized in the presence of inhibitors of protein as well as DNA synthesis, indicating that they do not correspond to separate immediate- and delayed-early classes as defined for other DNA viruses. The proportion of viral RNAs reached 25–55% of the total at 4 h. This rapid change, resulting in a relative decrease of the vast majority of host mRNAs, can contribute to the profound shutdown of host protein synthesis and blunting of antiviral responses. At 2 h, however, a minority of cellular mRNAs was increased. The overrepresented functional categories of the up-regulated RNAs were NF-κB cascade, apoptosis, signal transduction, and ligand-mediated signaling, which likely represent the host response to invasion.

human transcriptome | poxvirus transcriptome | RNA-seq | vaccinia virus mRNA | virus–host interaction

Viral infections are frequently accompanied by the perturbation of cellular gene expression in addition to the activation of viral genes. The development of new DNA-sequencing technologies should allow the simultaneous high-resolution analysis of viral and cellular mRNAs from infected cells. Vaccinia virus (VACV), the prototypic member of the poxvirus family, provides an interesting system to test this powerful approach, as viral gene expression is temporally regulated and host protein synthesis is profoundly affected (1). Poxviruses are large double-stranded DNA viruses with members that infect vertebrate or invertebrate species. Variola virus, the causative agent of smallpox, was eradicated from nature following human immunization with VACV, with which it has ≈90% sequence identity. In addition to serving as a vaccine against other poxviruses, VACV has been developed as a recombinant expression vector that has demonstrated utility for immunological studies in animals and as a platform for veterinary and human vaccines (2).

The double-stranded VACV DNA genome is ≈200 kbp long with more than 200 closely spaced open reading frames (ORFs), of which 12 are present in inverted terminal repetitions and therefore diploid (1). VACV transcription occurs in the cytoplasm and is regulated in a cascade fashion by stage-specific transcription factors that recognize distinct early, intermediate, and late promoter sequences (3–5). RNA polymerase and early transcription factors are packaged in infectious virus particles, allowing synthesis of early mRNAs within minutes after infection. In contrast, synthesis of intermediate mRNAs requires de novo synthesis of viral early proteins and DNA; synthesis of late mRNAs additionally requires intermediate gene expression. Diminution of host protein synthesis follows infection, and a profound shutdown correlates with the onset of VACV late gene expression. Determination of the expression kinetics of numerous VACV genes has been carried out over the years by analysis of protein synthesis or Northern blotting of RNAs and in some cases refined by primer extension or nuclease digestion. Recently, VACV genome-wide transcription maps were constructed using tiling microarrays (6, 7). In separate studies, microarrays were used to analyze cellular mRNAs following VACV infection (8, 9). However, ultrahigh-throughput DNA sequencing promises genome-wide transcriptome analysis at higher resolution, greater sensitivity, and without problems due to background and the need for corrections related to GC content and efficiency of hybridization to short microarray probes.

A recent report describes mRNA sequencing of mimivirus-infected amoeba cells in which 633,346 reads were analyzed (10). Here, we obtained almost 500 million short cDNA sequences for simultaneous temporal analysis of the VACV and cellular transcriptomes. Viral transcripts were detected at 0.5 h and 25–55% of the total mRNAs were virus-encoded at 4 h. These data allowed us to construct VACV genome-wide, strand-specific maps at single-base resolution as a function of time postinfection and to analyze over 14,000 cellular mRNAs.

Results

Deep Sequencing of Total Polyadenylated RNA from VACV-Infected Cells. HeLa cells were infected with purified VACV at a multiplicity sufficient to infect all cells (*SI Materials and Methods*). Total polyadenylated RNA was isolated at 0, 0.5, 1, 2, and 4 h postinfection in two independent experiments (WTA-A and WTA-B), and cDNAs were prepared and sequenced with an Applied Biosystems SOLiD analyzer. At each time about 40 million reads were obtained, of which approximately half could be aligned to the viral and human genomes (Table S1). There was a rapid increase in viral mRNA numbers and a gradual decrease in the majority of cellular mRNAs during a 4-h period (Fig. S1 *A* and *B*). In WTA-A, ≈25% of the sequence reads at 4 h were viral, whereas 55% were viral at that time in WTA-B

Author contributions: Z.Y. and B.M. designed research; Z.Y. and D.P.B. performed research; S.F.P. contributed new reagents/analytic tools; Z.Y., C.A.M., and B.M. analyzed data; and Z.Y. and B.M. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The cDNA sequences have been deposited in the Sequence Read Archive of the National Library of Medicine.

[1]To whom correspondence should be addressed. E-mail: bmoss@niaid.nih.gov.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1006594107/-/DCSupplemental

MICROBIOLOGY

(Fig. S1*C*). This difference may be due to variation in the metabolic state of the cells or the efficiency of infection. In both experiments, however, the cDNA sequences complemented 97% of the nucleotides of each viral DNA strand at 4 h (Fig. S1*D*). At 2 h, in the presence of cytosine arabinoside (AraC), an inhibitor of DNA replication, or cycloheximide (CHX), an inhibitor of mRNA translation, ≈60% of the VACV genome was transcribed, which was similar to the amount in the absence of inhibitors at 2 h (Fig. S1*D*).

Genome-Wide VACV Transcriptome Map. The reads were aligned with the genome to construct single-base-resolution maps of the VACV transcriptome. The number of reads (counts) per nucleotide are displayed along with the annotated ORFs in Fig. 1. Read counts above and below the line represent the abundance of the cDNAs prepared from mRNAs that are transcribed from the upper strand in the rightward direction and the lower strand in the leftward direction, respectively. Viral mRNAs were detected at 0.5 h and increased from 1 to 2 h, although the pattern remained unchanged. Up to 2 h, the majority of transcripts mapped near the ends of the genome, with those on the left predominantly on the lower strand and those on the right predominantly on the upper strand. However, at 4 h the transcription pattern drastically changed, correlating with the onset of DNA replication followed by late mRNA synthesis (Fig. 1).

No Evidence for a Delayed-Early Class of VACV Genes. Some viruses and phage have distinct immediate and delayed-early classes of genes that are expressed before DNA replication. In those cases, transcription of delayed-early genes is dependent on prior synthesis of a small number of immediate-early proteins. If this paradigm holds for VACV, then both classes of early mRNAs would be made in the presence of AraC, but only immediate-early mRNAs would be produced in the presence of CHX. However, the VACV genome-wide transcriptome patterns observed after infections with AraC and CHX were qualitatively similar (Fig. 2) and indistinguishable from the pattern obtained at 2 h in the absence of drug (Fig. 1). No genes expressed only or significantly more in the presence of AraC than CHX were identified in side-by-side comparisons (Fig. S2), indicating all VACV early genes could be classified as immediate-early. Some RNAs were more abundant in the presence of CHX than AraC (Fig. S2), perhaps because they were particularly sensitive to degradation induced by an early protein(s) encoded by VACV, such as the VACV early decapping enzyme D9 (11).

Analysis of Reads That Mapped to a Well-Characterized 16-kbp Genome Segment. The transcripts of relatively few clusters of VACV genes have been mapped in detail by high-resolution methods such as primer extension and nuclease protection. An exception are the 13 ORFs in the HindIII D fragment (12, 13). To assess the accuracy of the transcriptional map obtained by deep RNA sequencing, we compared the present data with those previously obtained for the HindIII D transcripts. In the presence of AraC, transcripts aligned to the D1R, D4R, D5R, D7R, D9R, and D12L ORFs (Fig. 3*A*). Reads complementary to the entire length of the ORF were obtained in each case, although the read counts per nucleotide were not uniform. In contrast, there were no or very few reads aligned to D2L, D3R, D8L, and D13L in the presence of AraC. Even though there were sequences corresponding to D6R, D10R, and D11L in the presence of AraC, those reads only covered the region immediately adjacent to the 3′ ends of highly expressed upstream early genes, suggesting that they are read-through transcripts. RNAs corresponding to D2L, D3R, D8L, and D13L as well as the entire D6R, D10R, and D11L ORFs were abundant at 4 h in the absence of AraC (Fig. 3*A*). Thus, we

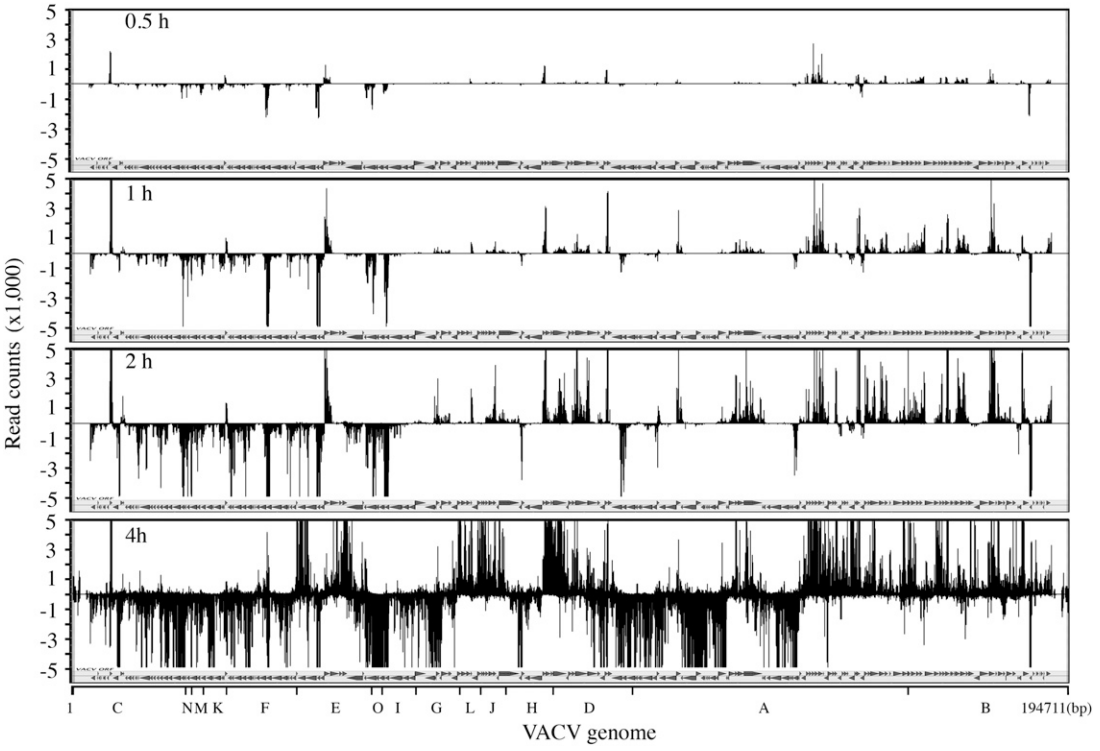

Fig. 1. VACV genome-wide transcriptome maps during the first 4 h of infection. The number of sequence reads per nucleotide was determined from the WTA-B time course and displayed over the ORF map of the entire VACV genome. The counts above the line map to the upper (rightward) DNA strand and counts below the line map to the lower (leftward) DNA strand. The highest read counts are off-scale in the 1- to 4-h samples for display purposes. The HindIII restriction map of the VACV genome is shown at the bottom for reference purposes.

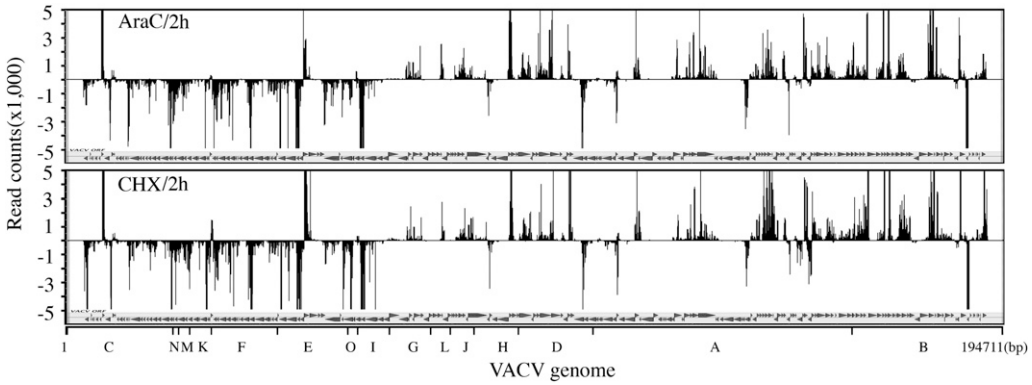

Fig. 2. VACV transcriptome of cells infected in the presence of AraC or CHX. HeLa cells were infected for 2 h in the presence of AraC or CHX, and genome-wide transcriptome maps are displayed as in Fig. 1.

classified D1R, D4R, D5R, D7R, D9R, and D12L as early and D2L, D3R, D6R, D8L, D10R, and D11L as postreplicative (PR). These results match previous characterization of this region by conventional molecular methods.

The large number of short sequence reads allowed us to zoom to single-base resolution. Fig. 3*B* shows the agreement obtained by superimposing the RNA start sites preceding the D1R ORF determined by primer extension (12) over the deep RNA sequencing map in the presence of AraC (Fig. 3*B*). Agreements were also obtained by comparing the transcriptome map with the previously determined start sites for the D4R, D7R, D9R, and D12L ORFs (See below). The accord with previous mapping of individual early transcripts in the HindIII D region provided confidence for the deep RNA sequencing method for the entire genome.

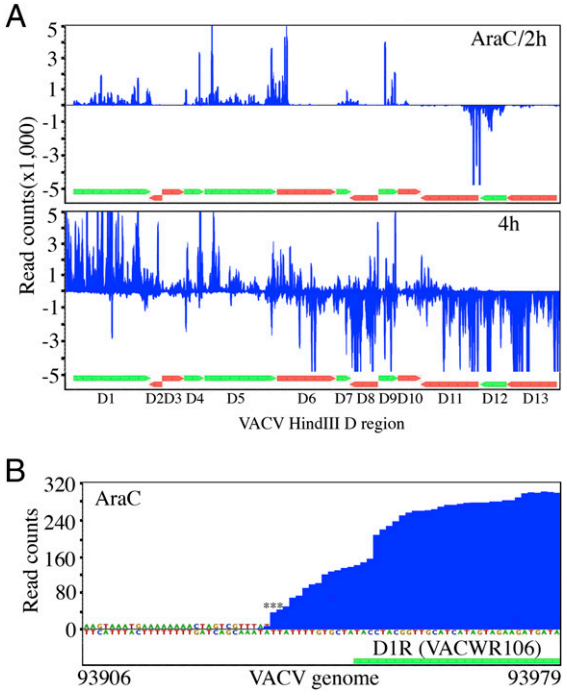

Fig. 3. HindIII D region of VACV transcriptome. (*A*) Enlargement of the HindIII D region of the AraC and 4-h transcriptome maps in Figs. 1 and 2. VACVWR ORFs 106–118 correspond to D1–D13 using the Copenhagen nomenclature. The ORFs in green indicate early genes and the ORFs in red indicate late genes. (*B*) Single-base-resolution plot of the start region of ORF 106 (D1R). Asterisks indicate the transcriptional start sites of D1R previously identified by primer extension.

Cluster Analysis of VACV Gene Expression. Cluster analysis was performed by an unbiased hierarchical method to identify temporal classes of viral RNAs. The read counts of the VACV ORFs at each time point in WTA-A and WTA-B are listed in Table S2. An initial heat map revealed two major clusters (Fig. S3). Major cluster 1 contained 131 ORFs, including most of those previously characterized as early; major cluster 2 contained 80 ORFs, including most of those previously characterized as intermediate and late. However, cluster 1 also contained the ORFs D6R, D10R, and D11L in the HindIII D fragment that exhibited read-through from an upstream early gene in the presence of AraC. To identify additional read-through RNAs, the transcription pattern of each ORF was carefully examined. Thirteen ORFs in major cluster 1 (Table S3) were designated as read-through and classified as PR, because in each case the count coverage at the 5′ end of the ORF was continuous from upstream but there were no or few counts mapping to the 3′ end of the ORF in the presence of AraC, whereas each was highly expressed at 4 h in the absence of drug. After removing the ORFs that appeared to contain read-through RNAs, the cluster analysis was repeated. The refined heat map is shown in Fig. 4*A*.

Cluster 1 was resolved into two subclusters, C1.1 and C1.2. RNAs of C1.1 were expressed earlier and at higher levels than those of C1.2, as seen in the heat map (Fig. 4*A*) and line plot (Fig. 4*B*). Nevertheless, C1.1 and C1.2 both contain early genes that are expressed in the presence of CHX as well as AraC, so that the distinction is solely based on temporal expression. We refer to the RNAs in the C1.1 and C1.2 subclusters as E1.1 and E1.2. The PR RNAs in cluster 2 could also be grouped into subclusters; however, these did not segregate the previously characterized intermediate and late RNAs. The corresponding line plot of C2 (omitting the "false-positive" read-through ORFs) is shown in Fig. 4*B*. The expression-class assignments of individual ORFs are listed in Table S4 with a comparison with previous results.

Most of the E1.1 and E1.2 genes are located at the right terminus of the upper strand and left terminus of the lower strand of the ORF map (Fig. 4*C*). The central region of the genome also contains some E1.2 genes and fewer E1.1 genes. The PR genes (cluster C2) are mainly in the central region, with predominance on the lower strand. In Fig. 4*D*, the genes encoding E1.1, E1.2, and PR RNAs were grouped into functional categories based on literature annotations or listed as "unknown" if no information was available. The genes encoding proteins for transcription were predominantly E1.2 and PR, whereas the genes encoding enzymes and factors for genome replication, which precedes intermediate and late gene expression, were mostly E1.1 and E1.2. An exception was the gene encoding the nuclease that resolves DNA concatemers, which is expressed late (14). The genes encoding proteins involved in host interactions

MICROBIOLOGY

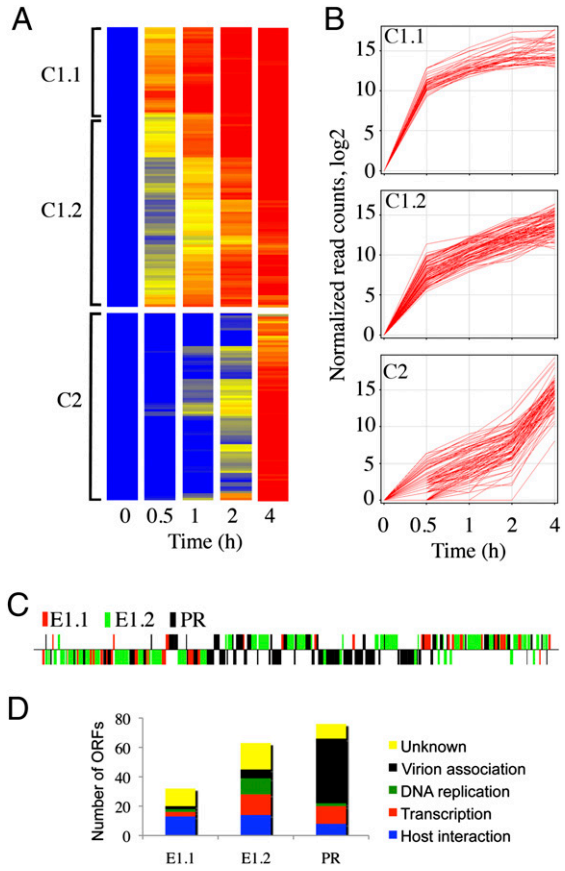

Fig. 4. Cluster analysis of the temporal expression of VACV mRNAs. (*A*) Heat-map representation of the normalized read counts of VACV ORFs from 0 to 4 h in the absence of drugs. The 13 ORFs that showed extensive read-through from upstream ORFs in the presence of AraC and classified as PR (Table S3) were excluded from the analysis. Colors from blue to yellow to red indicate increased numbers of the normalized read counts of each ORF. The clusters labeled C1.1 and C1.2 contain genes expressed early, whereas C2 contains genes expressed after DNA replication. (*B*) Line plots of read counts corresponding to *A*. (*C*) Map of VACV genome showing ORFs transcribed to the right (above line) and to the left (below line) colored red, green, or black to represent cluster 1.1 (E1.1), cluster 1.2 (E1.2), and cluster 2 (PR). (*D*) Assigned functions of genes in the different expression classes.

were predominantly in E1.1 and E1.2, and most of the virion components were in the PR class.

Analysis of the 5′ Ends of VACV Transcripts and Prediction of Promoter Motifs. As shown in Fig. 3*B*, the 5′ end of a transcript could be discerned by zooming to single-nucleotide resolution. In the same manner, we could align the 5′ ends of 83 of the 118 early RNAs of which only 19 had been previously mapped to individual nt (Table S5). Overlapping of upstream RNAs prevented the identification of the start sites of other early and PR RNAs. The distance from the RNA start to the initiation codon of the ORF varied from 1 to 320 nt, indicating a mean length of 40 nt for RNA untranslated leaders, excluding several RNAs with start sites that mapped within the annotated ORFs. The RNA start sites for VACWR061 and VACWR174 were 27 and 64 nt, respectively, downstream of the predicted AUG initiation codon. However, in each case, we noted a second AUG codon several nt downstream of the newly determined transcriptional start sites. Furthermore, the second AUG of both ORFs is conserved in all orthopoxviruses, whereas the first is conserved in only a subset. For VACWR046, the RNA start site was mapped 2 nt downstream of the start of the annotated ORF. In this case, there is a more highly conserved AUG start codon 11 nt downstream, again suggesting that the actual ORF is shorter than annotated. Although the VACWR101 ORF is expressed late in VACV infection (15, 16), we detected an early RNA start site within the middle of the ORF (Table S5), confirming a previous report (17). There are two conserved AUG sequences near the RNA start site within the VACWR101 ORF: The one 3 nt upstream would result in a protein of 102 amino acids; the other is downstream and would result in a 73-amino acid protein.

In addition to confirming or revising ORF annotations, determination of the RNA start sites allowed us to locate the core promoter regions of early genes. We used the motif discovery program MEME (18) to analyze the 50 bp upstream of the 83 RNA start sites. The program derived a 15-nt consensus sequence (AAAA-TGAAAA—A; Fig. S4) which closely corresponded to the core promoter originally described (4). The degree to which each predicted promoter matched the consensus is shown in Table S5. Overall, the promoters from the E1.1 ORFs were closer to the consensus than those from E1.2, although the *P* value of 0.11 obtained by the two-tailed Student's *t* test was not highly significant.

Cellular Transcriptome Profile Following VACV Infection. Cellular reads, obtained at the same time as the viral reads, were aligned to the RefSeq database (National Center for Biotechnology Information). RNAs corresponding to more than 50% of the 14,000 genes analyzed were relatively decreased at 4 h in experiment WTA-A, whereas >75% were decreased in WTA-B. In contrast, fewer host mRNAs were increased. At 2 h, before the massive down-regulation of cell mRNAs and viral late gene expression, 119 genes that were increased 2-fold or more in both WTA-A and WTA-B were sorted into functional groups using the PANTHER biological process classification system. The observed number of genes versus the numbers expected by chance within a certain annotation group was determined. Overrepresented categories or subcategories with *P* values less than 0.05 were NF-κB cascade, apoptosis, inhibition of apoptosis, signal transduction, and ligand-mediated signaling (Fig. 5*B*). At 4 h, the only statistically significant overrepresented category was chromatin packaging and remodeling.

Discussion

Analysis of the VACV transcriptome is complicated by the close packing of genes and is exacerbated by the occurrence of mRNAs that read through downstream ORFs. Termination of early transcripts is signaled by the RNA sequence UUUUUNU, in which N represents any nucleotide (19, 20). The efficiency of termination is ≈80% after a single termination motif (21); although some transcripts have multiple termination signals, most have one, others none, and in some, UUUUUNU is unrecognized because of secondary structure (22). Accordingly, RNA polymerase may continue to transcribe downstream ORFs until encountering functional termination sequences. If two early genes were adjacent, then the read-through would not impact gene classification. However, if the transcript of an early gene overlaps an adjacent late gene, then the latter could be misclassified as early. The high resolution provided by deep RNA sequencing allowed us to discern 13 ORFs that were partially overlapped by read-through transcripts made in the presence of AraC and were grouped with the early class by cluster analysis (Table S3). The absent or very low sequence reads mapping to the 3′ ends of these ORFs indicated that they are not true early genes, which was supported by their complete transcription at 4 h in the absence of AraC. In addition, we suspect that a few very short ORFs with relatively high level counts in the presence of AraC may be completely overlapped by read-through transcripts and therefore difficult to classify. After correcting for the read-through transcripts, we concluded that 118 early genes were expressed in the absence of

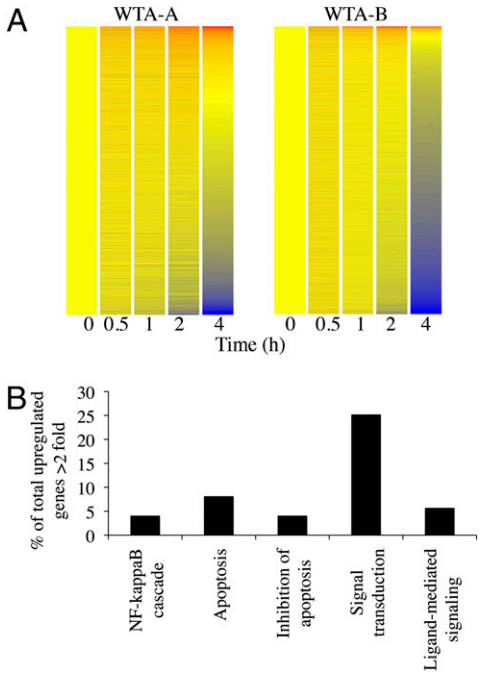

Fig. 5. Analysis of cellular mRNA changes during VACV infection. (*A*) Heat maps. The read counts of each cellular mRNA were normalized by the sum of the total reads, and the fold changes from 0 time were calculated. The heat maps of WTA-A and WTA-B were generated according to the order of the fold changes at 4 h postinfection. Colors from yellow to blue indicate down-regulated cellular genes; colors from yellow to red indicate up-regulated cellular genes. (*B*) Functional classification of genes with RNAs that increased at least 2-fold in both WTA-A and WTA-B. The up-regulated genes were analyzed in DAVID 6.7β (*SI Materials and Methods*), and the PANTHER biological process terms with a P value < 0.05 were chosen and displayed. The percentages of the genes of total up-regulated genes in overrepresented categories are shown at 2 h. Up-regulated genes are: NF-κB cascade: CARD9, LIF, NF-KB2, TNFRSF21, RELB; apoptosis: RELT, CARD9, CXCR4, EMP1, LIF, NFKB2, SPHK1, TNFRSF10D, TNFRSF21, RELB; inhibition of apoptosis: RELT, LIF, NFKB2, SPHK1, RELB; signal transduction: EPHA8, GPR3, GPR87, RELT, S100A13, STARD13, ARMC5, CAPN2, CLCF1, CARD9, CXCR4, DGKD, DOK3, EGFR, GJB3, GNRH1, HBEGF, IL1RAP, IRAK2, JUN, KALRN, LIF, NFKB2, PLG2G16, PTPRH, SLC1A5, SPHK1, TNFRSF10D, TNFRSF21, UCN2, RELB; and ligand-mediated signaling: GPR3, CLCF1, GNRH1, HBEGF, SLC1A5, UCN2, LIF.

DNA replication. Also noted were some small antisense transcripts, the significance of which remains to be determined.

Temporal analyses indicated that the early genes formed two subclusters, which differed slightly when two independent experiments were analyzed separately. Importantly, the RNA species within the two early subclusters were synthesized in the presence of CHX, an inhibitor of protein synthesis, as well as AraC, indicating they do not comprise separate immediate- and delayed-early classes as defined for other DNA viruses and phage (23–25). In those other systems, immediate-early RNAs are made in the absence of de novo protein synthesis, whereas delayed-early RNAs are dependent on synthesis of one or a few immediate-early proteins. Indeed, based on such a classification, all VACV early transcripts belong to the immediate class, as previously proposed (26). Similar although not identical early clusters were found by tiling array (7), but expression in the presence of CHX was not investigated in that study and the important distinction between immediate- and delayed-early gene expression was not made. To preclude misinterpretation, we avoid the immediate- and delayed-early terminology and refer to two temporal early classes (E1.1 and E1.2) of VACV mRNAs. The E1.1 and E1.2 genes were not distinct with regard to their functional class; for example, members of both are involved in transcription, DNA replication, and host interactions, and they are interspersed within the genome. However, the predicted core promoter motifs of the E1.1 genes more closely corresponded to the consensus than did those of E1.2, although the difference was not statistically highly significant. Some possibilities to account for their different kinetics include higher-affinity transcription factor binding sites for E1.1 promoters, other DNA–protein interactions, and DNA packing within the core.

It is more difficult to map PR transcripts than early mRNAs, as the early termination motif is not recognized despite its frequent presence and consequently they are heterogeneous in length and overlap adjacent ORFs (27, 28). Indeed, at 4 h, we found that the RNA sequences covered both DNA strands nearly in their entirety, consistent with the annealing of late RNA to form long duplex structures (29). For this reason, we did not analyze mRNAs at later times. Our classification of 93 PR genes was based on the kinetics of expression and dependence on DNA replication. We recognize two limitations of this analysis. First, with only one time point following DNA replication, we could not distinguish intermediate- and late-class RNAs. Second, because some genes have early and late or early and intermediate promoters (30, 31), these would be classified as early in our study. We intend to rectify the first ambiguity by further experiments with a VACV mutant defective in late gene expression (32) and by detailed kinetic analysis following the synchronization of VACV DNA replication. In addition, sequence analysis of the capped ends of VACV mRNAs should allow identification of distinct RNA start sites when tandem promoters precede an ORF. Table S4 shows that the deep sequencing identification of early and PR genes agreed with 145 of 151 literature annotations, with the caveat that we did not distinguish between genes with intermediate and late promoters or genes with intermediate or late in addition to early promoters; using the same criteria, the agreement was 122 of 140 (7) and 86 of 149 (6) in the previous microarray studies. Some of the differences were due to read-through transcription (Table S3) and sensitivity of RNA detection (Table S4).

Another notable aspect of this study was the determination of the 5′ boundaries of a large number of early transcripts, which allowed us to confirm the annotation of many ORFs and to correct others. The distance between 5′ boundaries of early RNA and the first AUG codon varied considerably, with a mean value of 40 nt. Localization of the 5′ ends of 83 mRNAs also allowed the prediction of their critical core promoter region. As there is no evidence for enhancer elements, it is likely that the level and perhaps timing of early gene expression is regulated by the promoter sequence.

Several previous studies provided evidence for relative or absolute decreases in cellular mRNA following VACV infection. Thus, by 7 h after high-multiplicity infection of HeLa cells with VACV, 80–90% of the RNA hybridized to VACV DNA (33). Recent microarray studies also found decreases in host mRNAs after infection of A549 cells with rabbitpox virus (9) and HeLa cells with VACV (8). We noted a rapid increase in the proportion of VACV mRNAs by direct sequencing, that is, 25–55% viral at 4 h. Because the amount of polyadenylated RNA recovered was similar between 0 and 4 h, the change in proportion of RNA is due to a combination of host RNA degradation accompanied by robust synthesis of viral RNA. The rapid decrease in thousands of cellular mRNAs indicates a global mechanism. Good candidates for regulating this process are the two VACV-encoded decapping enzymes, one with an early promoter and the other with a late promoter (11, 34). Deletion of the gene encoding the latter enzyme resulted in increased stability of both cellular and early viral mRNAs accompanied by a 10-fold reduction in the yield of progeny virus (35). Degradation of cellular mRNA can benefit VACV by blunting cellular antiviral responses and removing competition for the translation machinery. Analysis of individual host mRNAs, however, indicated

that a small subset of mRNAs showed an increase in relative amount. At 2 h, before the massive reduction in cellular mRNAs, the overrepresented functional categories of the up-regulated RNAs were NF-κB cascade, apoptosis, inhibition of apoptosis, signal transduction, and ligand-mediated signaling. The inhibition of apoptosis is a subcategory of apoptosis, suggesting that up-regulation of the genes inhibiting apoptosis was more notable than those inducing apoptosis. It seems likely that relative increases of these mRNAs represent a response of the host to a viral invader. At 4 h, the only overrepresented category was chromatin packaging and assembly, containing several histone mRNAs even though they should lack poly(A) tails as also observed in microarray studies of rabbitpox virus infection of A549 cells (9), and modified vaccinia virus infection of HeLa cells (36).

In summary, this study demonstrates the power of deep RNA sequencing to analyze virus–host interactions and viral transcriptomes in a more comprehensive way than achievable by previous methodologies. It will be interesting to apply this technology to infection of different cell types including resting cells and with other cytolytic and noncytolytic viruses.

Materials and Methods

Preparation of mRNA. Polyadenylated mRNA was isolated from VACV-infected cells by two rounds of selection with the Dynabeads mRNA Direct Kit (Invitrogen).

Generation of Whole-Transcriptome cDNA Library and SOLiD Sequencing. The strand-specific cDNA library was generated with the Whole Transcriptome Analysis Kit from Ambion using either adaptor mix A or B. The DNA library was sequenced with the Applied Biosystems SOLiD 3 system with 50-bp reads following the manufacturer's recommendations. The cDNA sequences were deposited in the Sequence Read Archive (SRA) of the National Library of Medicine under study and submission accession No. SRP002493 and No. SRA 017695, respectively.

Visualization of the VACV Transcriptome. VACV reads covering each base and VACV genome annotation were visualized with Mochiview (http://johnson-lab.ucsf.edu/sj/mochiview-software).

Cluster Analysis of Viral Gene Expression. Cluster analysis was performed with Genespring 11 (Agilent Technologies) using the hierarchical method and Chebychev metric.

ACKNOWLEDGMENTS. We thank Mariam Quinones (Bioinformatics and Computational Biosciences Branch, National Institute of Allergy and Infectious Diseases) and Kimmo Virtaneva (Research Technologies Branch, Research Technologies Section, National Institute of Allergy and Infectious Diseases) for technical suggestions and discussions, Alison McBride and Sandra Dunn (Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases) for technical help with the 2100 Bioanalyzer, and Catherine Cotter in our section for cell culture. The study was supported by the Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health.

1. Moss B (2007) Poxviridae: The viruses and their replication. *Fields Virology*, eds Knipe DM, Howley PM (Lippincott Williams & Wilkins, Philadelphia), Vol 2, pp 2905–2946.
2. Moss B (1996) Genetically engineered poxviruses for recombinant gene expression, vaccination, and safety. *Proc Natl Acad Sci USA* 93:11341–11348.
3. Baldick CJ, Jr, Keck JG, Moss B (1992) Mutational analysis of the core, spacer, and initiator regions of vaccinia virus intermediate-class promoters. *J Virol* 66:4710–4719.
4. Davison AJ, Moss B (1989) Structure of vaccinia virus early promoters. *J Mol Biol* 210: 749–769.
5. Davison AJ, Moss B (1989) Structure of vaccinia virus late promoters. *J Mol Biol* 210: 771–784.
6. Rubins KH, et al. (2008) Comparative analysis of viral gene expression programs during poxvirus infection: A transcriptional map of the vaccinia and monkeypox genomes. *PLoS One* 3:e2628.
7. Assarsson E, et al. (2008) Kinetic analysis of a complete poxvirus transcriptome reveals an immediate-early class of genes. *Proc Natl Acad Sci USA* 105:2140–2145.
8. Guerra S, et al. (2003) Cellular gene expression survey of vaccinia virus infection of human HeLa cells. *J Virol* 77:6493–6506.
9. Brum LM, Lopez MC, Varela JC, Baker HV, Moyer RW (2003) Microarray analysis of A549 cells infected with rabbitpox virus (RPV): A comparison of wild-type RPV and RPV deleted for the host range gene, SPI-1. *Virology* 315:322–334.
10. Legendre M, et al. (2010) mRNA deep sequencing reveals 75 new genes and a complex transcriptional landscape in Mimivirus. *Genome Res* 20:664–674.
11. Parrish S, Moss B (2007) Characterization of a second vaccinia virus mRNA-decapping enzyme conserved in poxviruses. *J Virol* 81:12973–12978.
12. Lee-Chen GJ, Bourgeois N, Davidson K, Condit RC, Niles EG (1988) Structure of the transcription initiation and termination sequences of seven early genes in the vaccinia virus HindIII D fragment. *Virology* 163:64–79.
13. Lee-Chen GJ, Niles EG (1988) Map positions of the 5′ ends of eight mRNAs synthesized from the late genes in the vaccinia virus HindIII D fragment. *Virology* 163:80–92.
14. Garcia AD, Aravind L, Koonin EV, Moss B (2000) Bacterial-type DNA Holliday junction resolvases in eukaryotic viruses. *Proc Natl Acad Sci USA* 97:8926–8931.
15. da Fonseca FG, Wolffe EJ, Weisberg A, Moss B (2000) Characterization of the vaccinia virus H3L envelope protein: Topology and posttranslational membrane insertion via the C-terminal hydrophobic tail. *J Virol* 74:7508–7517.
16. Lin CL, Chung CS, Heine HG, Chang W (2000) Vaccinia virus envelope H3L protein binds to cell surface heparan sulfate and is important for intracellular mature virion morphogenesis and virus infection in vitro and in vivo. *J Virol* 74:3353–3365.
17. Rosel JL, Earl PL, Weir JP, Moss B (1986) Conserved TAAATG sequence at the transcriptional and translational initiation sites of vaccinia virus late genes deduced by structural and functional analysis of the HindIII H genome fragment. *J Virol* 60: 436–449.
18. Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 2:28–36.
19. Yuen L, Moss B (1987) Oligonucleotide sequence signaling transcriptional termination of vaccinia virus early genes. *Proc Natl Acad Sci USA* 84:6417–6421.
20. Shuman S, Moss B (1988) Factor-dependent transcription termination by vaccinia virus RNA polymerase. Evidence that the *cis*-acting termination signal is in nascent RNA. *J Biol Chem* 263:6220–6225.
21. Earl PL, Hügin AW, Moss B (1990) Removal of cryptic poxvirus transcription termination signals from the human immunodeficiency virus type 1 envelope gene enhances expression and immunogenicity of a recombinant vaccinia virus. *J Virol* 64:2448–2451.
22. Luo Y, Shuman S (1991) Antitermination of vaccinia virus early transcription: Possible role of RNA secondary structure. *Virology* 185:432–436.
23. Salser W, Bolle A, Epstein R (1970) Transcription during bacteriophage T4 development: A demonstration that distinct subclasses of the "early" RNA appear at different times and that some are "turned off" at late times. *J Mol Biol* 49:271–295.
24. Ross L, Guarino LA (1997) Cycloheximide inhibition of delayed early gene expression in baculovirus-infected cells. *Virology* 232:105–113.
25. Honess RW, Roizman B (1975) Regulation of herpesvirus macromolecular synthesis: Sequential transition of polypeptide synthesis requires functional viral polypeptides. *Proc Natl Acad Sci USA* 72:1276–1280.
26. Cooper JA, Moss B (1979) In vitro translation of immediate early, early, and late classes of RNA from vaccinia virus-infected cells. *Virology* 96:368–380.
27. Cooper JA, Wittek R, Moss B (1981) Extension of the transcriptional and translational map of the left end of the vaccinia virus genome to 21 kilobase pairs. *J Virol* 39: 733–745.
28. Mahr A, Roberts BE (1984) Arrangement of late RNAs transcribed from a 7.1-kilobase EcoRI vaccinia virus DNA fragment. *J Virol* 49:510–520.
29. Boone RF, Parr RP, Moss B (1979) Intermolecular duplexes formed from polyadenylylated vaccinia virus RNA. *J Virol* 30:365–374.
30. Cochran MA, Puckett C, Moss B (1985) In vitro mutagenesis of the promoter region for a vaccinia virus gene: Evidence for tandem early and late regulatory signals. *J Virol* 54: 30–37.
31. Hirschmann P, Vos JC, Stunnenberg HG (1990) Mutational analysis of a vaccinia virus intermediate promoter in vivo and in vitro. *J Virol* 64:6063–6069.
32. Zhang YF, Keck JG, Moss B (1992) Transcription of viral late genes is dependent on expression of the viral intermediate gene G8R in cells infected with an inducible conditional-lethal mutant vaccinia virus. *J Virol* 66:6470–6479.
33. Boone RF, Moss B (1978) Sequence complexity and relative abundance of vaccinia virus mRNA's synthesized in vivo and in vitro. *J Virol* 26:554–569.
34. Parrish S, Resch W, Moss B (2007) Vaccinia virus D10 protein has mRNA decapping activity, providing a mechanism for control of host and viral gene expression. *Proc Natl Acad Sci USA* 104:2139–2144.
35. Parrish S, Moss B (2006) Characterization of a vaccinia virus mutant with a deletion of the D10R gene encoding a putative negative regulator of gene expression. *J Virol* 80: 553–561.
36. Ludwig H, et al. (2005) Role of viral factor E3L in modified vaccinia virus Ankara infection of human HeLa cells: Regulation of the virus life cycle and identification of differentially expressed host genes. *J Virol* 79:2584–2596.