Analysis of HIV-1 Expression Level and Sense of Transcription by High-Throughput Sequencing of the Infected Cell[†]

Gregory Lefebvre,¹ Sébastien Desfarges,² Frédéric Uytendaele,² Miguel Muñoz,² Niko Beerenwinkel,^{3,4} Jacques Rougemont,¹ Amalio Telenti,^{2,‡*} and Angela Ciuffi^{2,‡*}

Bioinformatics and Biostatistics Core Facility, School of Life Sciences, EPFL, Lausanne, Switzerland¹; Institute of Microbiology, University Hospital Center and University of Lausanne, Lausanne, Switzerland²; Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland³; and Swiss Institute of Bioinformatics, Basel, Switzerland⁴

Received 3 February 2011/Accepted 8 April 2011

Next-generation sequencing offers an unprecedented opportunity to jointly analyze cellular and viral transcriptional activity without prerequisite knowledge of the nature of the transcripts. SupT1 cells were infected with a vesicular stomatitis virus G envelope protein (VSV-G)-pseudotyped HIV vector. At 24 h postinfection, both cellular and viral transcriptomes were analyzed by serial analysis of gene expression followed by high-throughput sequencing (SAGE-Seq). Read mapping resulted in 33 to 44 million tags aligning with the human transcriptome and 0.23 to 0.25 million tags aligning with the genome of the HIV-1 vector. Thus, at peak infection, 1 transcript in 143 is of viral origin (0.7%), including a small component of antisense viral transcription. Of the detected cellular transcripts, 826 (2.3%) were differentially expressed between mock- and HIV-infected samples. The approach also assessed whether HIV-1 infection modulates the expression of repetitive elements or endogenous retroviruses. We observed very active transcription of these elements, with 1 transcript in 237 being of such origin, corresponding on average to 123,123 reads in mock-infected samples (0.40%) and 129,149 reads in HIV-1-infected samples (0.45%) mapping to the genomic Repbase repository. This analysis highlights key details in the generation and interpretation of high-throughput data in the setting of HIV-1 cellular infection.

There is significant interest in determining the level of HIV-1 transcription in the context of the infected cell. The impact of HIV-1 infection on cellular gene expression has been investigated in the past by gene expression arrays (3, 7, 9, 10, 15–18, 20, 21, 29, 31, 36, 37, 42, 47, 48, 51). However, the use of these arrays is limited by the set of probes that is included on the chip. Recently, it has been proposed that the unbiased analysis of transcription activity by deep sequencing will soon replace gene expression profiling by microarrays (6, 45, 49). The first application of the novel technologies in the field of HIV-1 has been in the assessment of viral sequence variation, in particular mutations present at low frequency in complex (quasispecies) populations (2, 8, 12, 19, 22, 23, 43, 46, 50, 52, 55). Pyrosequencing approaches also have been instrumental in the assessment of small noncoding RNAs in HIV-1-infected cells (54). The next goal is the joint analysis of the viral and host transcriptome of the infected cell.

Two general methods of deep sequencing are defined on the basis of the mode of sample preparation (28). One approach uses the fractionation of polyadenylated RNAs [also valid for

non-poly(A)⁺ RNAs] in short fragments, followed by reverse transcription using hexamers. Adapters are ligated to both ends and used for sequencing. This method allows the detailed characterization of the transcript structure. A second approach uses the serial analysis of gene expression (SAGE) to allow precise quantitation of poly(A)⁺ RNA and to yield information on strandedness, a key for the understanding of the antisense transcription of the HIV-1 genome. An important consideration regarding these technologies is the growing consensus that they will represent the new gold standard in the analysis of transcription; deep sequencing-based expression analysis has been shown to represent a major advance in robustness, resolution, and interlaboratory portability on multiple microarray platforms (6, 45, 49).

In the present study, we used Super-SAGE, followed by high-throughput sequencing (SOLiD; Life Technologies), also known as digital gene expression (DGE) tag profiling, 3' tag DGE, tag sequencing (Tag-Seq), or SAGE-Seq (32, 33, 38, 53), to analyze the transcriptome of a T-cell line 24 h postinfection.

MATERIALS AND METHODS

Cells. HEK 293T cells were cultured in Dulbecco's modified Eagle medium (DMEM; Invitrogen) supplemented with 10% heat-inactivated fetal calf serum (FCS) and 50 µg/ml gentamicin (D-10 culture medium). SupT1 cells (a T-cell line) were cultured in RPMI 1640 (Invitrogen) supplemented with 10% heat-inactivated fetal calf serum (FCS) and 50 µg/ml gentamicin (R-10 culture medium).

HIV vector production. To produce HIV-based vector particles, 293T cells were cotransfected with two plasmids (20 µg total) using the calcium phosphate method (Invitrogen) and according to the manufacturer's instructions. One plasmid, pNL4-3Δenv-eGFP, codes for all viral proteins except the envelope, which was disrupted and replaced by green fluorescent protein (GFP) (56). The second plasmid, pMD.G, encodes the vesicular stomatitis virus G envelope protein

* Corresponding author. Mailing address for A. Ciuffi: Institute of Microbiology, University Hospital Center and University of Lausanne, Bugnon 48, CH-1011 Lausanne, Switzerland. Phone: 4121314-40-99. Fax: 4121314-40-60. E-mail: Angela.Ciuffi@chuv.ch. Mailing address for A. Telenti: Institute of Microbiology, University Hospital Center and University of Lausanne, Bugnon 48, CH-1011 Lausanne, Switzerland. Phone: 4121314-40-97. Fax: 4121314-40-60. E-mail: Amalio.Telenti@chuv.ch.

[†] Supplemental material for this article may be found at <http://jvi.asm.org/>.

[‡] These authors contributed equally.

[▽] Published ahead of print on 20 April 2011.

FIG. 1. Overview of experimental and analytical procedures. (A) Experimental pipeline. (B) Bioinformatics pipeline.

(VSV-G) (39). Forty-eight hours after transfection, culture supernatant containing viral particles was collected, centrifuged to pellet cell debris, filtered through 0.22- μ m filters, concentrated using Centricon-Plus 70-100K (Millipore), treated with 100 U/ml DNase I (Roche), and stored frozen at -80°C . Virion concentration was assessed by measuring the CA (p24) antigen by enzyme-linked immunosorbent assay (ELISA) (Murex HIV Ag MAB; Abbott) according to the manufacturer's instructions.

HIV infection. SupT1 cells (5×10^6) were infected with 15 μ g p24 equivalent of HIV NL4-3 Δ env-eGFP/VSV-G in the presence of 5 μ g/ml Polybrene (Sigma) by centrifugation for 30 min at $1,500 \times g$. Mock-infected cells were treated

similarly but without HIV-based particles. Cells were washed with culture medium, resuspended at 10^6 cells/ml in R-10, and further incubated.

Transcriptional profiling by SAGE-Seq and bioinformatic analysis. At 24 h postexposure, cells were washed once with phosphate-buffered saline (PBS) and resuspended in 0.5 ml RNeasy Lysis Buffer (Qiagen). Total RNA was purified using an miRvana isolation kit (Ambion) or TRIzol (Invitrogen) according to the manufacturer's recommendations. RNA quality was assessed by electrophoresis on a Bioanalyzer 2100 (Total RNA Nano; Agilent). SAGE libraries were prepared with a SOLiD SAGE kit (Applied Biosystems) at the Functional Genomics Center Zurich. High-throughput sequencing was performed using SOLiD3 tech-

nology (Applied Biosystems). Sequence mapping was done with bowtie software (27), using defined analysis parameters (read length [rl], number of mismatches [n], and alignments authorized [m]). Tag hits (i.e., successfully aligned reads) were normalized according to the number of locations to which they mapped. Alignments used the human genome GRCh37, the human transcriptome RefSeq database, and the HIV-1 genome *HIV NL4-3Δenv-eGFP*. Differential expression analysis was performed using the R programming language (14) and was based on the negative binomial modeling of count data as described for the Bioconductor package DESeq (1). Genes with significantly modulated expression in HIV-1 samples compared to that of mock-infected samples were selected with a false discovery rate (FDR) of 0.05, and annotation was retrieved using the biomaRt package (11).

Northern hybridization. Total RNA (10 μg) from mock-treated or HIV vector-infected cells was supplemented with RNA loading buffer (Bioline), heated for 5 min at 65°C, and separated by electrophoresis on a 1.2% agarose-morpholine propanesulfonic acid-formaldehyde gel (44). The gel was washed three times for 10 min each with 10× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate). RNA transfer from gel to nitrocellulose membrane (Hybond N+; GE Healthcare) was performed overnight by capillarity using 20× SSC. The membrane was rinsed with 6× SSC, cross-linked with UV exposure (GeneLinker; Bio-Rad), rinsed briefly with water, and added to 10 ml preheated QuickHyb hybridization solution (Agilent) for 30 min at 65°C. Two-hundred-pmol oligonucleotide probes were labeled with 20 μCi [γ -³²P]dATP with 10 U T4 polynucleotidyl kinase (T4 PNK; New England BioLabs) for 1 h at 37°C. After the heat inactivation of the enzyme for 10 min at 95°C, probes were purified with a Qiaquick nucleotide removal kit (Qiagen), mixed with 100 μl sonicated salmon sperm DNA (10 mg/ml; Eppendorf), and added to the membrane in QuickHyb hybridization solution for 1 h at 65°C. Following hybridization, the membrane was washed twice for 15 min each with 2× SSC–0.1% SDS at room temperature, washed once for 30 min in 2× SSC–0.1% SDS at 52°C, exposed to an intensifying screen overnight at –80°C, and revealed using ImageQuantTL v2005 software on a Typhoon scanner (GE Healthcare). For reuse, the membrane was stripped with boiling 0.1× SSC–0.1% SDS twice for 15 min. Oligonucleotide probes were 5'-TCTCTCTCAGGGTCATCCATTCCA-3' for peak 1 sense transcripts, 5'-G CTCGTCCTTGACAGCTCGTCCA-3' for peak 3 sense transcripts, 5'-ATG GTGTTTACTAATCTTTTCCATGTGTT-3' for peak 4 sense transcripts, 5'-CTTGTTACACCCTGTGAGCCTGCA-3' for peak 1 antisense transcripts, 5'-CCGCCGCCGGGATCACTCTCGGCA-3' for peak 3 antisense transcripts, and 5'-GTAGACAGGATGAGGATTAACACATGGAAA-3' for peak 4 antisense transcripts. As a positive control for hybridization, 2 μg pNL4-3Δenv-eGFP digested with BglII and AflII was used.

RESULTS AND DISCUSSION

SupT1 cells (5×10^6 cells) were mock treated or were infected with a VSV-G-pseudotyped HIV-based vector expressing GFP (56) in duplicate (Fig. 1A). At 36 h postinfection, fluorescent-activated cell sorter (FACS) analysis revealed that 93% of cells were successfully transduced and expressed GFP; other cell types were tested and found to be less efficient in transduction (see Fig. S1A in the supplemental material). The use of the 24-h time point in the experimental model takes into consideration cell loss and viability, the initiation of translation as measured by GFP expression (as soon as 16 h), and the completion of the viral cycle as measured by the production and release of p24 (as soon as 20 h), reflecting viral particle release in SupT1 cells (see Fig. S1B in the supplemental material). Cells were collected for total RNA extraction, and poly(A)⁺ RNAs were reverse transcribed and processed for SAGE library preparation. cDNAs were digested with NlaIII, a frequent 4-bp cutter, and ligated to an adaptor sequence containing the recognition site of EcoP15I, an enzyme that cuts DNA asymmetrically 25 and 27 bases away. A second adapter was ligated, leading to a 27-bp transcript-specific tag surrounded by two distinct adapter sequences. High-throughput sequencing was performed using the SOLiD3 system with a universal primer annealing to the adaptor sequence. Reads

TABLE 1. Tags mapping to the human genome, human transcriptome, and HIV-1 vector genome

Sample	No. of tags for:			
	Human		HIV-1 vector ^c	
	Genome ^a	Transcriptome ^b	HIV specific	GFP specific
Mock rep1	30,668,095	38,730,420	45	69
Mock rep2	31,611,573	43,691,287	2,334	2,862
HIV rep1	27,894,076	33,337,832	117,391	135,487
HIV rep2	29,484,387	36,666,667	109,122	126,344

^a Alignment/mapping to GRCh37 allowed 1 mismatch (n = 1), 1 hit (m = 1), and a read length of 24 (rl = 24).

^b Alignment/mapping to RefSeq allowed 1 mismatch (n = 1), 10 multiple hits (m = 10), and a read length of 24 (rl = 24).

^c Alignment/mapping to *HIV NL4-3Δenv-eGFP* allowed 2 mismatches (n = 2), 2 multiple hits (m = 2), and a read length of 24 (rl = 24).

were mapped to the human genome (GRCh37), the human RefSeq database (40), the human repetitive element database (24), or the HIV-1 vector genome (*HIV NL4-3Δenv-eGFP/VSV-G*) (Fig. 1B).

Assessment of sequence alignment parameters. The variation of mapping parameters for HIV-1 tags was tested to maximize the number of tags mapping to the viral genome while minimizing the number of tags spuriously mapping to the host genome (see Fig. S2 in the supplemental material). After shortening the reads by right-trimming, we used read lengths of 21, 24, and 27 bp. In addition, we tested different numbers of mismatches (n = 0, 1, or 2) and multiple hits (m = 1, 2, or 10). We retained the conditions rl = 24, n = 1, and m = 1 as optimal for read mapping to the human genome; rl = 24, n = 1, and m = 10 for mapping to the human transcriptome; and rl = 24, n = 2, and m = 2 for mapping to the HIV-1 genome.

HIV-1 RNA tags identify sense and antisense transcription. Read mapping resulted in 33 to 44 million tags aligning with the human transcriptome and 0.23 to 0.25 million tags aligning with the genome of the HIV-1 vector (Table 1). These data suggest that, at peak infection, 1 transcript in 143 is of viral origin (0.7%). Overall, 0.33% were HIV-1 specific (1 in 309) and 0.37% were vector GFP specific (1 in 267).

The distribution of tags on the HIV-1 vector genome emphasized several relevant aspects of HIV-1 transcription (Fig. 2A). Five viral genomic regions carried 87.8% of total tags detected. A high proportion of tags (24.2%) mapped to the 3' end of the HIV-1 vector genome (Fig. 2A, peak 1), corresponding to sense transcription with a known functional polyadenylation site of HIV-1 (genome positions 9455 to 9460) and the first upstream NlaIII restriction site (position 9150). The second signal (Fig. 2A, peak 2), carrying 5.1% of total tags, corresponds to sense transcription and the same known functional polyadenylation site of HIV-1 or an alternative putative signal (at position 9108 to 9113) up to the second upstream NlaIII restriction site (position 8757). The third signal (Fig. 2A, peak 3), with 51.6% total tags, mapped at the end of the *gfp* open reading frame (ORF), with sense transcription, ending at a putative polyadenylation signal at positions 7133 to 7138 and reaching an upstream NlaIII restriction site at position 7072. The massive presence of these *gfp* tags suggests that viral transcription is skewed in favor of stable *gfp* transcripts, rendering this vector particularly suited as a reporter of HIV infection.

FIG. 2. Distribution of HIV-specific tags along the viral genome. (A) The HIV vector genome is depicted at the top, and nucleotide positioning, NlaIII restriction sites (vertical black bars), the common sense transcription start site (TSS; red arrow), and the polyadenylation signal [poly(A);

Nevertheless, it seems unlikely that this putative poly(A) signal (located in the *env* ORF) is used during the transcription of wild-type HIV-1. Further investigation of viral transcripts using wild-type HIV-1 should confirm this. The fourth signal (Fig. 2A, peak 4), representing 4.8% of total tags, is consistent with antisense viral transcription using a polyadenylation signal at positions 4908 to 4903, reported previously by Landry et al. (26), up to the first downstream *Nla*III restriction site at position 5099. The fifth signal (Fig. 2A, peak 5), with 2.1% of total tags, corresponds to antisense transcription ending at an unidentified polyadenylation site and up to the first downstream *Nla*III restriction site at position 1226. The remaining 12.2% of tags mapping elsewhere to the genome correspond mostly to the use of further *Nla*III restriction sites or alternative, putative poly(A) sites.

The technical procedure, from SAGE library preparation to high-throughput sequencing, keeps track of strand specificity. Sequences are read from a universal primer in the first adapter [on the poly(A) side] toward the *Nla*III site. Sense tags will map 3' from the *Nla*III site, while antisense tags will map 5' of the *Nla*III site. Globally, 20.5% of tags mapped to the HIV-1 vector genome in the antisense orientation (Fig. 2A). Peaks 4 and 5 are comprised mostly of antisense tags (88 and 92%, respectively); however, antisense tags also were present at the 3' end of the viral genome (40.5 and 49.0% for peaks 1 and 2, respectively). In contrast, a negligible amount of tags in the antisense orientation mapped to the third peak, corresponding to GFP (only 1 out of 125,970 tags). With the exception of peak 4, we did not identify canonical poly(A) signals (AATAAA, ATTAAA, AGTAAA, and AAGGAA) that would be associated with the observed accumulation of antisense tags. Confirming and understanding HIV-1 antisense transcription is important because of the possible roles for such transcripts in the regulation of viral expression (54), the generation of antisense proteins (26, 30, 34, 35), and the production of cryptic epitopes (4). However, the assessment of antisense transcription and cognate poly(A) motifs is not straightforward. Specifically, Landry et al. indicated that reverse transcription-PCR (RT-PCR) cannot reliably separate sense and antisense transcripts due to endogenous RT priming and therefore would require the use of 5' long terminal repeat (LTR)-deleted pNL4-3 constructs and the cloning and sequencing of the 3' rapid amplification of cDNA end (3' RACE) products (26). First, we repeated a deep sequencing experiment using bar-coded adapters (initially designed for sample multiplexing) during the library preparation of SAGE-Seq, as it was suggested that barcoding samples would generate fewer artifacts

(53). Although the overall distribution profile of the tags was comparable, only about 0.9% antisense tags were detected. To further estimate the proportion of viral antisense transcription, we performed Northern blotting using specific sense and antisense probes (Fig. 2B). Viral sense transcripts of 2-, 4-, and 9-kb classes were detected according to the specific probe used. However, none of the probes designed to anneal to antisense transcripts revealed any detectable signal. These data suggest that viral antisense transcription is more consistent with 0.9% abundance than with 20.5% abundance, calling for caution when analyzing antisense transcription by novel high-throughput technologies and highlighting the necessity to validate data by alternative methods.

Three additional technical aspects deserve comment. The use of the HIV-1 vector containing a GFP ORF in the place of *env* could disturb transcription and splicing. The Northern blot shown in Fig. 2B excludes significant defects in viral transcription and splicing. The integrity of transcription was further assessed by RT-quantitative PCR using primers flanking the major introns of HIV to confirm the detection of multiply, singly, and unspliced transcripts (data not shown). We examined the degree of conservation of the tag sequence to assess whether the tags displayed a perfectly matched sequence corresponding to the HIV-1 vector DNA. We tested 0, 1, and 2 mismatches to ensure that we would capture the putative error rate of the viral reverse transcriptase as well as the base miscalling of the SOLiD sequencing software. Upon HIV vector-specific tag analysis, 96.97% of the tags displayed a perfect match with the HIV vector genome, 2.92% presented 1 mismatch, and only 0.11% presented 2 mismatches. We also identified a limited number of reads (<0.01%) that mapped to the HIV-1 vector genome in the mock-treated cells (Table 1). Upon inspection, all tags were HIV specific, indicating that a small level of contamination occurred during the experimental procedure (from culture to library preparation).

Cellular transcriptome during infection. The high-throughput sequencing analysis of SAGE-RNA allowed the detection of 36,271 genes expressed by humans (79.8%) according to the RefSeq database (Table 1). On average, each transcript was represented by 815 tags (ranging from 0 to 3.8 million and with a median of 14.7). While the overall distribution in expression levels was not different for mock- and HIV-1-infected cells (Fig. 3), the identity of transcripts contained in mock- or HIV-infected samples was significantly modified. Of the detected transcripts, 826 (2.3%) were significantly differentially modulated between mock- and HIV-infected samples at a stringent adjusted *P* value of 10^{-4} at 24 h postinfection. Ingenuity path-

red vertical bar] are shown below. Multiple antisense transcription start sites (distributed between positions 9175 and 8714; only two are drawn here for figure simplicity; green arrows) and the poly(A) signal (vertical green bar), as described in Landry et al. (26), are indicated. HIV-1 open reading frames also are indicated, including that of the putative antisense protein (ASP). The lower portion of this panel indicates the density of tags (vertical axis) distributed along the HIV-1 vector genome (*x* axis) for both HIV-1 samples (HIV rep1 and HIV rep2) and mock infection samples (Mock rep1 and Mock rep2). The five major peaks are numbered at the bottom, displaying the number of tags at each peak as well as the proportion of these tags representing sense or antisense transcription. (B) Total RNA from mock-infected or HIV-based vector-infected cells was extracted using miRvana or Trizol at 22 and 24 h postexposure and subjected to Northern blot hybridization. Left panel, total RNA gel electrophoresis; right two panels, Northern blots using strand-specific oligonucleotide probes aligning with the corresponding peak of tags detected by SAGE-Seq. Probes were designed to anneal with sense transcripts (red probes; top) or antisense transcripts (green probes; bottom) with the nucleic acid sequence complementary to the identified peak tag sequence. The plasmid encoding the HIV vector genome digested with *Bgl*II and *Afl*II was used as a positive control for hybridization (+).

FIG. 3. Analysis of cellular transcripts in mock- and HIV-infected cells. Average tag density analysis in mock-infected (black lines) and HIV-1-infected (red lines) samples for total cellular transcripts (lines), repetitive elements (dashed lines), and HERV elements (dotted lines) is shown. Most cellular transcripts were detected with 1 to 1,000 tags (log 0 to 3; average, 2.9 log) across the complete transcriptome. Only a few transcripts were identified in larger numbers (e.g., above 10,000; log 4). A smaller number of tags was observed for the expressed repetitive elements (average, 2.3 log), including HERVs (average, 1.9 log). Statistical differences between mock infection and HIV-1 infection distributions were assessed by Wilcoxon test.

way analysis identified “cellular growth and proliferation” and “RNA posttranscriptional modification” as prominent modulated functions, and it indicated that networks related to cell cycle, DNA replication and repair, gene expression, and cell death were relevant features. There was downregulation in HIV-infected cells of many of the genes associated with those networks, which is consistent with marked cellular compromise and stress. The complete set of differentially expressed genes is included in Fig. S3 in the supplemental material. Of note, some of the differentially expressed genes include non-protein-coding transcripts, such as spliceosomal and small nucleolar RNAs, that are not generally thought of as polyadenylated and should not have been captured by SAGE. However, there is increasing evidence that poly(A) tails can be added to such RNAs, possibly marking them for degradation (5, 25).

We assessed the overlap between the sets of genes that were identified as differentially expressed in the current study and those from various microarray studies in the literature that investigated differential expression in CD4⁺ T cells (42), CD8⁺ T cells (41), monocytes (17), and lymph nodes (29) during HIV-1 infection *in vivo*, as well as with a set of validated genes compiled by Giri et al. (16) from microarray studies published between 2000 and 2006. The overlap varied from 4% in the study of circulating monocytes in HIV-infected individuals (17) to 8% in the analysis of lymphatic tissue in the setting of various stages of HIV-1 infection *in vivo* (29). There also was a 6% overlap with a set of validated genes curated by Giri et al. (16). The final shortlist of 52 genes common between the present and one or more of the previous studies (see Fig. S4 in the supplemental material) also was enriched for genes involved in cell cycle, DNA replication and repair, and cell proliferation. However, it should be underscored that differences

in the nature of the techniques (deep-sequencing versus microarray), a lack of recent studies using new-generation microarrays for the transcriptome analysis of *in vitro* infection of T-cell lines (comparable to the experiment completed herein), and general differences in study design makes comparison between studies of unclear significance.

The approach also assessed whether HIV-1 infection modulates the expression of repetitive elements or endogenous retroviruses. We observed very active transcription of these elements, with 1 transcript in 237 being of such origin, corresponding to an average of 123,123 reads in mock infection (0.40%) and 129,149 reads in HIV-1 (0.45%) mapping to the genomic Repbase repository (Fig. 3). Of these, 4,021 and 4,451 average reads mapped to human endogenous retroviruses (HERV) in mock- and HIV-infected cells, respectively (Fig. 3). The modest increase in HERV in the HIV-infected cell results mainly from the contribution of HERVK, which increases from 3,331 to 3,814 average tags. The biological role of the prominent transcriptional activity of endogenous retroelements, i.e., the regulatory activity or generation of translated products, needs further analysis (13).

Conclusions. The analysis of deep sequence data using SAGE allowed the precise measurement of expression levels of the proviral genome in HIV-1-infected cells and identified a small component of antisense viral transcription. Improvements in RNA-Seq will increasingly deliver information on both strand specificity and the nature of the transcript, including splice forms and sequence variation, that will facilitate the study of the dynamics of virus-host interactions (28). Extending this approach to replication-competent HIV-1 isolates and to different cellular backgrounds may reveal differences in virus-host interactions due to specific strain or cellular factors.

ACKNOWLEDGMENTS

We thank Marzanna Künzli and Sirisha Aluri from the Functional Genomics Center Zurich (FGCZ) facility for SAGE library preparation and SOLiD sequencing, with great help and support from Gerrit Kuhn (Life Technologies). The following reagent was obtained through the NIH AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: pNL4-3-deltaE-EGFP (no. 11100) from Haili Zhang, Yan Zhou, and Robert Siliciano (56).

This work was supported by the Swiss National Science Foundation, grant 310030-130699.

G.L. and J.R. performed the bioinformatics analyses, F.U. and M.M. carried out HIV-1 production, cell infections, and RNA extraction, S.D. performed Northern blot analyses, A.T. and A.C. led the project and wrote the manuscript, and N.B. edited and proofread the manuscript.

We declare that we have no competing interests.

REFERENCES

- Anders, S., and W. Huber. 2010. Differential expression analysis for sequence count data. *Nat. Precedings* **2010**:1–5.
- Archer, J., et al. 2009. Detection of low-frequency pretherapy chemokine (CXC motif) receptor 4 (CXCR4)-using HIV-1 with ultra-deep pyrosequencing. *AIDS* **23**:1209–1218.
- Arendt, C. W., and D. R. Littman. 2001. HIV: master of the host cell. *Genome Biol.* **2**:REVIEWS103.
- Bansal, A., et al. 2010. CD8 T cell response and evolutionary pressure to HIV-1 cryptic epitopes derived from antisense transcription. *J. Exp. Med.* **207**:51–59.
- Bayne, E. H., S. A. White, and R. C. Allshire. 2007. DegrAAAAd into silence. *Cell* **129**:651–653.
- Blow, N. 2009. Transcriptomics: the digital generation. *Nature* **458**:239–242.
- Borjabad, A., A. I. Brooks, and D. J. Volsky. 2010. Gene expression profiles of HIV-1-infected glia and brain: toward better understanding of the role of

- astrocytes in HIV-1-associated neurocognitive disorders. *J. Neuroimmune Pharmacol.* **5**:44–62.
8. **Bushman, F. D., et al.** 2008. Massively parallel pyrosequencing in HIV research. *AIDS* **22**:1411–1415.
 9. **Carroll, E. E., et al.** 2006. Altered gene expression in asymptomatic SHIV-infected rhesus macaques (*Macaca mulatta*). *Viol. J.* **3**:74.
 10. **Chung, H. K., et al.** 2008. Cellular gene expression profiles in rhesus macaques challenged mucosally with a pathogenic R5 tropic simian human immunodeficiency virus isolate. *Viral Immunol.* **21**:411–423.
 11. **Durinck, S., et al.** 2005. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**:3439–3440.
 12. **Eriksson, N., et al.** 2008. Viral population estimation using pyrosequencing. *PLoS Comput. Biol.* **4**:e1000074.
 13. **Garrison, K. E., et al.** 2007. T cell responses to human endogenous retroviruses in HIV-1 infection. *PLoS Pathog.* **3**:e165.
 14. **Gentleman, R. C., et al.** 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**:R80.
 15. **George, M. D., D. Verhoeven, Z. McBride, and S. Dandekar.** 2006. Gene expression profiling of gut mucosa and mesenteric lymph nodes in simian immunodeficiency virus-infected macaques with divergent disease course. *J. Med. Primatol.* **35**:261–269.
 16. **Giri, M. S., M. Nebozhyn, L. Showe, and L. J. Montaner.** 2006. Microarray data on gene modulation by HIV-1 in immune cells: 2000–2006. *J. Leukoc. Biol.* **80**:1031–1043.
 17. **Giri, M. S., et al.** 2009. Circulating monocytes in HIV-1-infected viremic subjects exhibit an antiapoptosis gene signature and virus- and host-mediated apoptosis resistance. *J. Immunol.* **182**:4459–4470.
 18. **Harman, A. N., et al.** 2009. HIV-1-infected dendritic cells show 2 phases of gene expression changes, with lysosomal enzyme activity decreased during the second phase. *Blood* **114**:85–94.
 19. **Hoffmann, C., et al.** 2007. DNA bar coding and pyrosequencing to identify rare HIV drug resistance mutations. *Nucleic Acids Res.* **35**:e91.
 20. **Hyrca, M. D., et al.** 2007. Distinct transcriptional profiles in ex vivo CD4+ and CD8+ T cells are established early in human immunodeficiency virus type 1 infection and are characterized by a chronic interferon response as well as extensive transcriptional changes in CD8+ T cells. *J. Virol.* **81**:3477–3486.
 21. **Imbeault, M., M. Ouellet, and M. J. Tremblay.** 2009. Microarray study reveals that HIV-1 induces rapid type-I interferon-dependent p53 mRNA up-regulation in human primary CD4+ T cells. *Retrovirology* **6**:5.
 22. **Ji, H., et al.** 2010. HIV drug resistance surveillance using pooled pyrosequencing. *PLoS One* **5**:e9263.
 23. **Jojic, V., T. Hertz, and N. Jovic.** 2008. Population sequencing using short reads: HIV as a case study. *Pac. Symp. Biocomput.* **2008**:114–125.
 24. **Jurka, J., et al.** 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**:462–467.
 25. **LaCava, J., et al.** 2005. RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* **121**:713–724.
 26. **Landry, S., et al.** 2007. Detection, characterization and regulation of antisense transcripts in HIV-1. *Retrovirology* **4**:71.
 27. **Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg.** 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**:R25.
 28. **Levin, J. Z., et al.** 2010. Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat. Methods* **7**:709–715.
 29. **Li, Q., et al.** 2009. Microarray analysis of lymphatic tissue reveals stage-specific, gene expression signatures in HIV-1 infection. *J. Immunol.* **183**:1975–1982.
 30. **Ludwig, L. B., et al.** 2006. Human immunodeficiency virus-type 1 LTR DNA contains an intrinsic gene producing antisense RNA and protein products. *Retrovirology* **3**:80.
 31. **Martinez-Mariño, B., H. Foster, Y. Hao, and J. A. Levy.** 2007. Differential gene expression in CD8(+) cells from HIV-1-infected subjects showing suppression of HIV replication. *Virology* **362**:217–225.
 32. **Matsumura, H., D. H. Kruger, G. Kahl, and R. Terauchi.** 2008. SuperSAGE: a modern platform for genome-wide quantitative transcript profiling. *Curr. Pharm. Biotechnol.* **9**:368–374.
 33. **Matsumura, H., et al.** 2010. High-throughput SuperSAGE for digital gene expression analysis of multiple samples using next generation sequencing. *PLoS One* **5**:e12010.
 34. **Michael, N. L., et al.** 1994. Negative-strand RNA transcripts are produced in human immunodeficiency virus type 1-infected cells and patients by a novel promoter downregulated by Tat. *J. Virol.* **68**:979–987.
 35. **Miller, R. H.** 1988. Human immunodeficiency virus may encode a novel protein on the genomic DNA plus strand. *Science* **239**:1420–1422.
 36. **Mitchell, R., C. Y. Chiang, C. Berry, and F. Bushman.** 2003. Global analysis of cellular transcription following infection with an HIV-based vector. *Mol. Ther.* **8**:674–687.
 37. **Montano, M., et al.** 2006. Gene-expression profiling of HIV-1 infection and perinatal transmission in Botswana. *Genes Immun.* **7**:298–309.
 38. **Morrissey, A. S., et al.** 2009. Next-generation tag sequencing for cancer gene expression profiling. *Genome Res.* **19**:1825–1835.
 39. **Naldini, L., et al.** 1996. In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science* **272**:263–267.
 40. **Pruitt, K. D., T. Tatusova, and D. R. Maglott.** 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35**:D61–D65.
 41. **Rotger, M., et al.** Comparative transcriptomics of extreme phenotypes of human HIV-1 infection and SIV infection in sooty mangabey and rhesus macaque. *J. Clin. Invest.*, in press.
 42. **Rotger, M., et al.** 2010. Genome-wide mRNA expression correlates of viral control in CD4+ T-cells from HIV-1-infected individuals. *PLoS Pathog.* **6**:e1000781.
 43. **Rozera, G., et al.** 2009. Massively parallel pyrosequencing highlights minority variants in the HIV-1 env quasispecies deriving from lymphomonocyte subpopulations. *Retrovirology* **6**:15.
 44. **Sambrook, J. F., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed., vol. 1, p. 7.39–37.52. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
 45. **Shendure, J.** 2008. The beginning of the end for microarrays? *Nat. Methods* **5**:585–587.
 46. **Simen, B. B., et al.** 2009. Low-abundance drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naïve patients significantly impact treatment outcomes. *J. Infect. Dis.* **199**:693–701.
 47. **Sirois, M., et al.** 2008. R5 and X4 HIV viruses differentially modulate host gene expression in resting CD4+ T cells. *AIDS Res. Hum. Retrovir.* **24**:485–493.
 48. **Solis, M., et al.** 2006. Gene expression profiling of the host response to HIV-1 B, C, or A/E infection in monocyte-derived dendritic cells. *Virology* **352**:86–99.
 49. **'t Hoen, P. A., et al.** 2008. Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. *Nucleic Acids Res.* **36**:e141.
 50. **Tsibris, A. M., et al.** 2009. Quantitative deep sequencing reveals dynamic HIV-1 escape and large population shifts during CCR5 antagonist therapy in vivo. *PLoS One* **4**:e5683.
 51. **Vahey, M. T., et al.** 2008. CD4+ T-cell decline after the interruption of antiretroviral therapy in ACTG A5170 is predicted by differential expression of genes in the ras signaling pathway. *AIDS Res. Hum. Retrovir.* **24**:1047–1066.
 52. **Wang, G. P., A. Ciuffi, J. Leipzig, C. C. Berry, and F. D. Bushman.** 2007. HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. *Genome Res.* **17**:1186–1194.
 53. **Wu, Z. J., et al.** 2010. Gene expression profiling of human breast tissue samples using SAGE-Seq. *Genome Res.* **20**:1730–1739.
 54. **Yeung, M. L., et al.** 2009. Pyrosequencing of small non-coding RNAs in HIV-1 infected cells: evidence for the processing of a viral-cellular double-stranded RNA hybrid. *Nucleic Acids Res.* **37**:6575–6586.
 55. **Zagordi, O., R. Klein, M. Daumer, and N. Beerenwinkel.** 2010. Error correction of next-generation sequencing data and reliable estimation of HIV quasispecies. *Nucleic Acids Res.* **38**:7400–7409.
 56. **Zhang, H., et al.** 2004. Novel single-cell-level phenotypic assay for residual drug susceptibility and reduced replication capacity of drug-resistant human immunodeficiency virus type 1. *J. Virol.* **78**:1718–1729.